PLANNING AND EVALUATION SERVICE

Does Professional Development Change Teaching Practice? Results from a Three-Year Study

2000

U.S. DEPARTMENT OF EDUCATION   OFFICE OF THE UNDER SECRETARY

DOC #2000-04

Does Professional Development Change Teaching Practice? Results from a Three-Year Study

October 2000

***Prepared by*:**

Andrew C. Porter, University of Wisconsin, Madison
Michael S. Garet, American Institutes for Research
Laura Desimone, American Institutes for Research
Kwang Suk Yoon, American Institutes for Research
Beatrice F. Birman, American Institutes for Research

***Project Management*:**

Principal Investigator, Andrew C. Porter
Project Director, Beatrice F. Birman
Deputy Project Director, Michael S. Garet

Prepared for:

U.S. Department of Education
Office of the Under Secretary

This report was prepared for the U.S. Department of Education under Contract Number EA 970001001. The views expressed herein are those of the contractor. No official endorsement by the U.S. Department of Education is intended or should be inferred.

U.S. Department of Education
Richard W. Riley
Secretary

Office of the Under Secretary
Judith A. Winston
Under Secretary (A)

Planning and Evaluation Service
Alan L. Ginsburg
Director

Elementary and Secondary Education Division
Ricky T. Takai
Director

October 2000

To order copies of this report, write:

ED Pubs
Education Publications Center
U. S. Department of Education
P. O. Box 1398
Jessup, MD 20794-1398;

via fax, dial (301) 470-1244;

or via electronic mail, send your request to: edpubs@inet.ed.gov.

You may also call toll-free: 1-877-433-7827 (1-877-4-ED-PUBS). If 877 service is not yet available in your area, call 1-800-872-5327 (1-800-USA-LEARN); Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY), should call 1-800-437-0833.

To order online, point your Internet browser to: www.ed.gov/pubs/edpubs.html.

This report is also available on the Department's Web site at: www.ed.gov/offices/OUS/eval/.

On request, this publication is available in alternative formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center (202) 260-9895 or (202) 205-811

CONTENTS

EXHIBITS

EXECUTIVE SUMMARY

INTRODUCTION

SECTION I: THE STUDY OF PROFESSIONAL DEVELOPMENT AND TEACHER CHANGE: BUILDING ON NATIONAL, CROSS-SECTIONAL FINDINGS WITH LONGITUDINAL DATA

SECTION II: TEACHER PARTICIPATION IN PROFESSIONAL DEVELOPMENT

SECTION III: THE EFFECTS OF PROFESSIONAL DEVELOPMENT ON TEACHING PRACTICE IN MATHEMATICS AND SCIENCE

SECTION IV: TRENDS IN TEACHING PRACTICE

APPENDIX A: STUDENT ACHIEVEMENT TRENDS IN THE 30 SCHOOLS

APPENDIX B: SAMPLE DESIGN AND RESPONSE RATES FOR THE LONGITUDINAL TEACHER SURVEY

APPENDIX C: ANALYSIS METHODS AND RESULTS FOR VARIATION IN TEACHERS' PROFESSIONAL DEVELOPMENT EXPERIENCES

INTRODUCTION

What are the characteristics of professional development that improve teaching practice? This report addresses that question, using data from the National Evaluation of the Eisenhower Professional Development Program. The Eisenhower program is part of the federal government's efforts to support education reform based on high standards. The success of standards-based reforms depends on teachers' ability to foster both basic knowledge and advanced thinking and problem solving among their students (Loucks-Horsley, Hewson, Love, & Stiles, 1998; National Commission on Teaching & America's Future, 1996), and such effective practices require teachers to have a deep understanding of the content they teach (Ma, 1999). Professional development is considered an essential mechanism for deepening teachers' content knowledge and developing their teaching abilities. As a result, it is a cornerstone of systemic reform efforts designed to increase teachers' capacity to teach to high standards (Smith & O'Day, 1991).

The Eisenhower Professional Development Program, Title II of the Elementary and Secondary Education Act (ESEA), is the federal government's largest investment that is solely focused on developing the knowledge and skills of classroom teachers. The program is a primary means for helping schools and school districts across the nation meet the U.S. Department of Education's objective of ensuring that a "talented and dedicated teacher is in every classroom in America" (U.S. Department of Education, 1999c). Part B of the program, with a FY 2000 appropriation of $335 million, provides funds through state education agencies (SEAs) to school districts and through state agencies for higher education (SAHEs) to institutions of higher education and nonprofit organizations (SAHE grantees). These funds primarily support professional development in mathematics and science, but also in other content areas. The goal of the Eisenhower Professional Development Program is to support professional development experiences for teachers that enhance classroom teaching and, ultimately, improve student learning.

This report focuses on the effects of professional development on improving classroom teaching practice.[1] Using a purposefully selected sample of teachers in 30 schools, in 10 districts, in 5 states, we examine the quality of teachers' professional development in Eisenhower and other professional development activities and its effects on changing teaching practice in mathematics and science from 1996–1999.

This is the third in a series of reports based on a multiyear evaluation of the Eisenhower Program, conducted by the American Institutes for Research (AIR) under contract with the U.S. Department of Education's Planning and Evaluation Service.[2] The national Evaluation of the

[1] In this report, we use the terms *teaching practice*, *classroom practice*, *classroom instruction*, and *instruction* interchangeably.

[2] The first report was based on six exploratory case studies of school districts conducted at the beginning of the evaluation, in the spring of 1997. See *The Eisenhower Professional Development Program: Emerging Themes from Six Districts*, by B. F. Birman, A. L. Reeve, and C. L. Sattler, 1998, Washington, DC: U.S. Department of Education. The purpose of that report was to obtain an initial description of the Eisenhower program and the issues that it faced in different local contexts. The second report described the status of the program on several dimensions, such as features of quality and management and implementation; the report also linked these dimensions to characteristics of the professional development and to teachers' self-reported outcomes. It was based primarily on data from three national probability samples: (1) district Eisenhower coordinators, (2) Eisenhower project directors in SAHE grantees (i.e., the institutions of higher education and nonprofit organizations supported through the SAHE component of the program), and (3) teachers participating in Eisenhower-assisted professional

Eisenhower Program, begun in 1996, includes three strands of data collection, each with unique strengths:

(1) The *National Profile*, which collected data from national probability samples of district Eisenhower coordinators, SAHE-grantee project directors, and teachers who participated in Eisenhower-assisted professional development (i.e., activities sponsored in full or in part by Eisenhower funds). This component of the evaluation provided data that are generalizable to all districts receiving Eisenhower funds, all teachers who participate in Eisenhower-assisted professional development, and all SAHE-grantee projects.

(2) The *Case Studies*, which provided detailed information about how the Eisenhower program operates in 10 schools districts—two school districts in each of five states: Kentucky, New York, Ohio, Texas, and Washington. Data from this component provided a detailed context for interpreting the quantitative findings.

(3) The *Longitudinal Study of Teacher Change*, which surveyed all mathematics and science teachers in 30 schools—three schools (one elementary, one middle and one high school) in each of the 10 case-study districts—at three points in time.[3] These data allow us to examine teachers' professional development and teaching practice over time.

This report draws on the survey data from our Longitudinal Study of Teacher Change (LSTC) and augments the results from our earlier work based on the National Profile and Case Studies. Although the National Eisenhower Evaluation as a whole was designed to focus on several research questions addressing the type and quality of Eisenhower activities, who participates in them, how they fit into other reform efforts, and how they are managed and implemented, this report focuses on one particular research question:

Do teachers' experiences in Eisenhower-assisted professional development activities, in the context of other professional development activities, contribute to changes in teaching practice?

The LSTC was not based on a national sample; it was a purposefully selected sample of teachers in 30 schools, in 10 districts, in 5 states. The LSTC examined the quality of teachers' professional development in Eisenhower and other professional development activities and the effects of professional development on changing teaching practice in mathematics and science from 1996 to 1999.

Exhibit 1 illustrates the conceptual framework for the entire national evaluation and highlights where the Longitudinal Study of Teacher Change fits into the overall study. Starting with the box on the far right, we show that improving teaching practice is the goal of the Eisenhower legislation. From the next box on the left, we see that teacher experiences in Eisenhower-assisted professional development activities are intended to improve teaching practice. The quality of the activities that districts and SAHE grantees make available, and the ways that districts and SAHE

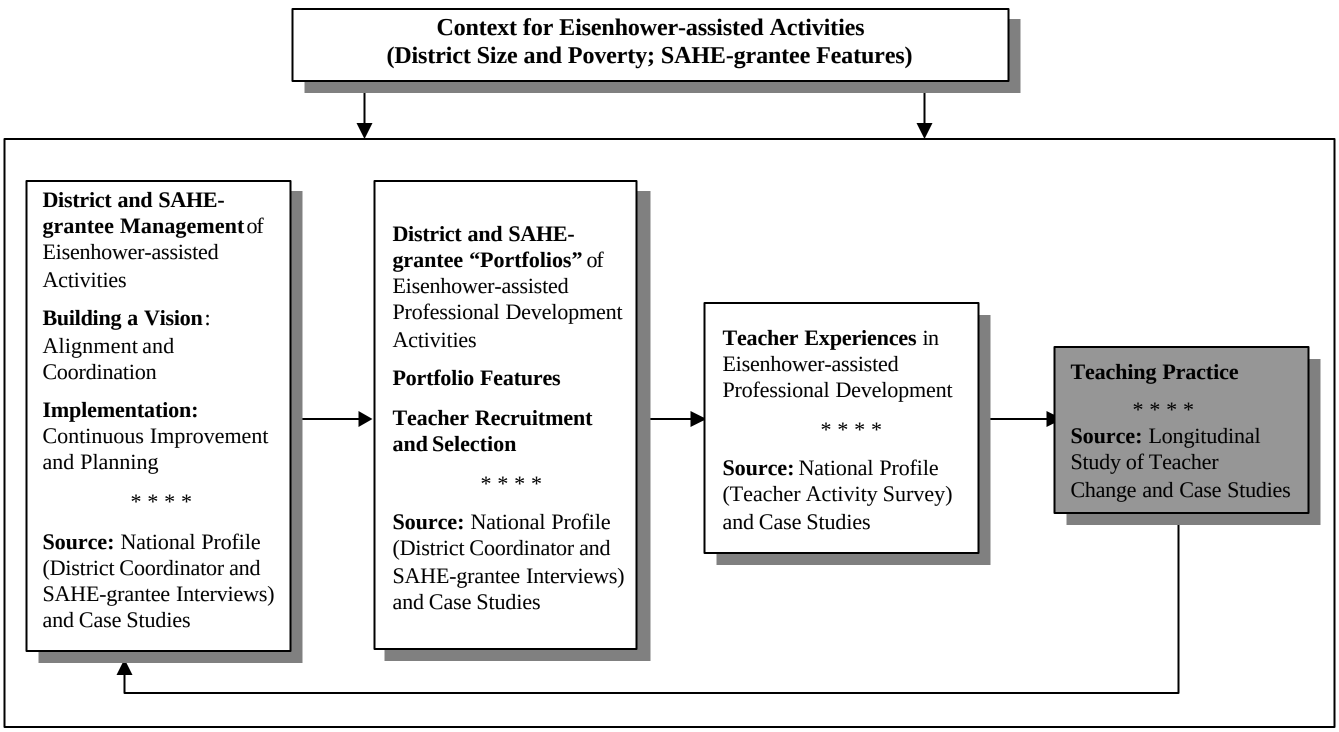
development (i.e., professional development that was sponsored, at least in part, by Eisenhower funds). In addition, the second report drew on data from 10 in-depth case studies in five states. See *Designing Effective Professional Development: Lessons from the Eisenhower Program*, by M. Garet, B. Birman, A. Porter, L. Desimone, and R. Herman, R. with K. Suk Yoon, 1999, Washington, DC: U.S. Department of Education.

[3] The Longitudinal Study of Teacher Change also included interviews and classroom observations of teachers in the 30 schools. Results of these data are reported in Garet et al., (1999).

grantees select teachers to participate, in turn influence teacher experiences in Eisenhower-assisted professional development. We hypothesized that the overall quality of Eisenhower-assisted activities is shaped by the degree of integration of the Eisenhower-assisted activities with other professional development and systemic reform efforts, as well as by how districts and SAHE grantees plan and evaluate Eisenhower-funded activities. This report describes the part of the evaluation that focuses on classroom teaching practice.

EXHIBIT 1

Conceptual Framework for the National Evaluation of the Eisenhower Professional Development Program



To describe the results of our Longitudinal Study of Teacher Change, we divide our analysis and reporting into five sections. First, we summarize the results from our national study and describe the design of the Longitudinal Study of Teacher Change and the way it builds on our national findings. Second, we describe the quality of professional development experienced by teachers in our longitudinal sample. Third, we explore the effects of professional development on teaching practice. Fourth, we examine trends in teaching practice and discuss how they inform our findings on the effectiveness of professional development in changing teachers' instruction. The fifth and last section of the report summarizes our results and suggests implications for the Eisenhower and other professional development programs to increase their effectiveness in fostering teacher change.

SECTION I

The Study of Professional Development and Teacher Change: Building on National, Cross-Sectional Findings with Longitudinal Data

Over the past decade, a large body of literature has emerged on professional development, teacher learning, and teacher change.[4] The research literature contains a mix of large- and small-scale studies, including intensive case studies of classroom teaching, evaluations of programs designed to improve teaching and learning, and surveys of teachers about their preservice preparation and in-service professional development experiences.[5] In addition, there is a considerable amount of literature describing "best practices" in professional development, drawing on expert experiences (e.g., Loucks-Horsley et al., 1998). Despite the amount of literature, however, relatively little systematic research has been conducted on the *effects* of professional development on improving teaching or on improving student outcomes.

Although relatively little research has been conducted on the effects of alternative forms of professional development, the research that has been conducted, along with the experience of expert practitioners, provides some preliminary guidance about the characteristics of high-quality professional development. Characteristics of professional development that are identified as "high quality" or "effective" include a focus on content; in-depth, active learning opportunities; links to high standards, opportunities for teachers to engage in leadership roles; extended duration; and the collective participation of groups of teachers from the same school, grade, or department. (See, in particular, Garet et al., 1999; Hiebert, 1999; Loucks-Horsley et al., 1998; U.S. Department of Education, 1999b.)

Although lists of characteristics such as these commonly appear in the literature on effective professional development, there is little direct evidence on the extent to which these characteristics are related to better teaching and increased student achievement. Some studies conducted over the past decade suggest that professional development experiences that share all or most of these characteristics can have a substantial, positive influence on teachers' classroom practice and student achievement.[6] Several recent studies have begun to examine the relative importance of specific dimensions or characteristics of professional development. For example, a number of recent studies

[4] See Richardson & Placier (in press) for a comprehensive review of the literature on teacher learning and professional development.
[5] See, for example, Cohen (1990) for an intensive case study of change in mathematics teaching; Carey and Frechtling (1997) for a program evaluation of exemplary professional activities in science; and U.S. Department of Education (1999a) for a national survey of teachers focused on teacher preparation and qualifications.
[6] See, for example, Fennema et al. (1996), an experimental study examining the effects of Cognitively Guided Instruction, an intervention in elementary school mathematics; Wilson and Ball (1991), an intensive case study of two teachers who participated in the Summer Math program; and Cohen and Hill (1998), which describes the relationship between participation in professional development, teaching practice, and student achievement, using survey data from California. See Kennedy (1998) for a review of available randomized studies examining the effects of teacher professional development on student achievement in mathematics and science. See Shields, Marsh, and Adelman (1998) for a recent examination of the effects of the National Science Foundation (NSF) Statewide Systemic Initiatives (SSIs) on classroom practice in mathematics and science; and Weiss, Montgomery, Ridgway, and Bond (1998) for an examination of the effects of the NSF Local Systemic Change (LSC) initiatives.

suggest that the intensity and duration of professional development is related to the degree of teacher change (Shields, Marsh, & Adelman, 1998; Weiss, Montgomery, Ridgway, and Bond 1998). Furthermore, there is some indication that professional development that focuses on specific mathematics and science content and the ways students learn such content is especially helpful, particularly for instruction designed to improve students' conceptual understanding (Cohen & Hill, 1998; Fennema et al., 1996). However, few studies have explicitly compared the effects of different forms of professional development on teaching and learning.[7] Further, most studies of professional development have not examined its effects on a national scale.

Given the need for new, systematic research on the effectiveness of alternative strategies for professional development, we designed our evaluation of the Eisenhower Professional Development Program to enable us to examine the relationship between professional development and change in teaching practice in both a cross-sectional, national probability sample of teachers and a smaller, longitudinal sample of teachers. The Eisenhower program can then be evaluated in terms of the frequency with which program funds are used to provide professional development with features found to be effective. The results from our national sample of teachers are described in detail in our second-year report, *Designing Effective Professional Development: Lessons from the Eisenhower Program* (Garet et al., 1999). Below we summarize these results and explain how they serve as the foundation for our longitudinal study of teachers.

What We Know About Eisenhower Professional Development and Teacher Outcomes: Lessons from Our National Data

The Longitudinal Study of Teacher Change is designed to build on the national, cross-sectional data that we examined in detail in our second-year report. In Garet et al. (1999), we described results from our Teacher Activity Survey, a mail survey of a national probability sample of 1,027 teachers who participated in 657 Eisenhower-assisted activities during the 1997–98 school year.[8] We used the survey of teachers' professional development activities to assess the effectiveness of Eisenhower-assisted activities, examine the quality of Eisenhower-assisted activities, and assess the strength of the relationships between features of the activities in which teachers participated and teachers' self-reported outcomes.

To measure the quality of Eisenhower-assisted activities, we integrated and operationalized the ideas in the literature on "best practices" in professional development. We focused on three "structural features," or characteristics of the structure of a professional development activity. These structural features include the form or organization of the activity—that is, whether the activity is organized as a **reform type,** such as a study group, teacher network, mentoring relationship, committee or task force, internship, individual research project, or teacher research center, in contrast to a traditional workshop, course, or conference; the **duration** of the activity, including the total number of contact hours that participants are expected to spend in the activity, as well as the span of time over which the activity takes place; and the degree to which the activity emphasizes the

[7] Kennedy (1998) and Cohen and Hill (1998) are among the few examples of studies that compare the relative effectiveness of different forms of professional development. Both studies conclude that professional development focused on the teaching and learning of specific mathematics and science content is more effective than more general professional development.
[8] The mail survey of teachers represents a response rate of 72 percent of sampled teachers. Details regarding sampling design and methodology are provided in Garet et al., 1999.

collective participation of groups of teachers from the same school, department, or grade level, as opposed to the participation of individual teachers from many schools.

In addition to these structural features, we focused on three "core features" or characteristics of the substance of the professional development experience itself: the extent to which the activity offers opportunities for **active learning—**that is, opportunities for teachers to become actively engaged in the meaningful analysis of teaching and learning, for example, by reviewing student work or obtaining feedback on their teaching; the degree to which the activity promotes **coherence** in teachers' professional development, by incorporating experiences that are consistent with teachers' goals, aligned with state standards and assessments, and encourage continuing professional communication among teachers; and the degree to which the activity has a **content focus—**that is, the degree to which the activity is focused on improving and deepening teachers' content knowledge in mathematics and science.

- **We found that the six key features of high-quality professional development led to increases in teachers' self-reported knowledge and skills and changes in teaching practice: three structural features (characteristics of the structure of the activity)—reform type, duration, and collective participation— and three core features (characteristics of the substance of the activity)—active learning, coherence, and content focus.**

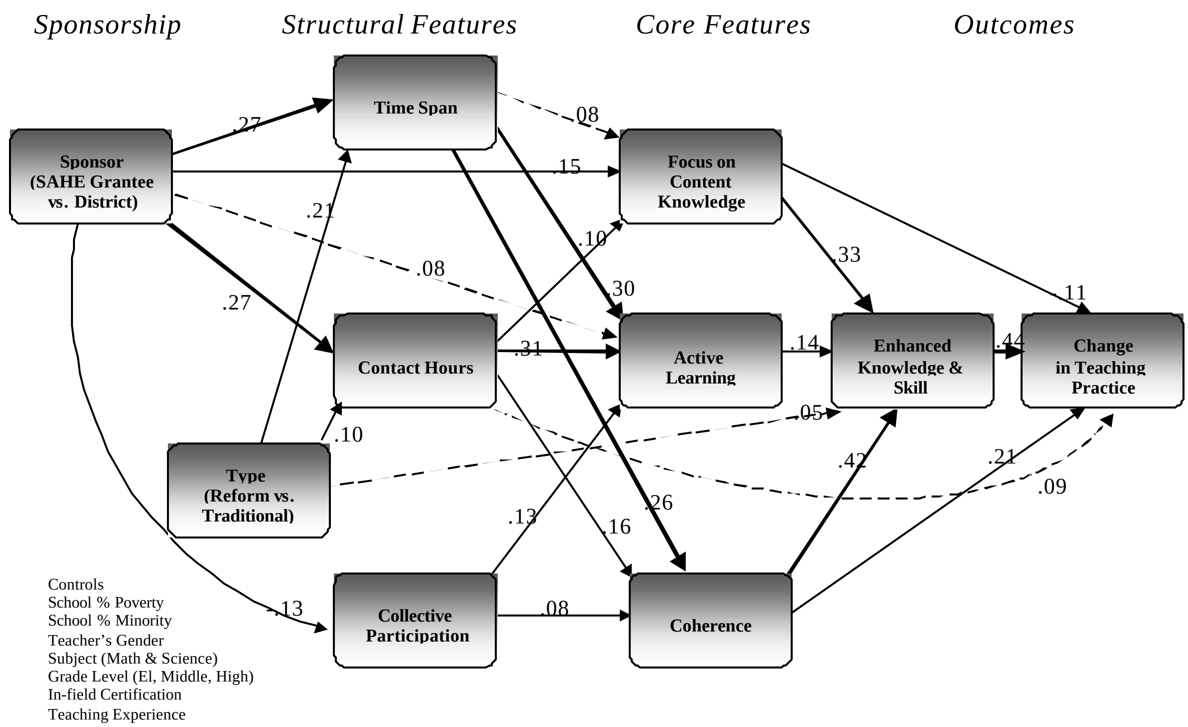
Our national data allowed us to examine how these features of professional development operate to affect teacher outcomes. We used a statistical technique, ordinary least squares regression (OLS), to estimate a formal causal model, which showed that the structural features of professional development activities influenced the core features of the activities and that the core features, in turn, influenced how successful the experience was in increasing teacher-reported growth in knowledge and skills and changes in teaching practice. For example, as Exhibit 2 shows, activities of longer duration tended to place more emphasis on deepening teachers' content knowledge, provide more opportunities for teachers to engage in active learning experiences, and provide activities that are more coherent. Similarly, activities with greater collective participation of teachers also tended to place more emphasis on content, provide more opportunities for active learning, and offered more coherent professional development than other activities. In turn, professional development that was content-focused and coherent and had active learning was more successful in improving teacher knowledge and eliciting changes in teachers' classroom practices.

The features of high-quality professional development identified in our national data, while consistent with ideas articulated in the Eisenhower legislation, deepen and extend those ideas by providing details about what makes professional development effective. For example, the Eisenhower legislation promotes professional development that is linked to other reform efforts in a coherent, systematic way. The results from our national data show the effectiveness of specific dimensions of coherence, such as discussing professional development experiences with colleagues, and participating in follow-up activities that build on previous activities.

Further, with our national data from district Eisenhower coordinators, we found significant differences between districts in the quality of professional development they provide. We found these differences both in the features of the activities provided—such as active learning, collective participation, and the span of time over which the activities extend—and in district management strategies, including alignment with standards and assessments, frequency of co-funded projects, and a commitment to continuous improvement. Generally, we found that larger districts are more likely to provide high-quality professional development than are smaller districts. (See Garet et al., 1999, for more details on these findings.)

EXHIBIT 2

The Relationship Between Features of Professional Development and Teacher Outcomes



The Purpose and Design of the Longitudinal Study of Teacher Change

Our confidence in these results is strong, given that the data are from a national probability sample. And although the data are based on teacher self-reports, we have confidence in the validity of the data because we did not ask teachers to judge the characteristics of the activities that influenced their effectiveness; instead we asked teachers to describe the characteristics of the activities they experienced, and we asked them whether the activities had an effect on their knowledge, skills, and classroom practice. Then, through data analysis techniques (e.g., ordinary least squares (OLS) regression), we identified characteristics that were associated with the effectiveness of the activities. Because teachers were not asked to judge the quality of the professional development in which they participated, the study minimizes self-report bias (e.g., Mullens & Gayler, 1999; Mullens, 1998). In addition, the substantial variation in the responses that teachers and district administrators provided to these behavioral items, as well as the consistency in teacher and district administrator responses, provides support for the validity of the data.

Although these data showed significant relationships between professional development and changes in teaching practice, the data are cross-sectional (i.e., they were collected at only one point in time). A stronger method of attributing changes in teaching practice to professional development experiences is to gather longitudinal data on teaching practice and experiences in professional development. The Longitudinal Study of Teacher Change was designed to build on these findings from our national, cross-sectional data. With longitudinal data, we can add to our knowledge drawn from the national data. The longitudinal data enable us to document teaching practice before and after a professional development activity and to examine the extent to which changes in teaching practice can be attributed to participation in the professional development activity.

The purpose of the Longitudinal Study of Teacher Change is to examine the effects of Eisenhower-assisted and other professional development on teaching practice in mathematics and science. We do not hypothesize, and so do not test, direct effects of professional development on student achievement; rather, we examine the direct effects of professional development on teachers' instruction. In the LSTC, we use detailed measures of teaching practice that we collected by surveying teachers at three points in time: the fall of 1997, the spring of 1998, and the spring of 1999. Although our study does not measure the effects of professional development on student achievement directly, the measures of teaching practice that we use have been associated with gains in student achievement. (We discuss the measures in more detail in Section III of this report.)

The Sample of Schools

We expected systematic differences in results by school level, so we chose one elementary school, one middle school, and one high school in each of the 10 districts we studied to allow the analysis of results by school level. Further, by design, the sample of 30 schools in the Longitudinal Study of Teacher Change is disproportionately high-poverty—17 of the sample schools, or 57 percent, are high-poverty; nationwide, 25 percent of schools are high-poverty (defined as 50 percent or more students eligible for free lunch).[9] This feature of the sample is useful in an evaluation of the Eisenhower program because the program targets teachers in high-poverty schools.

In addition, we sought schools in which teachers were likely to participate in Eisenhower-assisted activities over the 1997–98 school year, the year in which we conducted site visits to all 30 schools.[10] We selected states, districts, and schools in the sample that had adopted diverse approaches to professional development in addition to traditional workshops and conferences. If such professional development is more effective than traditional approaches, then the teachers' instruction in the sample schools might be better than that of the average teacher. A few of the 30 schools experienced achievement gains in 4th and 8th grade mathematics during the study period (1996–99), some experienced a decline in scores, and others remained at the same level. (See Appendix A for a list of the 4th and 8th grade achievement scores for 1996–99 for each of the 30 schools).[11]

[9] We used poverty data from the Common Core of Data (CCD).

[10] As part of our site visits to the 30 case study schools, we conducted one-time classroom observations of two teachers in each school—usually one mathematics teacher and one science teacher. In conjunction with the observations, we conducted a brief pre-observation interview and a somewhat longer post-observation interview with each of the 60 teachers we observed. The results of these observations are discussed in Garet et al., 1999.

[11] The achievement data were collected from existing data at the sites. Scores were not always available for 4th and/or 8th grade for every year. Where 4th and/or 8th grade scores are not available, we provide the scores for the grades closest to 4th and 8th grade.

In sum, the longitudinal sample was selected to maximize the opportunity to investigate important differences in approaches to professional development using Eisenhower funds. The sample is not nationally representative, but neither is it extremely unusual. It allows an exploratory, in-depth examination of the characteristics of professional development that foster teacher change. Our longitudinal data complement our earlier nationally representative data. The national data documented the frequency with which Eisenhower professional development has specific characteristics, and our longitudinal data allow us to look at the effectiveness of these specific characteristics over time.

The Sample of Teachers

We surveyed all the teachers who taught mathematics and science in each of the 30 schools in the sample (i.e., all of the elementary-school teachers, and the middle and high school teachers who teach mathematics and/or science classes). We focus on mathematics and science teachers because they are the primary participants in Eisenhower-assisted activities. In elementary schools, we randomly administered mathematics surveys to half the teachers and administered science surveys to the other half. The three waves of the survey provide data pertaining to the 1996–97, 1997–98, and 1998–99 school years.

Description of the sample. Four hundred thirty (430) teachers responded to the 1996–97 survey; 429 teachers responded to the 1997–98 survey; and 452 teachers responded to the 1998–99 survey. [12] (See Appendix B for a complete discussion of the response rates.) Some teachers who responded did not teach mathematics or science during the 1996–97, 1997—98, or 1998–99 school year, either because they were not employed as teachers in one or more of these years or because they taught other subjects, and thus they are not included in the analyses of classroom teaching. In addition, we excluded some teachers from particular analyses because they did not complete a minimum necessary set of items on the survey. For most analyses, we rely on the sample of 287 teachers who responded to all three waves of the survey. For some analyses (those focusing only on professional development experiences), we rely on a sample of 318 teachers who responded to at least the second and third waves. And for some analyses, we restrict the dataset to teachers who taught the same course in each of the three years of the study (n=207). The response rate for the first wave was 75 percent; for the second wave, it was 74 percent; and for the final wave in 1998, 75 percent. (See Appendix B for more details on sample sizes and response rates.)[13]

The sample is 74 percent female and 18 percent minority. Ninety-three percent of the sample are certified teachers. Twelve percent of mathematics teachers and 18 percent of science teachers in the sample are novice teachers, or teachers who have taught the surveyed subject for three or fewer years.[14] (See Appendix B for a more complete description of the sampling, response rates, design, and methodology.)

[12] The response rate of high school teachers was higher than those of elementary and middle school teachers, perhaps because principals and department chairs in high school were more involved in administering the survey.
[13] We compared responses from teachers who responded only to wave one, teachers who responded to waves two and three, and teachers who responded to all three waves and found no significant differences ins gender, teaching experience, certification, poverty, and all of our measures of teaching practice. The one significant difference we found was that teachers who responded to wave one only were overrepresented in high-poverty schools, compared with those who participated in all three waves.
[14] We asked teachers about personal background information, such as gender and years of experience, only in the baseline wave of the survey.

The data in this report are unique in that they provide detailed information on teaching practice and professional development over a three-year period for all teachers of mathematics and science in a school. These data enabled us to analyze relationships between teachers' professional development experiences and classroom practice, while controlling for prior differences in their classroom practice.

To set the context for examining the effects of professional development on instruction, in the next section we describe the professional development experienced by teachers in our longitudinal study.

SECTION II

Teacher Participation in Professional Development

The effectiveness of professional development depends to a large extent on its characteristics and quality (e.g., Cohen & Hill, 1998; Garet et al., 1999; Kennedy, 1998). Therefore, before we turn to assessing the effectiveness of professional development, we describe the professional development experiences of teachers in our longitudinal sample. We first describe our methods and measures and benchmark the quality of the professional development experienced by teachers in our longitudinal sample by comparing it with the professional development experienced by teachers in our national sample. We then describe our findings on the continuity of teachers' professional development from one year to the next.

Our examination of the means (averages) and standard deviations (a measure of variation) for several features of quality of the professional development activity showed a great deal of variation in professional development experiences among teachers in our longitudinal sample. This is consistent with our national data. Further, our longitudinal data enabled us to examine why the quality of teachers' professional development varied. We examined whether the variation was due to differences between teachers' subjects and school levels (i.e., mathematics vs. science, and elementary vs. middle vs. high school); differences between teachers in the same school; differences between schools; and differences in the average quality of professional development between one year and the next. We found the following:

- **Teachers in the same school appear to have quite different professional development experiences.**
- **There is a great deal of variation in the professional development that individual teachers experience from one year to the next.**
- **There is more consistency in the core features of teachers' professional development experiences—active learning, coherence, and content focus—than in the structural features—reform type, duration, and collective participation.**

Below we describe these results in more detail.

How We Describe Professional Development and Measure Its Quality

Identifying and describing a professional development activity is complex. Teachers experience many different types of professional development throughout their careers, both preservice and in-service, which serve a variety of purposes. Further, the quality of professional development can be described on many dimensions. In this section, we explain the instructions we gave teachers to identify the professional development activity on which to report, and we describe the dimensions on which we base our measures of quality.

Choosing a Professional Development Activity to Describe

In waves two and three of the survey (pertaining to the 1997–98 and 1998–99 school years), we asked teachers to describe the professional development activities in which they had participated.

For example, in the third wave, we asked teachers who received the mathematics form of the survey to "report all mathematics-related professional development you participated in over the previous year, including the summer of 1998 and the 1998–99 school year." We then asked them to choose one of the activities to describe in the following manner:

> If one of the organized professional development experiences you participated in was *particularly helpful to the class you reported [on earlier*], please pick that activity. If not, pick any organized professional development activity. You may choose an activity that began before the summer of 1998, if you continued to participate in that activity during the summer of 1998 or the 1998–99 school year. In answering questions about the activity you have chosen, please include all components of the activity, even if they occurred at different times during the year. (For example, if you attended a summer institute with follow-up activities during the school year, include both the summer institute and the follow-up activities in your answers.)

Thus, the activity reported on in many cases was not a typical professional development activity, but one that teachers thought was the most helpful to them.[15]

Some teachers did not participate in any mathematics or science professional development in 1997–98 or 1998–99. The percentage of teachers who did not participate in any science professional development in 1997–98, for example, ranged from less than 7 percent for high school teachers to nearly 25 percent for elementary teachers (See Appendix C).[16]

Measuring Features of Quality

On the basis of findings from our national data, we use six key features to describe the quality of professional development—reform vs. traditional, duration (time span and contact hours), collective participation, active learning, coherence, and content focus. Below we describe how we measured and scaled each of these dimensions of professional development.[17]

Reform vs. traditional. On the Longitudinal Teacher Survey (LTS), we asked teachers to describe the type of activity on which they were reporting, using eight categories. We classified three types of activities as traditional in form: (1) within-district workshops or conferences, (2) courses for college credit, and (3) out-of-district workshops or conferences. We classified the remaining five types of activities as reform activities: (1) teacher study groups, (2) teacher collaboratives, networks, or committees, (3) mentoring, (4) internships, and (5) resource centers.[18]

Contact hours. We asked teachers the total number of contact hours that they spent in the professional development activity, including all components of the activity that were held during a one-year period. The measure is a continuous variable indicating the number of contact hours the teachers spent in the activity on which the teacher reported.

[15] We did not ask teachers the total number of professional development activities in which they participated during the year; we did, however, ask them the number of *types* of professional development activities in which they participated, as well as the number of hours spent on each type.
[16] We excluded from our analyses teachers who did not participate in professional development.
[17] These scales are identical to the scales used in the analysis of our national data.
[18] The survey included a final category, "other organized forms of professional development," and asked the teacher to describe the form. We reclassified all responses into one of the eight forms.

Time span. In addition to asking about the hours of professional development provided, we asked about the span of the activity, or the period of time in days, weeks, months, or years over which the activity was spread. The nine options were (1) less than a day, (2) one day, (3) two to four days, (4) a week, (5) a month, (6) 2 to 5 months, (7) 6 to 9 months, (8) 10 to 12 months, and (9) more than a year. The composite for time span was coded on a 9--point scale, where 1=less than a day and 9=more than a year.

Collective participation. We asked each teacher in our longitudinal sample to indicate whether the activity in which the teacher participated was designed for all teachers in a school or set of schools or all teachers in the teacher's department or grade level.[19] We combined and averaged responses to these two questions to create an index of the extent to which the activity provided opportunities for collective participation. We coded the scale as 0=not collective, 1=somewhat collective, and 2=collective.

Active learning. To measure active learning, our survey included four items to measure opportunities for observing and being observed teaching; five items that measured planning for classroom implementation; four questions that focused on reviewing student work; and five items that asked questions about presenting, leading, and writing.[20] Since simply summing the 18 types of active learning opportunities would give more weight to planning and presenting/writing than to observing and reviewing student work, in computing the index, we weighted each of the four items pertaining to observation and the four items pertaining to student work by 1.25. This produced an index from 0 (no opportunities were provided for active learning) to 20 (all types of active learning were provided).

Coherence. We measured three dimensions of coherence. First, we asked each teacher to report the extent to which the activity the teacher attended was consistent with the teacher's goals for professional development, was based explicitly on what the teacher had learned in earlier professional development experiences, and followed up with activities that built on what the teacher learned in the professional development activity. Second, we asked each teacher to indicate the extent to which the activity was aligned with state or district standards and curriculum frameworks and with state and district assessments. Third, we asked teachers whether they had discussed what they learned with other teachers in their school or department who did not attend the activity;

[19] Teachers were also given the following options: teachers as individuals; teachers as representatives of their departments, grade level, or schools; and other configurations. Teachers could check all that applied.

[20] The observation questions asked (1) whether the teacher received coaching or mentoring in the classroom as part of the Eisenhower-assisted activity; (2) whether the teacher's teaching was observed by the activity leader(s) and feedback was provided; (3) whether the teacher's teaching was observed by other participants and feedback was provided; and (4) whether the activity was evaluated in part on the basis of an observation of the teacher's classroom. The classroom implementation questions asked whether, as part of the activity in which the teacher participated, the teacher (1) practiced under simulated conditions, with feedback; (2) met formally with other activity participants to discuss classroom implementation; (3) communicated with the leader(s) of the activity concerning classroom implementation; (4) met informally with other participants to discuss classroom implementation; and (5) developed curricula or lesson plans that other participants or the activity leader reviewed. The questions pertaining to reviewing student work were (1) whether the teacher reviewed student work or scored assessments as part of the activity; (2) whether work completed by students in the teacher's classroom was reviewed by other activity participants or (3) the activity leader; and (4) whether student outcomes were examined as part of an evaluation of the activity. The questions for presenting, leading, and writing were whether, as part of the activity, the teacher (1) gave a lecture or presentation; (2) conducted a demonstration of a lesson, unit, or skill; (3) led a whole-group discussion; (4) led a small-group discussion; or (5) wrote a paper, report, or plan.

whether they had discussed or shared what they learned with administrators (e.g., the principal or the department chair); and whether they had communicated, outside of formal meetings held as part of the activity, with participants in the activity who teach in other schools. Because there are three items for the first and third of these dimensions, and only two items for the second dimension, we weighted the items for the second dimension by 1.5. This produced a scale from 0 (the activity did not include any of the types of coherence that we measured) to 9 (the activity provided all of the forms of coherence that we measured).

Content focus. To examine the content focus of teachers' activities, we asked each teacher to indicate the degree of emphasis the activity gave to deepening content knowledge in mathematics and science, the extent to which the activity emphasized how students learn specific content, and the extent to which it focused on methods of teaching specific content. To measure this, we used a 3-point scale, where 0=no emphasis, 1= minor emphasis, and 2=major emphasis.

Characteristics of Teachers' Professional Development in Our Longitudinal Study

Eisenhower Sponsorship

To set the context for examining the effects of professional development, we first describe the type and quality of professional development that teachers in our longitudinal sample experienced. We describe activities on the following dimensions: (1) the extent to which the activities were funded by the Eisenhower program, (2) the differences between activities funded by Eisenhower and those funded by other sources, and (3) the differences in the quality of activities experienced by our longitudinal sample versus our national sample.

Teachers generally do not know the funding source of the activities in which they participate. Therefore, to determine whether activities were funded by the Eisenhower program, we used a three-step process. First, using the case-study reports written for each of the 30 schools in the study, we identified the list of professional development activities supported by Eisenhower funds during the 1997–98 school year. Second, we asked the teacher to write the name of the activity the teacher chose to describe on the survey. Specifically, we asked teachers to "please describe the activity in one or two sentences. Include the name or title of the activity (for example, the name of the workshop or title of the course attended)." Third, we matched activities that teachers named with activities listed in the case studies. There was some ambiguity in the matching process, since the Eisenhower coordinators and teachers sometimes used different terminology and descriptions of the professional development activities. Further, while we were able to identify which activities were supported through the district component of the Eisenhower program, we could not determine which activities were supported through the SAHE component of the program.

Using this process, we determined that 21 to 28 percent of the professional development activities described by teachers in our longitudinal sample were funded by the district components of the Eisenhower Professional Development Program in 1997–98. The conservative estimate of 21 percent includes only those activities for which there were clear matches on both lists; 28 percent is a more liberal estimate and includes activities that were *probably* funded by Eisenhower, but for which making an exact match was not possible.

Comparison of Eisenhower-Funded and Non-Eisenhower-Funded Activities in the Longitudinal Sample

To examine whether Eisenhower funding made a difference in the quality of the activities in which our longitudinal sample of teachers participated, we examined whether there are significant differences in the quality of activities according to whether or not the activity was funded by the Eisenhower program. Specifically, we used a statistical technique called "hierarchical linear modeling" to analyze differences in the six dimensions of quality. In these comparisons, we controlled for the teachers' subject (i.e., mathematics or science), the school level (i.e., elementary, middle, or high school), and the specific school and district in which the teachers' worked.[21] We did this so that we could strengthen our ability to show that any differences we found in the quality of professional development activities— that is, in reform type, time span, contact hours, collective participation, active learning, coherence, and content focus—were related to whether or not they received Eisenhower support and was not due to these other factors.

In our longitudinal sample, there are no significant differences in quality between Eisenhower-funded and non-Eisenhower-funded activities. This indicates that for the activities on which teachers chose to report, those funded by Eisenhower are not of higher quality than those funded by other sources. (See Appendix C for more detailed data.)

Comparison of Professional Development in the Longitudinal Sample to the National Sample

One way to judge the quality of the professional development activities experienced by teachers in our longitudinal sample is to compare them with activities experienced by our national sample of teachers. In our national sample, SAHE-sponsored activities were of significantly higher quality than district-sponsored activities on most dimensions. The longitudinal sample focuses on district- and school-level professional development; it does not provide information about professional development offered through the SAHE component of the Eisenhower program. That is, the activities of teachers in our longitudinal sample were virtually all district-funded, although only 21 to 28 percent were funded by the Eisenhower program.

When we compared the activities of teachers in our longitudinal sample with district-sponsored Eisenhower-assisted activities in our national sample on the six dimensions of quality discussed above, we found that the quality of the activities that teachers in our longitudinal sample experienced is about the same as the quality of activities experienced by teachers in our national district sample. The only significant differences were that teachers in our national sample experienced activities with significantly more contact hours and more content focus than teachers in our longitudinal sample.[22] Exhibit 3 shows these comparisons.

Our national data indicated the following about district-supported Eisenhower activities: an average of only 23 percent of teachers participating in Eisenhower-assisted professional development were in reform types of professional development; the average time span of a professional development activity was less than a week; the average number of contact hours was 25 and the median was 15 hours; most activities did not have collective participation or a major emphasis on content; and most activities had limited coherence and a small number of active learning

[21] The interaction between subject and school level was insignificant, so it was removed from the analyses.
[22] These analyses control for school level and subject.

opportunities (see Garet et al.,1999 for more details). In short, nationwide, the typical professional development experience was not high-quality.

Although only 21 to 28 percent of the professional development activities described by teachers in our longitudinal sample were funded by the Eisenhower Professional Development Program in 1997, the findings are directly relevant to the Eisenhower program. That is, our longitudinal data demonstrate trends and effects that we would expect from district-sponsored Eisenhower-assisted activities, given the common characteristics of activities in our longitudinal sample and Eisenhower-assisted activities nationwide.

Teacher and School-Level Variation in the Quality of Professional Development

Exhibit 3 allows us to examine the differences between teachers in our longitudinal sample in their professional development experiences. Consistent with our national data, the exhibit shows large differences in teachers' professional development in our longitudinal sample. We see this by looking at the "standard deviations" of the features of quality (the numbers in parentheses are the standard deviations). For example, if we look at the row labeled "contact hours," Exhibit 3 indicates that in 1998–99, teachers in our longitudinal sample participated in the selected professional development activity for about 18 hours, on average. However, the exhibit also indicates a great deal of variation in contact hours. The standard deviation (SD) measures this variation around the mean. For contact hours, the standard deviation is 21.7 hours; if we add the SD to or subtract it from the mean number of contact hours, we see that most of the longitudinal sample of teachers experienced a professional development activity between 1 hour (the lowest response category) and 50 hours.[23] Thus, although teachers' professional development was about 18 hours on average, many teachers' professional development activity lasted only an hour or two, whereas for many others, the activity lasted close to 50 hours.

[23] Subtracting the SD of 21.7 from the mean of 18 will result in a negative number of contact hours, which obviously is not appropriate. The fact that the standard deviation is greater than the mean indicates that the distribution for contact hours (and other features of professional development) is highly skewed.

EXHIBIT 3

Comparison of Professional Development for 1997–98 and 1998–99 Longitudinal Sample vs. National Sample

		Eisenhower-Sponsored National Sample	Longitudinal Teacher Survey	
		District	1997–98	1998–99
Reform Type				
	Mean	22.8	15.9	18.7
	(SD)	(41.8)	(36.7)	(39.1)
	N[24]	767	182	182
Scale: Percent of reform activities, 0–100%				
Time Span				
	Mean	3.99	4.20	3.81
	(SD)	(2.2)	(2.3)	(2.3)
	N	782	186	186
Scale: 1 to 9 where 1=less than a day and 9=more than a year				
Contact Hours				
	Mean	25.50	21.23	18.20*
	(SD)	(35.0)	(24.1)	(21.7)
	N	783	194	194
Scale: Number of hours of professional development activity, from 1 to unlimited maximum				
Collective Participation				
	Mean	0.38	0.36	0.33
	(SD)	(0.6)	(0.6)	(0.5)
	N	783	191	191
Scale: 0–2, where 0=not collective, 1=somewhat collective, and 2=collective				
Active Learning				
	Mean	3.53	3.89	3.43
	(SD)	(3.5)	(4.0)	(3.3)
	N	783	176	176
Scale: 0–20, where 0=no opportunities for active learning, and 20=all types of active learning (i.e., 18 types with weighting)				
Coherence				
	Mean	4.79	5.66	5.33
	(SD)	(2.1)	(2.0)	(1.9)
	N	781	173	173
Scale: 0–9, where 0=activity provided no types of coherence and 9=activity provided all 9 of the forms of coherence that we measured				
Content Focus				
	Mean	1.36	1.09*	1.09*
	(SD)	(0.7)	(0.8)	(0.8)
	N	765	191	191
Scale: 0–2, where 0=no emphasis on content, 1=minor emphasis on content, and 2=major emphasis on content				

*significantly different from the national sample at $p \leq .05$

How to read this exhibit Twenty-two point eight (22.8) percent of teachers in Eisenhower-assisted district-sponsored professional development activities were in reform-type activities. Fifteen point nine (15.9) percent of teachers in our longitudinal sample were in reform-type activities in 1997–98, and 18.7 percent were in reform activities in 1998–99. The higher percentage of reform types for teachers in the national sample compared with that for teachers in our longitudinal study is not statistically significant.

Note: For more details on the scales for each variable, see previous section on **Measuring Features of Quality**.

[24] The number of teachers included in the longitudinal analyses is below 287 (the number of teachers who completed the survey for all three years) because we could include only teachers who responded to the 1997-98 and 1998-99 waves (which included the professional development questions) *and* who participated in mathematics or science professional development in both years. Many teachers did not participate in professional development in both years; for example, about one quarter of elementary teachers to whom we sent the science survey did not participate in any professional development in science in either year.

Similarly, Exhibit 3 indicates that in the1998–99 wave of the longitudinal survey, the time span for teachers' selected professional development activity was between two –to four days and one week, on average. However, many teachers' activities lasted less than one day, and many others lasted as long as two to five months. As shown in Exhibit 3, the response category "3" indicates that the professional development lasted from two –to four days, and the response category "4" indicates a time span of one week. The score of 3.81 falls between the response categories 3 and 4, showing that teachers on average reported the time span of their professional development as between two –to four days and one week. The standard deviation of 2.3 indicates that most teachers' activities last between less than one day (response category 1) and one day (response category 2) and two to five months (response category 6). Thus, similar to the number of contact hours, there is also a great deal of variability in the time span of teachers' professional development activities. In fact, Exhibit 3 shows that there is a great deal of variation in teachers' experiences with almost all of the features that describe the quality of their professional development activities. This is consistent with findings from our national data on Eisenhower activities.

What accounts for the variation in teachers' professional development experiences? Our longitudinal data enable us to answer many questions to explain this variation. For example, do teachers who teach different subjects or levels of school have different professional development experiences? Do teachers in some schools experience higher-quality professional development than teachers in other schools? Do teachers within the same school have professional development experiences that differ from one another?

Our longitudinal data also enable us to answer questions about variation in teachers' professional development experiences over time. For example, does the average quality of teachers' professional development differ across years? Does the quality of professional development differ across years for individual teachers? That is, did individual teachers who experienced professional development of unusually high or low quality in 1997–98 also experience the same quality of professional development in 1998–99?

With our longitudinal data, we used hierarchical linear modeling to estimate what proportion of the variation in the quality of teachers' professional development experiences can be attributed to (1) average differences between teachers who teach different subjects or levels of school, (2) average differences between schools, (3) average differences between teachers in the same school, and (4) average differences between one year and the next. (See Appendix C for an explanation of the analyses undertaken for this section of this report.) For each feature of high-quality professional development shown in the first column of Exhibit 4, we show the percent of the variation in teachers' professional development experiences that we can attribute to each of these differences.

The last column of Exhibit 4, labeled "unexplained year-to-year variation in professional development of individual teachers," shows the percent of the variation between years in individual teachers' professional development experiences that cannot be explained by any of the differences listed above. If teachers' experiences in professional development were consistent from one year to the next, all the variation in their experiences would be due to differences between teachers' subjects and school levels; differences between teachers in the same school; differences between schools; and differences in the average quality of professional development between one year and the next. None of the variation would be left unexplained.

EXHIBIT 4

Percent of Variation in the Quality of Professional Development (1997–98 and 1998–99)

Features of Professional Development	Percent of Variation in the Quality of Professional Development Explained by Differences Across				Unexplained Year-to-Year Variation in Professional Development of Individual Teachers
	Subject and School Levels	Schools	Teachers in the Same School	Years	
Structural Features					
Reform Type	0.0%	0.0%	7.1%	0.0%	93.0%
Span	0.8	0.0	19.9	0.4	79.0
Contact hours	2.8	2.6	15.6	0.2	79.0
Collective participation	0.0	13.8	10.3	0.0	75.9
Core Features					
Active learning	0.0	7.5	45.7	0.0	46.8
Coherence	3.1	2.9	46.5	0.3	47.3
Content focus	5.0	3.3	30.0	0.0	61.7

How to read this exhibit Only 0.8 percent of the variation in the time span of professional development activities can be attributed to differences between teachers' subjects and grade levels; 0.4 percent of the variation lies between years; none of the variation is between schools; 19.9 percent of the variation in time span is between teachers in the same school; and nearly 80 percent of the variation is an unexplained year-to-year variation.

The data reported in Exhibit 4 indicate that some of the variation in teachers' professional development experiences can be explained by differences in teachers' subjects, levels, and schools and in the average quality of professional development across years.[25] But, the data also show that a considerable percentage of the variation in each feature of professional development is left unexplained.

For example, if we examine the row labeled "time span," we see that about 21 percent of the variation between teachers in the time span of their professional development activities can be explained by three differences, combined. These are the differences between teachers' subjects and school levels (0.8 percent), or differences between teachers within the same school (19.9 percent), or differences in average professional development quality across years (0.4 percent). By far, the largest percent of the variation in time span that our analyses can explain can be attributed to differences across teachers in the same school. About 79 percent of the variation in time span of professional development activities cannot be explained by these differences, however.

By contrast, we can explain a lot more of the variation in the level of coherence of teachers' professional development experiences. More than half of the variation in the coherence of teachers' professional development experiences can be explained by the differences between teachers' subjects and school levels (3.1 percent), across years in average professional development quality (0.3

[25] See Appendix C for tables showing the specific effects of school level and subject. We found that compared with middle and high school teachers, elementary school teachers participate in professional development activities that have higher-quality features and are more effective in increasing teachers' self-reported knowledge and skills. We found no significant differences in the quality of professional development according to subject. In our initial analyses, we examined differences by poverty level and found that it had no effect.

percent), across schools (2.9 percent), and across teachers within the same school (46.5 percent). About 47.3 percent of the variation in the coherence of professional development is not explained by these differences. Exhibit 4 also shows the following

- **Teachers in the same school appear to have quite different professional development experiences.**

Exhibit 4 shows that of the variation in teachers' professional development experiences that our analyses can explain, a substantial amount is accounted for by the differences between teachers in the same school. Very little of the variation in the features of quality of professional development experienced by teachers in our longitudinal sample can be explained by the other differences we examined. This means that the differences in the quality of professional development experiences across teachers in the same school are greater than the differences between teachers who teach different subjects or grade levels, or the differences in the average quality of professional development across years, or the differences in professional development across schools.

These results suggest that schools generally do not have systematic approaches to planning and providing professional development for their teachers. If they did, we would expect that a larger proportion of the variation in the quality of professional development would be explained by differences across schools. A substantial amount of variation between schools in the quality of professional development would indicate that teachers within the same schools were experiencing professional development with similar qualities.

One exception to the lack of differences between schools in the quality of professional development is collective participation—the extent to which the professional development was designed for all teachers in a school or set of schools or all teachers in a department or grade level. For this feature of high-quality professional development, almost 15 percent of the variation lies between schools. This finding lends confidence to the validity of this measure, in that collective participation is inherently a school-level construct, intended to measure whether all teachers in a school, department, or grade level participate in professional development together. Therefore, we would expect it to have more between-schools variation than other measures.

We also found some significant differences between districts on several features of professional development. Our sample size of districts is small (n=10), but these findings are consistent with results from our national data, which showed the importance of district management and implementation strategies in fostering the provision of high-quality professional development.[26] (See Appendix C for data on district differences.)

[26] Results from our national data suggest why some districts may be more likely than others to offer opportunities for collective participation in professional development activities. We know from the analysis of our national data that districts that coordinate funding streams; use continuous improvement strategies, such as conducting needs assessments; and involve teachers in planning for professional development are more likely to have activities that have opportunities for collective participation. Thus, districts that practice these management strategies would be more likely to offer professional development with opportunities for collective participation. Similarly, districts that coordinate funding streams and use continuous improvement efforts to plan and develop their professional development activities might be more likely to offer opportunities for teachers' collective participation. (See Garet et al., 1999, for more details on these findings.)

Continuity of Participation in High-Quality Professional Development

Our analyses indicate that there is only modest continuity in teachers' professional development experiences from one year to the next.

- **Teachers experience professional development that varies in quality from one year to the next.**

- **There is more consistency in the core features of teachers' professional development experiences—active learning, coherence, and content focus—than the structural features—reform type, duration, and collective participation.**

Exhibit 4 shows that for most dimensions of quality, more than half of the variation in professional development that teachers experience is not explained by the factors that we examined—subject taught and school level, differences between schools, differences across teachers within a school, and average differences between years in the quality of professional development. The "unexplained" variation is the variation between years that is not accounted for by the other factors in our analyses. It indicates that individual teachers appear to have very different experiences in their professional development from one year to the next. This is especially true for duration (i.e., time span and contact hours) and collective participation, two of the three structural features of the professional development. There is somewhat less variation across years in the degree of active learning, coherence, and content focus, the three core features. (As discussed earlier, the core features of professional development refer to the substance of the activity, such as active learning, coherence, and content focus. These can be contrasted to the features of the structure of the activity, such as whether it is a reform type, what its duration is, and whether it has collective participation.)

These findings are mirrored in an examination of the association between features of quality from one year to the next. We measure this relationship by computing the correlation coefficient; the higher the correlation coefficient, the stronger the relationship between features of professional development from one year to the next. The correlation coefficient ranges from –1.00 to +1.00. Exhibit 5 shows that the amount of active learning, coherence, and content focus that teachers reported in 1997–98 tend to correlate with the amount reported in 1998–99 (r=.56, .52, and .38, respectively). The time span, contact hours, and collective participation of teachers' reported professional development, however, tend to have much lower correlations across the two years (r=.21, .21, and .25, respectively). The correlation between reform type for 1997–98 and for 1998–99 is not significant. Thus, Exhibits 4 and 5 both show that there is greater variation across years (and thus weaker correlations) in the structural features of professional development—time span, contact hours, and collective participation—than in the core features of the professional development—active learning, coherence, and content focus.

EXHIBIT 5

Correlations of Professional Development Features Between 1997–98 and 1998–99

Features of Professional Development	Correlation Between Years 1997–98 and 1998–99*	N
Structural Features		
Reform Type	.09	199
Time Span	.21*	188
Contact Hours	.21*	196
Collective Participation	.25*	193
Core Features		
Active Learning	.56***	176
Coherence	.52***	192
Content Focus	.38***	193

*p<=.05, **p<=.01, ***p<=.001

How to read this exhibit Participating in professional development with a long time span in 1997–98 is correlated .21 with participation in professional development with a long time span in 1998–99. A correlation of one (1) would indicate a perfect correlation (i.e., all teachers would report exactly the same time span in 1997–98 and 1998–99.)

Perhaps the higher correlation from year to year in core features rather than in structural features of professional development occurs because structural features may be more subject to influences beyond the control of teachers. For example, some of the structural features of activities may be more in the hands of districts or states that may have requirements about the number of contact hours that teachers need to maintain their certification. In contrast, teachers may have more discretion over whether to participate in the active learning opportunities available in an activity and whether to discuss the activity with other teachers (a measure of coherence), and teachers may be consistent in these choices over time. The relatively higher consistency in coherence over the two years lends support to our measure of coherence, since it is intended to assess, at least in part, the degree to which professional development experiences build on one another.

Our examination of the variation in the qualities of professional development over time shows that the correlations from year to year in the quality of professional development are modest. Although there is some consistency in the professional development in which teachers participate from one year to the next, the year-to-year correlations for the quality of professional development are lower than the correlations for other variables frequently encountered in education (for example, student test scores or grades). Also, as Section IV will describe, there is considerably more consistency in teachers' classroom practice from year –to year than in the quality of their professional development. For our survey, the teacher had to select a single professional development experience to describe each year. This selection process may have had some randomness to it (e.g., because teachers may have varying, unrelated criteria for choosing a useful activity on which to report), and this randomness could also lower the correlations from one year to the next. Nevertheless, the professional development correlations imply *some* tendency for teachers to have consistent professional development over time.[27]

Consistency in teachers' professional development could be good or bad. Consistency in professional development is desirable only to the extent that the professional development is of high quality. The great variation in teachers' professional development experiences that we report indicate that although some teachers receive consistently high-quality, effective professional development, others receive consistently low-quality, ineffective professional development.

Summary: Teacher Participation in Professional Development

Our data analysis provides a description of teachers' experiences in professional development activities over a two-year period—1997–98 and 1998–99. In each of these years, teachers reported on the characteristics of one professional development activity that they found "particularly helpful" to a class that they described in another section of our survey. The Eisenhower program supported between 21 and 28 percent of the activities of teachers in our longitudinal sample in 1997–98. These activities were similar to district-sponsored Eisenhower activities in our national sample, but were generally of lower quality than SAHE-sponsored activities in our national sample.

We found a great deal of variation in the quality of professional development activities reported by the teachers in our longitudinal sample. Much of this variation appears to be random, although there is a small tendency for teachers to participate in activities of similar quality from one year to the next. In short, although teachers sometimes experience high-quality professional development, there is a great amount of variation in teachers' experiences—individual teachers in the same school often have very different professional development experiences, and individual teachers often experience very different professional development experiences from year to year.

The high variation in professional development across teachers and across years is consistent with the belief that professional development is very much an individual experience—and that even an individual teacher's experience is to a large extent inconsistent over time. This suggests that there are limited programmatic elements in the school or across years that help to shape teachers' experiences in professional development.

In the next section, we examine how teachers' experiences in professional development affect their teaching practice and how the quality of the professional development affects teaching practice.

[27] Again, these analyses do not include teachers who experienced no professional development during the study period.

SECTION III

The Effects of Professional Development on Teaching Practice in Mathematics and Science

Our longitudinal data provide a unique opportunity to further examine the effects of professional development experiences on teaching practice. Whereas our nationally representative data allowed us to identify the features of professional development that were related to increases in teachers' self-reported knowledge and skills and to changes in teaching practice, the longitudinal data afford us a stronger method of attributing changes in teaching practice to professional development experiences. Our longitudinal data allow us to document teaching practice before and after a particular professional development activity and then to determine the extent to which changes in teaching practice could be attributed to participation in the activity.

In this section, we describe a number of different approaches that we used to analyze our longitudinal data regarding the effects of professional development on teaching practice. Our first approach was to analyze the impact on classroom practice of the six features of high-quality professional development that we identified in our national data. We found the following:

- **Reform types of professional development, and professional development characterized by active learning and coherence, significantly increase teachers' use of active, project-centered instruction in their classrooms.**

We then turned to a more fine-grained analysis. We examined whether professional development that focused on specific teaching strategies, or sets of such strategies, had an effect on teachers' use of these strategies in their classrooms. In particular, we were interested in professional development that focused on teaching strategies intended to increase students' higher-order learning. We examined teaching strategies in three areas: technology use, instructional methods, and approaches to assessing student work. We found the following:

- **Professional development focused on specific, higher-order teaching strategies increases teachers' use of those strategies in the classroom.**

Because we conducted multivariate analyses, this finding takes into account other factors that could affect teachers' use of higher-order teaching strategies. Thus, our findings regarding the effects of professional development on the use of higher-order teaching strategies are independent of teachers' prior use of these teaching strategies, as well as of the teachers' subject matter (mathematics or science) and school level (elementary, middle, or high school).

We then turned back to our six features of high-quality professional development to determine whether these features would strengthen the impact of professional development that focused on specific higher-order teaching strategies. We found the following:

- **Features of high quality—reform types, collective participation, active learning, and coherence—increase the impact of professional development activities that focus on specific higher-order teaching strategies.**

Thus, our findings from the longitudinal sample of teachers support the findings from our national data, which identified the importance of features of high-quality professional development for improving teachers' knowledge and skills and their classroom practice.

The Effects of High-Quality Professional Development on Teaching Practice

Our first approach to examining the effects of professional development activities was to look at the relationships between the six features of high-quality professional development and the characteristics of effective teaching practice. For these analyses, effective practice consisted of key aspects of content and pedagogy. Our analysis of content included an examination of the alignment of content coverage with national standards. We begin with a description of good classroom practice in mathematics and science, which is followed by a description of our measures of content and pedagogy. Finally, we present the results of our analysis of the relationship between the features of high-quality professional development and good teaching practice, using these measures.

What Is Good Teaching Practice in Mathematics and Science?

An understanding of good teaching practice in mathematics and science begins with a vision for the classroom. Although this vision is difficult to capture empirically, research has identified some common elements of content and pedagogy that have been shown to improve student learning. Overall, effective instruction can be characterized by *content* that is aligned with high standards and *pedagogy* focused on active learning.

Content

Content—which includes both the topics of instruction, such as fractions or decimals, and the teacher's expectations for student performance, such as memorizing or understanding concepts—matters for student learning. Student achievement improves when the content of instruction is consistent with national standards and assessments (Cohen & Hill, 1998; Gamoran, Porter, Smithson, & White, 1997). National standards for mathematics and science specify crucial content areas that effective instruction should address: covering core topics, such as life science, and developing students' topic understanding in sophisticated ways, such as making connections to real-world situations.

The National Council of Teachers of Mathematics (NCTM) developed standards for mathematics curricula (NCTM, 1989) and instruction (NCTM, 1991) that included topics to be covered and performance goals for students.[28] The key content areas (i.e., topics and performance goals for students) differ by school level (i.e., K–4, 5–8, 9–12), but generally focus on (1) numbers and operations, (2) patterns and functions, (3) algebra, (4) geometry and spatial sense, (5) measurement, and (6) statistics and probability. The mathematics standards identify the following

[28] A draft of the revised 1989 standards was released in 1998. Major changes include (1) reorganizing the grade-level breakdown from K–4, 5–8, and 9–12 to preK–2, 3–5, 6–8, and 9–12; (2) relating process standards—expectations for student performance—more closely to content standards; (3) adding the process standards of representation; and (4) emphasizing the development of content strands (e.g., algebra) over the grade levels (Romberg, 1998). New NCTM Standards were published in 2000, but these were not available at the time of our study.

standards for student performance that apply across all grades: (1) problem solving, (2) reasoning, (3) making connections, and (4) communicating mathematical ideas.

Reform in science education has emphasized real-world problems, investigations of natural phenomena, and linkages to other subjects rather than abstract knowledge (Raizen, 1998). In setting content standards for science, the National Research Council (NRC) identified certain content areas as central to teaching and learning science: (1) physical, life, earth, and space science, (2) science and technology, (3) science in personal and social perspectives, and (4) the history and nature of science.[29]

In addition, the NRC identified some concepts and student performance standards that cross content areas, such as systems, order, and organization; evolution and equilibrium; and understanding of and ability to conduct scientific inquiry. In setting standards for student performance, the NRC emphasized developing skills useful for scientific inquiry, such as asking questions, collecting data, and developing explanations. An underlying premise of these standards is to focus less on "student acquisition of information" and more on "student understanding and use of scientific knowledge, ideas, and inquiry processes" (NRC, 1996, p. 52). Thus, the performance goal of memorizing material is less central than the goals of understanding concepts or making connections.

Pedagogy

In addition to content that encompasses both topics to be covered and performance goals for students to be reached, effective teaching practice in mathematics and science involves certain approaches to pedagogy. Pedagogy—the activities used to convey content, typically including such dimensions as whole-class versus individual instruction or project- versus text-based instruction—also matters for student learning. National mathematics and science standards are based on research indicating that teachers' use of active learning methods for their students is especially effective. Students learn science best when they are active participants engaged in activities, rather than passive recipients of lecture-style instruction (Raizen, 1998). Active learning calls for students to be involved in creating their own learning experiences. Pedagogical approaches that support active mathematics and science learning include using inquiry-based instruction, in which the teacher facilitates rather than informs; actively engaging students in complex problems for which there are no simple solutions; and incorporating multiple disciplines in activities (NCTM, 1998; NRC, 1996; Raizen, 1998).

National standards in mathematics and science, consistent with research on effective instruction, indicate that both content—especially core topics and complex performance goals—and pedagogy—especially instruction that fosters active learning—are important to student learning. Clearly, content and pedagogy are intertwined. Although active learning is especially student-driven, it is still coordinated around content—effective teachers set instructional goals and monitor activities, intervening when appropriate. For example, previous research has found that teachers sometimes focus so much on changing the process of instruction that they neglect the topics of the lesson. Roitman (1998), for example, described a case in which an observed teacher was so focused on providing active learning activities that her lesson was topic-free. Thus, although our analyses

[29] The National Science Teachers Association (NSTA), the American Association for the Advancement of Science (AAAS), and NRC each developed standards documents (see AAAS, 1993; AAAS, 1989; NSTA, 1992). The three organizations agreed that NRC would be responsible for developing broad standards for science (Raizen, 1998).

examine teachers' performance goals for students separately from pedagogy, we acknowledge that their interaction is also a strong contributor to effective instruction.

Measuring Content and Pedagogy

The complex vision of effective classroom practice in mathematics and science, embodied by national standards, is not easy to capture empirically. To examine teachers' content coverage, we collected unusually detailed information on the content covered by our sample of teachers. To measure pedagogy, we asked a series of questions about teachers' pedagogical strategies, from which we developed subscales to indicate types of pedagogy. This section provides a brief overview of these items and methods.

Choosing the Class to Describe

In each wave of the Longitudinal Teacher Survey (LTS), we asked teachers to select a year-long mathematics or science course to describe. We asked them to choose, if possible, a course they had taught in 1996–97, were continuing to teach in 1997–98, and expected to teach in 1998–99. The survey asked detailed questions about the topics they covered and their goals for student performance during the entire school year. In the 1997–98 and 1998–99 waves, we also asked teachers detailed questions about their participation in professional development activities. Thus, we gathered three years of data on teaching practices and two years of data on participation in professional development activities.

Measuring the Content Taught

We characterize the content taught in terms of both the *topics* covered and the *performance goals* teachers hold for students. In the content section of the survey, we asked teachers to describe the content they taught in the class they chose to describe, using a two-dimensional matrix. (Different forms of the matrix were used for elementary, middle, and high school mathematics and science. See Exhibit 6 for a sample section from the elementary mathematics form of the survey.) The matrix was initially developed by Porter, Kirst, Osthoff, Smithson, and Schneider (1993) in a comprehensive study of mathematics reform, and it was revised for use in the Eisenhower evaluation. The matrix has been used in several other studies that have shown that the content, as measured by the matrix, predicts student achievement gains (e.g., Gamoran et al., 1997).[30]

[30] Porter et al. (1993) present comprehensive information on the reliability and validity of data collected by using the content matrix, as well as data collected by using teacher logs and classroom observations.

EXHIBIT 6

Excerpt from Content Coverage Section of the Elementary School Mathematics Teachers Survey

Elementary School Topics	Coverage	Your Performance Goals for Students					
Whole Numbers	**<none>**	**Memorize**	**Understand Concepts**	**Perform Procedures**	**Generate Hypotheses**	**Collect Analyze/ Interpret**	**Make Connections**
Addition	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Subtraction	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Multiplication	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Division	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Combinations of add, subtract, multiply, and divide	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Equations (including missing addend, factor, etc.)	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Fractions	**<none>**	**Memorize**	**Understand Concepts**	**Perform Procedures**	**Generate Hypotheses**	**Collect Analyze/ Interpret**	**Make Connections**
Identify equivalent fractions	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Add	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Subtract	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Multiply	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Divide	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Combinations of add, subtract, multiply, and divide	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3

Source: Longitudinal Teacher Survey Elementary School Mathematics, Fall 1997 (1996–97 school year).
How to read this exhibit: The rows represent topics (in bold typeface) and their corresponding subtopics. The teachers were asked to complete the grid by (1) indicating whether the topic had not been covered during the school year by circling "none"; (2) noting the level of coverage for each subtopic by circling 0, 1, 2, or 3 to indicate not covered to sustained coverage; and (3) marking the emphasis for each subtopic on each of the performance goals by circling 0, 1, 2, or 3 to indicate no emphasis to sustained emphasis.

The rows of the matrix contain a comprehensive list of the topics and subtopics that teachers might cover. Algebra, for example, is a topic in mathematics, and multistep equations is a subtopic under algebra. Astronomy is a topic within science, and the Earth's moon is a subtopic under astronomy. Each subject area (i.e., mathematics and science) and each school level (i.e., elementary, middle, and high school) has a unique set of topics and subtopics. The matrix for middle school mathematics, for example, has nine topics and 84 subtopics, while the matrix for high school science has 28 topics and 191 subtopics.

The columns of the matrix contain performance goals for students. Performance goals are teachers' expectations for what students should know and be able to do. There are six performance goals in the matrix: (1) memorize; (2) understand concepts; (3) perform procedures; (4) generate hypotheses; 95) collect, analyze, and interpret data; and (6) make connections. (See Appendix D for definitions of the performance goals.) For example, when a teacher emphasizes memorizing, the teacher may expect students to be able to produce definitions or terms, facts, and formulas from memory. When a teacher emphasizes using information to make connections, the teacher may expect students to be able to use and integrate concepts, apply ideas to real-world situations, build or revise theory, and make generalizations.

The report *A Nation at Risk* (NCEE, 1983) has, in part, prompted a movement toward a more balanced emphasis on understanding and memorizing (Roitman, 1998). National mathematics and science standards emphasize the performance goal of understanding. Ideally, teachers would have performance goals for students that are consistent with the performance goals advocated in the national standards. Of the six performance goals that teachers could identify on the content matrix (i.e., memorize, understand concepts, perform procedures, generate hypotheses, collect/analyze/interpret data, make connections), four are especially relevant for the abstract thinking involved in developing complex understanding: understanding concepts, generating hypotheses, collecting/analyzing/interpreting data, and making connections. Two are especially relevant for developing concrete skills and knowledge: memorizing and performing procedures. Ideally, teachers will balance their emphasis on the six performance goals.

The content grid measures the two dimensions that make up a *content area*—defined as the intersection of the two dimensions, topics and performance goals. For example, if teachers emphasize memorizing facts about the Earth's moon, the content area incorporates the subtopic (the Earth's moon) and the performance goal (memorizing). Both elements—topics and performance goals—are integral to understanding the content of a lesson or course. For example, the student learning that would be likely to take place if the content were memorizing facts about the Earth's moon (e.g., its gravity or its distance from the Earth in miles) is very different from the student learning that would occur if the content were understanding the Earth's moon (e.g., the forces working to keep satellites in orbit).

Each teacher was asked to follow several steps in describing his or her yearlong course, using the matrix. First, the teacher indicated the amount of time given to each subtopic, using a scale where 0=no coverage, 1=slight coverage (less than one class/lesson), 2=moderate coverage (one to two classes/lessons), and 3=sustained emphasis (more than two classes/lessons). Then, the teacher indicated the relative amount of emphasis given to each performance goal when teaching the subtopic, using the following scale: 0=no emphasis, 1=slight emphasis, 2=moderate emphasis, and 3=major emphasis. The full matrix of data provided by each teacher can be used to calculate the percentage of the teacher's total yearlong class time—based on 180 days—devoted to each topic and subtopic, each performance goal, and each content area (i.e., the intersection of a subtopic and performance goal).

Performance goals and alignment with national standards. For this report, we analyzed the content matrix in two (of a number of possible) ways. First, we examined the percent of emphasis each teacher gave to each performance goal during the year of instruction on which he or she reported.[31] Second, we analyzed the extent to which the teachers' relative emphasis on particular topics and performance goals was aligned with high standards.

To report on the consistency of the content taught with high standards, we needed to identify an appropriate measure of high standards. The National Council of Teachers of Mathematics (NCTM) standards and the National Research Council (NRC) standards set a framework for important mathematics and science concepts that should be taught in the classroom. However, these standards are at a level of generality that makes quantitative content analysis difficult. Therefore, we used the National Assessment of Educational Progress (NAEP) to make explicit the content focus of the national standards. The NAEP provides items that reflect the NCTM and NRC frameworks and permits content analyses of items to determine relative emphases for mathematics and science content.[32, 33]

To compare how well instruction meets high standards, we report an overall measure of the alignment between the content areas taught by teachers in our sample and the content areas emphasized by the NAEP. We used the fine-grained information collected from the content matrix described earlier, and we developed a unique strategy of measuring alignment, drawing on the full set of items administered as part of the NAEP. The index combines the extent to which the topics covered by the teachers in our sample match the topics assessed by the NAEP, the extent to which the performance goals our teachers emphasize match the NAEP, and the extent to which the content

[31] We did not analyze each teacher's emphasis on particular topics because the sample size of teachers in our longitudinal study was too small. To analyze topics, the sample of teachers must be divided into subgroups, for example, elementary mathematics and middle school science. The number of teachers in each of these subgroups was too small to produce reliable results.

[32] To develop a test that would be perceived as national, the National Center for Education Statistics (NCES) has modeled the NAEP on the professional associations' standards (Reese et al., 1997). For example, 30 percent of the science assessment involves hands-on performance exercises and 50 percent involves open-ended questions (NAGB, 1997); these also are areas of emphasis for the standards. The high standards set by the test are evident in the scores reported for the 1996 science assessment; only 3 percent of students tested at the advanced level and 21 to 29 percent tested at or above the proficient level (Raizen, 1998).

[33] As "the nation's report card," the NAEP represents an appropriate standard, although admittedly not the only possible one. Many benchmarks are available for comparing actual and desired content. Using state standards as the benchmark would be more desirable in many ways; we might expect teachers to be more aligned with state standards than with national standards. In the ongoing *Moving Standards to the Classroom Study*, sponsored by the U.S. Department of Education's Planning and Evaluation Service and conducted by the American Institutes for Research, we are using state standards as the benchmark to examine aligned instruction in six states.

But to make the alignment analysis possible on a large scale, we used the NAEP. Because the NAEP focuses on content and performance goals consistent with standards developed by national professional associations, and because the NAEP establishes high expectations for achievement, it is reasonable to use the items on the NAEP tests as a measure of high instructional standards. Although the sets of standards developed by professional associations are not completely consistent with one another, there is a substantial amount of agreement about what constitutes high instructional and content standards. Mathematics and science generally are tested in every other NAEP administration, or every four years. The data used for these analyses were the 1996 mathematics and science NAEP tests.

areas (intersection of topics and performance goals) covered by the teachers match the topics assessed by the NAEP.

To determine the relative amount of emphasis given by the NAEP to each subtopic, performance goal, and content area in our elementary, middle, and high school science and mathematics matrices, we reviewed the full set of NAEP mathematics and science items for the 1996 tests for 4th, 8th, and 12th grade.[34,35,36] For each teacher, the index takes on a value ranging from zero (no agreement at all between the content areas the teacher emphasizes and those emphasized by the NAEP) to 100 percent alignment (complete agreement between the content areas emphasized by the teacher and the NAEP).[37]

High alignment indicates that teachers emphasize topics and performance goals that are similar to the NAEP's emphasis. For example, teachers might focus on understanding concepts (a performance goal) for motion and forces (a topic) by asking students to explain, in everyday terms, the relationship between motion and force. If the NAEP also emphasizes understanding concepts for motion and forces, there would be high agreement between instruction and the NAEP on that content area. If there was a pattern of such agreement across content areas, then the index of alignment would be high. It is difficult to define an appropriate expectation for the percentage of alignment with the NAEP. We would not expect perfect or even near perfect alignment; 25 to 50 percent alignment seems to be a reasonable goal, given that the NAEP may not be closely aligned with other standards to which teachers may be more likely to respond, such as state or district standards. Although we included all of the NAEP items, the items are still only a sample of the test domain, whereas our descriptions of instruction described 100 percent of the instruction. Further, some performance goals for students—such as carrying out sustained work—cannot be adequately measured by a timed, paper-and-pencil test such as the NAEP.

Measuring Pedagogy

We conducted analyses on a series of six questions in the Longitudinal Teacher Survey about teachers' pedagogical strategies. (See Appendix D for a list of the questions). Using a statistical technique called "factor analysis," we developed a set of four scales to measure the extent to which teachers rely on the four pedagogical strategies. Each scale is formed of subscales, and each subscale

[34] The NAEP 4th-grade test was compared with instruction for teachers in elementary school (grades K–5). The NAEP 8th-grade test was compared with instruction for teachers in middle school (grades 6–8). The NAEP 12th-grade test was compared with instruction for teachers in high schools (grades 9–12).

[35] We did not expect high alignment between the NAEP and high school mathematics and science instruction, given that in high school, the mathematics and science curricula involve separate courses focused on specific subtopics, whereas the NAEP is a comprehensive measure.

[36] We asked two curriculum experts in mathematics and two experts in science to review each NAEP item and to determine the specific subtopics and performance goals that each item was designed to tap. (Garet et al., 1999, provides information on the reliability of the expert ratings of the NAEP items.) The expert raters placed items in up to three cells of the content grid. Using this information for the full set of NAEP items, we computed the relative emphasis given by NAEP to each subtopic, performance goal, and content area.

[37] The index of alignment is computed as 1 minus the sum, across content areas, of the absolute value of the difference between the teacher's and the NAEP emphasis in each content area, divided by 2, the result is multiplied by 100. The absolute value is required because the index is designed to capture cells for which the teachers give more emphasis than the NAEP, as well as those for which they give less emphasis.

is standardized to have a mean of 50 and a standard deviation of 10.[38] Two scales capture traditional strategies: didactic instruction and individual seatwork. Didactic, or teacher-led, instruction includes the following teacher activities: lecturing and having students take a passive role; de-emphasizing group work, reading, writing, and student presentations; and not using concrete models or interdisciplinary lessons. Individual seatwork includes the following student activities: working on homework in class; working on pencil-and-paper exercises; reciting or drilling orally; taking quizzes; and working individually or in pairs.

Two scales capture nontraditional pedagogical strategies: active, project-centered instruction and discussion-oriented instruction. Project-centered instruction includes students' working on independent, long-term projects; problems with no immediate solution; technical writing skills; hands-on activities; and a de-emphasis on paperwork and individual work. Discussion-oriented instruction focuses on discussion and debate. (See Appendix D for internal consistency estimates of the scales and more information on how these scales were defined and constructed.)

The pedagogical approaches we identified here are consistent with the dimensions of pedagogy identified in the literature. The national standards, as well as research on pedagogy, indicate that effective instruction calls for an increased emphasis on nontraditional approaches, without fully abandoning traditional approaches. Teachers' balanced emphasis on specific pedagogical approaches is one indicator of the consistency of their teaching with national standards.

The Quality of Teaching Practice

We did not measure the quality of teaching practice in terms of how well teachers implement specific practices. For example, we know how often teachers use student-centered discussion, but not how well they use it, in terms of the amount of time they wait for student responses, the enthusiasm with which they facilitate discussions, and the extent to which they encourage and challenge students in their responses. We consider teachers' pedagogical strategies and the content taught, as measured by the content matrix, measures of teaching quality. This is supported by previous evidence, mentioned earlier, which has shown that teachers' content as measured on the content matrix is related to student achievement gains (e.g., Gamoran et al., 1997; Porter et al. 1993).

The Effects of Features of High-Quality Professional Development on Alignment, Performance Goals, and Pedagogy

We used information from the content matrix and the pedagogical questions on the Longitudinal Teacher Survey to develop 11 global measures of teaching practice: an index of alignment with the NAEP, the six performance goals (memorize; understand concepts; perform procedures; generate hypotheses; collect, analyze and interpret data; and make connections)' and the four pedagogical strategies (teacher-centered instruction, individual seatwork, active instruction, and discussion-oriented instruction).[39] As Exhibit 7 shows, in 1996–97, the average alignment of instruction with the NAEP for teachers in our sample was 21 percent. This did not change

[38] Because each scale was formed by averaging several subscales, the standard deviation of each scale is less than 10. The higher the correlation between the subscales that make up each scale, the closer the standard deviation is to 10. The standard deviation decreases as the correlation decreases.

[39] In our initial exploratory analyses, we examined the effects of professional development on topics covered, as indicated by the content matrix, but found no effects. As we indicated earlier, sample sizes for these analyses were very small, since the sample of teachers must be divided into subsets by subject and level; as a result, the analyses did not produce reliable results.

significantly in 1997–98 or 1998–99. Further, in 1996–97, on average, teachers gave 11 percent of their instructional time to generating hypotheses; 13 percent to collecting, analyzing, and interpreting data; 15 percent to memorization; 17 percent to making connections; 20 percent to performing procedures; and 23 percent to understanding concepts. Overall, teachers gave practically the same emphasis to each performance goal in 1998–99. Similarly, teachers' average use of the four pedagogical strategies, standardized to an average of 50 in 1996–97, did not change significantly in 1998–99. Our confidence in the validity of these results is based on the validity studies of the content grid with classroom observations and teachers logs, discussed earlier (e.g., Porter et al., 1993), and on the fact that the results are consistent with other studies that document the challenges and barriers to fostering teacher change (e.g., Cohen, 1990) (See Section IV, below, on *Trends in Teaching Practice* for more details on changes in teaching practice over time.)

EXHIBIT 7

Average Characteristics of Teachers' Instruction for 1996–97, 1997–98, and 1998–99

	Means by Year		
Dimension of Teaching Practice	**1996–97**	**1997–98**	**1998–99**
Alignment of Content Coverage with the NAEP *Scale:* From 0 to 100% alignment with the NAEP	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Alignment Index	21% (9)	20% (9)	20% (8)
Performance Goals *Scale:* From 0 to 100% emphasis on the performance goal			
Memorize	15% (7)	16% (6)	15% (6)
Understand Concepts	23% (6)	22% (5)	22% (5)
Perform Procedures	20% (6)	20% (6)	20% (7)
Generate Hypotheses	11% (5)	12% (5)	12% (5)
Collect, Analyze, & Interpret	13% (5)	13% (4)	13% (5)
Make Connections	17% (5)	17% (4)	17% (4)
Pedagogy *Scale:* Standardized scale where mean=50 and standard deviation=10, in 1996–97			
Didactic Instruction	50.1 (5.6)	50.6 (5.8)	49.9 (6.2)
Individual Seatwork	49.5 (5.4)	49.4 (5.5)	49.1 (5.6)
Active, Project-centered Instruction	49.7 (5.7)	49.5 (5.8)	49.1 (6.4)
Discussion-oriented Instruction	50.0 (10.8)	49.6 (9.3)	49.4 (8.7)

How to read this exhibit The percent alignment of teachers' content coverage with the NAEP decreased from 21 percent alignment in 199697 to 20 percent alignment in 1997-98 and 1998-99. These changes are not statistically significant. The percent emphasis that teachers give to the performance goal of memorization increased from 15 percent in 1996-97 to 16 percent in 1997-98; it decreased to 15 percent in 1998-99. These changes are not statistically significant. Teachers' use of the pedagogical strategy of didactic instruction increased from 50.1 to 50.6 in 1997-98; and decreases to 49.9 in 1998-99. These changes are not statistically significant. Pedagogy is on a standardized scale where 50 is the mean level of use of each of the four strategies; this number is not a percent, but rather a general metric designed to provide a point from which to show increases or decreases in teachers' use of a particular pedagogical strategy. The mean was standardized in 1996-97, but could vary in 1997-98 and 1998-99.

Exhibit 7 indicates virtually no overall change in teaching practice across our sample of teachers as a whole. However, this does not mean that individual teachers did not change their classroom practice. Individual teachers could change their practice substantially, but in ways that cancel each other out on our measures. For example, with regard to the performance goal, "performing procedures," some teachers might have changed their practice in the direction of increasing emphasis on this performance goal. Other teachers may in fact have decreased their emphasis on performing procedures, deciding to concentrate instead on understanding concepts. Thus, the overall appearance of stability in teaching practice from one year to the next could mask individual change over time.

As we previously described, we asked about teaching practice in 1996–97, then asked questions about teachers' experiences in a professional development activity in 1997–98, then asked about teaching practice again in 1998–99. With these data, we were able to predict teaching practice in 1998–99 on the basis of the features of teachers' professional development experiences in 1997–98, controlling for teaching practice in 1996–97.[40,41] We used ordinary least squares (OLS) regression to examine the effects of the six qualities of professional development described in Section II, as measured in 1997–98, on the 11 broad dimensions of teaching practice (i.e., alignment, the six performance goals, and the four types of pedagogical strategies), as measured in 1998–99, controlling for teaching practice in 1996–97. These analyses are shown in Exhibit 8. We found the following:

- **Reform types of professional development, and professional development characterized by active learning and coherence, significantly increased teachers' use of active, project-centered instruction in their classrooms.**

However, as is clear from Exhibit 8, there were very few effects of the quality of professional development on the 11 dimensions of teaching practice. We found seven significant effects in this analysis. From the number of regression analyses that we performed, we would expect some significant relationships just by chance (i.e., about four, based on p<.05).[42]

---

[40] We base our measures of teaching practice on data from our Longitudinal Teacher Survey. We did, however, conduct classroom observations in the second year of the study to validate the survey instrument. In 1997—98, we conducted classroom observations of 60 teachers in our sample. An observer completed the content matrix for one class session, and the teacher completed the content matrix for the same session, and we compared results. We also compared results for observer's versus teacher's perceptions of pedagogy. On many measures we found consistent ratings, and for some measures we found inconsistency between observer and teacher self-reports. Prior research, however, indicates that detailed, nonevaluative teacher surveys (such as the ones we used for the Eisenhower evaluation) have a high level of consistency with classroom observation results (e.g., Mullens & Gayler, 1999; Porter et al., 1993). Further, classroom observations would allow us to measure classroom practice only at several points in time, whereas the content matrix measures classroom practice for the entire school year. That is, using classroom observations would not allow us to measure the content of instruction for the entire year (since we could not observe classrooms every day). But with the information from the content matrix, we are able to measure teachers' instructional content for the entire year (since in the content matrix, teachers report on their instruction for the entire school year).

[41] We did not use professional development in 1998–99 to predict teaching practice in that year to reduce the "common method" bias that arises from asking about professional development and teaching practice on the same survey.

[42] We looked at 77 relationships (i.e., 7 features of quality and 11 teaching practices). We would expect 1/20 (i.e., 4) to be significant by chance.

EXHIBIT 8

Effects of the Quality of Professional Development on Alignment, Performance Goals, and Pedagogy

Independent Variable: Features of Quality of Professional Development	Dependent Variable										
	Alignment	Performance Goal Emphasis						Pedagogical Strategies			
	Alignment with the NAEP	Memorize	Understand Concepts	Perform Procedures	Generate Questions/ Hypotheses	Collect, Analyze, and Interpret Data	Use Information to Make Connections	Didactic Instruction	Individual Seatwork	Active, Project-centered Instruction	Discussion-oriented Instruction
Reform Type	*	NS	NS	NS	NS	NS	NS	NS	NS	*	NS
Contact Hours	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Span	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*
Collective Participation	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Active Learning	NS	NS	*	NS	NS	NS	NS	NS	NS	*	NS
Coherence	NS	NS	NS	*	NS	NS	NS	NS	NS	*	NS
Content Focus	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS

How to read this exhibit: Participating in professional development that is a reform type does not significantly affect teachers' emphasis on the six performance goals, but it does increase teachers' alignment with the NAEP and the use of active project-centered instruction. An "*" indicates that the coefficient is statistically significant at p<.05; "NS" indicates that the effects are not statistically significant at the p<.05 level.

It seems plausible that the effects of certain features of high-quality professional development on active, project-centered instruction can be explained by the close link between certain features of professional development (e.g., reform-type activities and active learning opportunities) and project-centered instruction. For example, reform types of professional development often focus on active learning for teachers and therefore would help promote active project-centered instruction for students. Similarly, one of the components of project-centered instruction might be considered a proxy for active project-centered instruction. For example, one component on the project-centered pedagogy scale is work on an independent, long-term project, which is a type of reform activity that teachers might experience in their professional development. Finally, participating in professional development with active learning opportunities is the feature of professional development most strongly correlated with teachers' use of active project-centered instruction. Their relationship indicates that teachers who experience active learning in their professional development activities are most likely to use active approaches to instruction.

We may have found only a small number of effects in our analysis because our measures of alignment, performance goals, and pedagogy are too broad to detect smaller, yet important changes in instructional practice. It may well be true that a teacher made a substantial change in one particular area of instruction, but if it was only one practice (e.g., use of computers to write reports), it might not be evident in broad measures of alignment, performance goals, or pedagogy. If we had found many significant relationships, this would suggest quite strong effects. As it is, we did find a few significant relationships that indicate the potential for professional development to affect general teaching practices. [43]

The Effects of Professional Development on Higher-Order Use of Technology, Instructional Practices, and Student Assessments

Next, we examined more closely some specific teaching strategies that were the focus of teachers' professional development to determine whether they could be linked to specific changes in teaching practice. We focused on three areas—technology use, instructional methods, and student assessments. We chose these because certain practices in each of these areas have been linked to higher-order learning by researchers or school reformers and because we had exactly parallel measures of both professional development and teaching practice in these three areas. After describing our measures, we turn to the results of these analyses regarding the impact of professional development.

Measures of Higher-Order Teaching Practice and Professional Development

On the longitudinal teacher survey, we asked teachers for detailed information about their use of specific teaching strategies in their classrooms. In a separate section of the survey, we asked exactly parallel questions about whether teachers' professional development activity focused on these specific strategies. From these questions, we identified three areas of teaching practice and

[43] The small number of effects might be due in part to differences in measurement in our national and longitudinal surveys. In our national sample, we asked teachers to report whether the professional development activity in which they participated enhanced their knowledge and skills in one of six areas; in our longitudinal sample, we asked teachers many more-detailed questions about changes in their instruction. If a teacher changed on one or two instructional practices, this would be recorded as one or two of the six options in the national survey, while it would be tapped by one or two out of more than 50 practices on our longitudinal survey. Therefore, when we combine the items from the longitudinal survey into scales (e.g., alignment, the six performance goals, and the four pedagogy scales), we would not expect to see much of an effect.

professional development for which we analyzed effects: (1) the use of technology for higher-order learning, (2) the use of instructional methods for higher-order learning, and (3) the use of assessment strategies for higher-order student learning. Below, we describe our measures for each of these areas.

Use of technology for higher-order learning. Much recent literature focuses on the potential benefits of certain uses of technology on students' learning (Birman et al., 1997; Means, 1994; Means et al., 1993). Researchers have examined various uses of technology to support multidisciplinary tasks; to help students learn critical thinking; to provide opportunities for authentic learning experiences, such as collecting and analyzing real-world data; and to provide opportunities for access to experts, resources, and information beyond the classroom (Cognition and Technology Group at Vanderbilt University, 1994; Sivin-Kachala & Bialo, 1996; Means & Olsen, 1995). These types of technology use for higher-order learning have been viewed as a key feature of enabling students to achieve at high levels of performance.

To measure the extent to which professional development activities focused on the uses of technology that are linked to higher-order learning, we asked teachers whether the professional development activity in which they participated focused on improving their capacity to use (1) calculators or computers to develop models or simulations; (2) calculators or computers for data collection and analysis; (3) computers to write reports; and (4) computers to access the Internet. Teachers responded yes or no. In Years 1 and 3, we asked teachers how often they used these strategies as part of their mathematics or science instruction. The response scale was 0=almost never, 1=some lessons, 2=most lessons, and 3=every lesson.

Exhibit 9 shows that in 1997–98, about a fourth of teachers participated in professional development that focused on each of the following: using calculators or computers to develop models, using calculators or computers for data collection and analysis, and using computers to access the Internet. Thirteen percent of teachers experienced professional development focused on using computers to write reports.

EXHIBIT 9

Percent of Teachers Whose Professional Development Focused on Specific Teaching Strategies in 1997–98

Teaching Strategy Used in the Professional Devleopment Activity	Percent of Teachers Whose Professional Development Focused on the Strategy
Use of technology for higher-order learning	
calculators or computers to develop models	28%
calculators or computers for data collection and analysis	28
computers to write reports	13
computers to access the Internet	25
Use of instructional methods for higher-order learning	
work on independent, long-term projects	32
work on problems for which there is no obvious solution	41
develop technical writing skills	34
work on interdisciplinary lessons	39
debate ideas or otherwise explain their reasoning	44
Use of student assessments for higher-order learning	
essay tests	10
performance tasks	58
systematic observation of students	41
math/science reports	18
math/science projects	29
portfolios	32

How to read this exhibit: In 1997–98, on average, 28 percent of the teachers in our longitudinal sample partcipated in professional development that focused on using calculators or computers to develop models.

In terms of teaching practice, Exhibit 10 shows that in both 1996–97 and 1998–99, teachers used higher-order technology strategies infrequently. On average, teachers reported that the frequency of their use of these strategies was between “almost never” and “some lessons.” In 1996—97, teachers’ mean use of higher-order technology strategies was between .35 and .57, depending on the strategy, where 0=almost never, 1=some lessons, 2=most lessons, and 3=every lesson. For 1998—99, the mean use was somewhat higher—between .42 and .65. There is considerable variation on these measures, however; standard deviations range from .59 to .73. This indicates that many teachers almost never use these technology strategies, while some teachers use these strategies in “most lessons.”

EXHIBIT 10

Teachers' Use of Specific Teaching Strategies for Higher-Order Learning

Teaching Strategy Used in the Classroom	1996–97 Mean	1996–97 SD	1998–99 Mean	1998–99 SD
Use of technology for higher-order learning[1] *Scale:* 0=almost never; 1=some lessons; 2=most lessons; 3=every lesson				
calculators or computers to develop models	.49	(.70)	.57	(.73)
calculators or computers for data collection and analysis	.57	(.65)	.61	(.73)
computers to write reports	.50	(.62)	.65	(.73)
computers to access the Internet	.35	(.59)	.42	(.66)
Use of instructional methods for higher-order learning[2] *Scale:* 0=almost never; 1=some lessons; 2=most lessons; 3=every lesson				
work on independent, long-term projects	.78	(.63)	.82	(.62)
work on problems for which there is no obvious solution	.88	(.61)	.91	(.58)
develop technical writing skills	1.20	(.79)	1.19	(.70)
work on interdisciplinary lessons	.98	(.78)	1.04	(.78)
debate ideas or otherwise explain their reasoning	1.48	(.75)	1.38	(.80)
Use of student assessments for higher-order learning[3] *Scale:* 0=not used; 1=minor importance; 2=moderate importance; 3=very important				
essay tests	1.06	(.96)	.97	(.94)
performance tasks	2.04	(.86)	1.93	(.86)
systematic observation of students	1.98	(.93)	1.97	(.91)
math/science reports	1.05	(.96)	1.09	(.89)
math/science projects	1.34	(1.04)	1.32	(.98)
portfolios	1.02	(1.07)	.88	(1.02)

How to read this exhibit: In 1996—97 , on average, teachers in our longitudinal sample had students in their class use calculators or computers to develop models between "almost never" and "some lessons" (.49); in 1998—99, teachers increased their use of this strategy (.57).

Note: The wording of the survey items was as follows:

1. About how often did students use the following as part of math/science instruction?
2. How often did you have students (during math/science instruction)?
3. How important were the following assessment strategies in determining students' grades in this math/science class?

Instructional methods for higher-order learning. National standards in mathematics and science and research literature have identified a number of instructional methods intended to increase students' higher-order learning. Research has shown that students learn best when instruction includes opportunities for them to engage in active learning (e.g., Raizen, 1998). Both the National Research Council and the National Council of Teachers of Mathematics to support instruction that includes inquiry-based instruction and active learning opportunities (NCTM, 1998; NRC, 1996).

To measure the extent to which professional development activities emphasized higher-order instructional methods, we asked teachers whether the professional development activity in which they participated focused on developing their capacity to use any of the following six instructional methods with students: (1) work on independent, long-term (at least one week) projects; (2) work on problems for which there is no immediately obvious method or solution; (3) develop technical or mathematical writing skills; (4) use equations, graphs, tables, and text together; (5) work on interdisciplinary lessons (e.g., writing journals in class); and (6) debate ideas or otherwise explain their reasoning. Teachers responded yes or no to these questions. In 1996–97 and 1998–99, we asked teachers how often they used these methods as part of mathematics/science instruction. The response scale was 0=almost never, 1=some lessons, 2=most lessons, and 3=every lesson.

In 1997–98, between about one-third and two-fifths of teachers participated in professional development that focused on an instructional method intended to foster higher-order learning, depending on the type of instructional method (e.g., 32 percent of teachers participated in professional development that focused on working on independent, long-term projects, and 44 percent of teachers participated in professional development that focused on debating ideas). This did not change much in 1998–99. These results are shown in Exhibit 9.

On average, teachers reported using these higher-order instructional methods in "some lessons," which is indicated by a response of 1 on our response scale. The mean use of higher-order instructional methods in teaching practice was between .78 and 1.48 in 1996–97, depending on the method, and between .82 and 1.38 in 1998–99. There is little change in these measures from one year to the next. The standard deviations range from .61 to .79, however, indicating that there is moderate variation in teachers' use of these methods. This means that many teachers "almost never" use these instructional methods, while some others use them in "most lessons." Exhibit 10 displays these means and standard deviations.

Use of assessment strategies for higher-order learning. Much recent literature has advocated the use of different forms of assessment to measure students' higher-order learning. The usual paper-and-pencil tests are viewed as perhaps adequate for assessing basic skills, but not for the higher-order skills that researchers and reformers would like students to master. Alternative assessments such as essay tests, portfolio assessments, and project-based assessments are more appropriate than paper-and-pencil tests for measuring students' higher-order learning (e.g., Koretz, Stecher, Klein, & McCaffrey, 1994; Mitchell, 1996).

To measure the extent to which professional development emphasized student assessment methods that are associated with higher-order learning, we asked teachers whether the professional development activity focused on developing their capacity to use any of the following six forms of student assessments in their classroom teaching: (1) essay tests; (2) performance tasks or events; (3) systematic observation of students; (4) math/science reports; (5) math/science projects; and (6) portfolios. Teachers responded yes or no to these questions. In Years 1 and 3, we asked teachers how important these assessment strategies were in determining students' grades in the mathematics/science course on which they were reporting. The response scale was0=not used, 1=minor importance, 2=moderate importance, and 3=very important.

As Exhibit 9 shows, the percentage of teachers in professional development with a focus on using higher-order student assessments varies considerably, from 10 percent to 58 percent, depending on the type of student assessment (e.g., 10 percent of teachers participate in professional development that focused on using essay tests, and 58 percent are in professional development that focused on performance tasks.) Participation rates generally remain constant from 1997–98 to 1998–99.

Exhibit 10 shows that on average, teachers report placing "minor importance" on using some approaches to assessing the performance of their students, while they place greater emphasis on other assessment methods. The mean response category of about 1 for essay tests (1.06), math/science reports (1.05), and portfolios (1.02) indicate that on average teachers relegate minor importance to these methods. Teachers appear to place greater emphasis on other approaches to assessing student performance. The mean response category of about 2 for teachers' use of performance tasks (2.04) and systematic observations of students (1.98) indicates that they place "moderate importance" on these methods. The changes in the uses of assessment approaches across the two years when we collected data are not statistically significant. The use of all these approaches to student assessments varies considerably; standard deviations range from 0.86 to 1.07. This indicates that for almost any method for assessing student performance, many teachers place only "minor importance" on the method, while many other teachers view the method as "very important."

The Effects of Focused Professional Development on the Use of Specific Teaching Strategies in the Classroom

We examined the effects on teachers' use of specific teaching strategies of professional development that focused on these strategies. For each group of higher-order strategies discussed above—in the areas of technology use, instructional methods, and assessment of students—we examined the relationship between focusing on a strategy in professional development and using that strategy in the classroom. We also analyzed the effect of professional development that focused on closely related teaching strategies on teachers' use of the strategies in the classroom. For example, we were interested in whether teachers whose professional development in 1997–98 focused on using computers and calculators to develop models used computers or calculators to develop models in their classrooms in 1998–1999. We also were interested in finding out whether teachers whose professional development activity focused on other related higher-order uses of technology—such as using computers or calculators for data collection and analyses, for writing reports, or for accessing the Internet—were also more likely to use computers or calculators to develop models.

We used a multivariate statistical technique called "hierarchical linear modeling" to determine the effects of professional development that in 1997–98 focused on specific strategies on the use of these specific teaching strategies in 1998–99. This technique allowed us to determine the effect of professional development on the use of specific teaching practices while controlling for other factors that could affect teaching practice. These factors were the use of these teaching strategies prior to the professional development (in 1996–97), teachers' subject (mathematics or science), and school level (elementary, middle, or high school). (See Appendix D for a detailed discussion of the statistical approach and a complete presentation of the results.)

We found the following:

- **Professional development focused on specific, higher-order teaching strategies increases teachers' use of those strategies in the classroom.**

Exhibits 11, 12, and 13 illustrate that in each of the three areas examined in our analysis—technology use, instructional methods, and student assessment—professional development that focuses on a particular teaching strategy increases teachers' use of that strategy. As described above, these effects control for the use of these teaching strategies in 1996–97, teachers' subject, and school level. In addition to expecting effects of strategy-focused professional development on teachers' use of a particular strategy, we hypothesized that professional development that focuses on other specific strategies within the same area of professional development also would increase teachers' use of a specific strategy within that area. In other words, we hypothesized a "spillover" in the effects of professional development on classroom uses of specific teaching strategies. Our results indicate that the spillover is in the expected direction, but results are not significant.

As discussed above, one of the strategies that we examined in the area of higher-order technology was the use of calculators or computers to develop models. We use this one strategy as an example to discuss the results, but this strategy is representative of the other strategies in the area of technology. That is, the findings apply to each of the strategies in technology.

Exhibit 11 illustrates the effects of two types of professional development—professional development that did not focus on this strategy at all and professional development that focused on this strategy. The analysis controls for subject area and school level and for teachers' use of the strategy before participating in the professional development. Reading from left to right, the first bar of Exhibit 11 (.54) indicates that without professional development that focused on using calculators or computers to develop models in 1997–98, on average, teachers reported using this strategy between "almost never" (response category of 0) and "in some lessons" (response category of 1). If the teachers participated in professional development that focused on this strategy, their use of the strategy in the classroom increased to .84, meaning that more teachers reported using the strategy in "some lessons." Participating in professional development that focused on other specific technology strategies (e.g., use of computers to write reports or use of the Internet) was even further in the direction of increased use of calculators or computers to develop models even further, but results were not significant. (See Appendix D for these results.)

EXHIBIT 11

Effects of Professional Development on the Use of Calculators or Computers to Develop Models

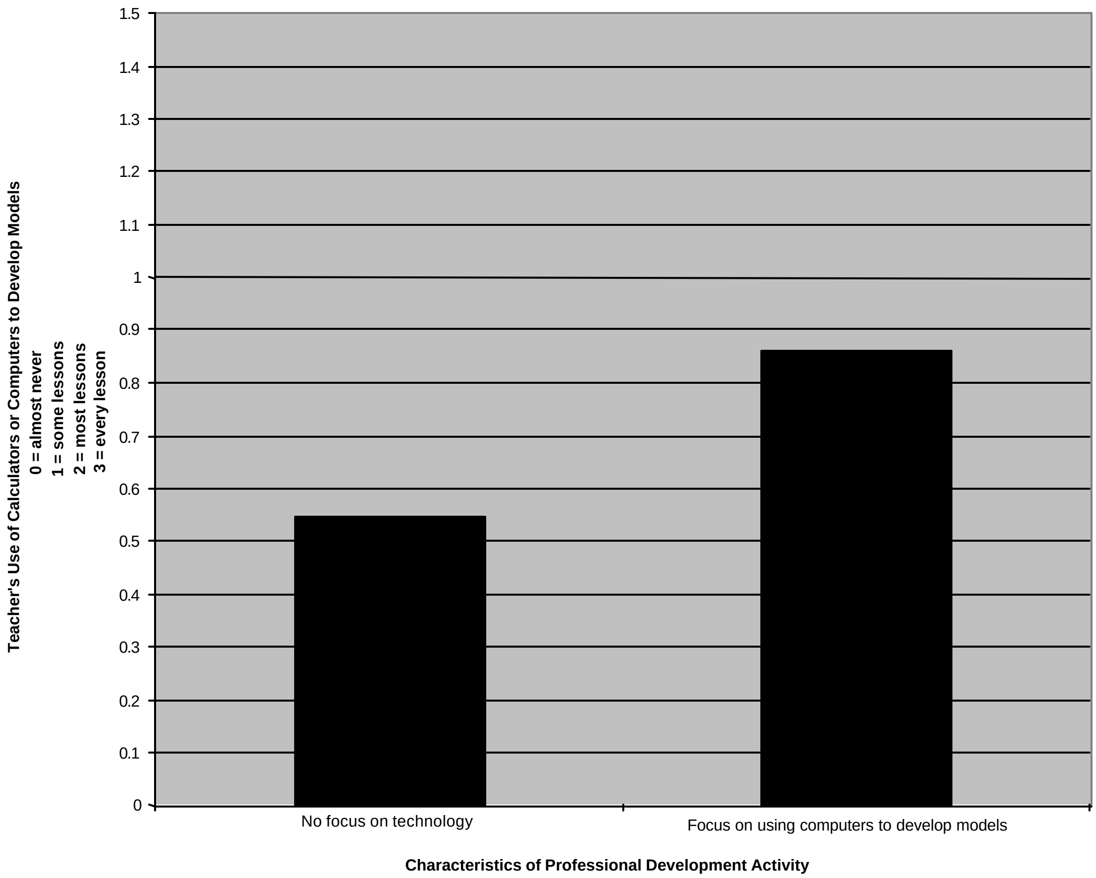


Exhibit 12 illustrates the same pattern, using an example from the area of higher-order instructional methods. Reading from left to right, the first bar indicates that, on average, teachers use problems with no obvious solutions in "some lessons," without any professional development regarding this teaching strategy. (The value of this bar is about 1.1, close to the response value of 1, which equals "some lessons.") The second bar indicates that participating in a professional development activity focusing on the use of problems with no obvious solutions boosts the use of this strategy in the classroom to more than 1.3—closer to the response category of 2, "most lessons." Although these changes may seem small, and it is difficult to assign a value to increasing teachers' use of a strategy from "some lessons" closer to use during "most lessons," our findings show that teachers whose professional development focuses on this strategy are likely to increase its use in their classrooms.

EXHIBIT 12

Effects of Professional Development on the Use of Problems with No Obvious Solution

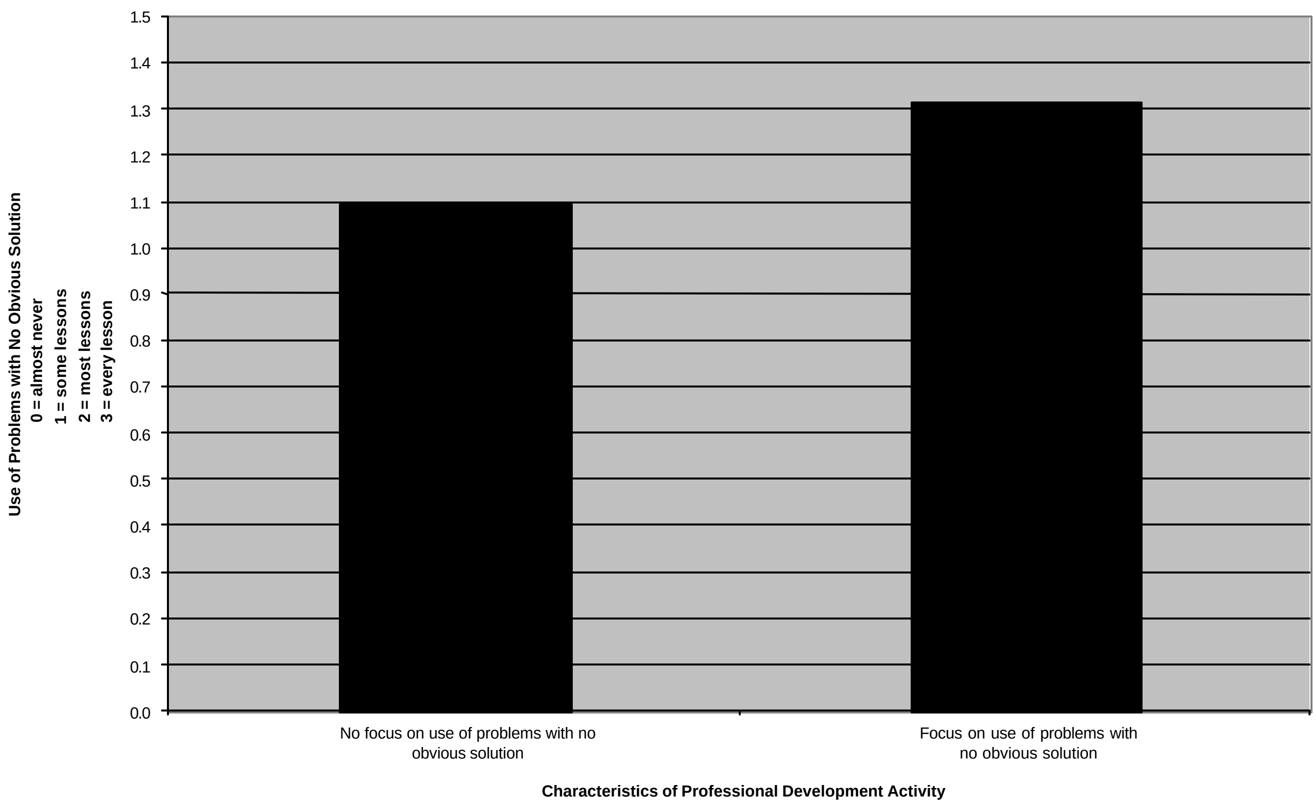
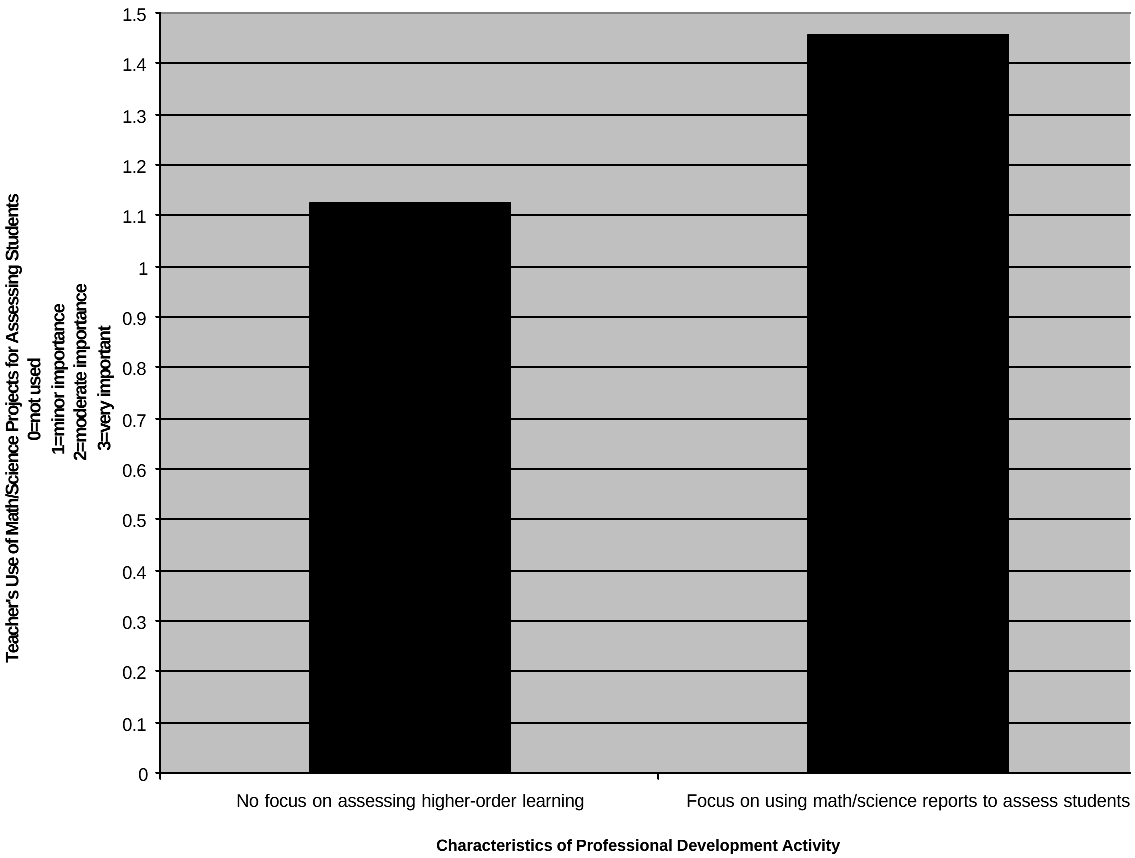


Exhibit 13 illustrates the same general pattern in our third area—assessing students' higher-order learning. The exhibit focuses on a single strategy in this area—using mathematics and science projects to determine student grades. The exhibit illustrates that without professional development on this specific assessment strategy, teachers attach minor importance to it. (The first bar in the left of the graph is at slightly over 1.1, where 1= minor importance of the strategy in determining students' grades in mathematics and science class.) If teachers participated in professional development that focused on this strategy, they were likely to attach more importance to it. (The second bar increases the effect to 1.4, where 1=minor importance, and 2= moderate importance.)

EXHIBIT 13

Effects of Professional Development on the Use of Mathematics and Science Projects to Determine Student Grades



These results show that professional development does have an effect on classroom practice. When professional development focuses on specific teaching strategies, teachers increase their use of these strategies in their classrooms. This effect is independent of teachers' prior use of these strategies, the subjects they teach, or the level of school. The results also suggest that there may be an added benefit for teachers who participate in professional development that focuses on a number of specific, related teaching strategies. Studies with a larger number of teachers are needed to fully test this hypothesis, however. If the spillover hypothesis is correct, this suggests that building teachers expertise in one specific area increases teachers' tendency to apply that expertise in similar areas.

The Increased Impact of Professional Development Activities with Specific Features of High Quality

Having found that professional development focusing on specific teaching strategies had effects on the use of these strategies in the classroom, we sought to examine the extent to which features of high quality increased the effectiveness of the professional development. As discussed above, our national data identified six features of professional development associated with high quality. Three "structural features" that describe the way the activity is designed—reform type, duration (including contact hours and time span), and collective participation—and three "core features" that describe the substance of the activity—active learning, coherence, and content focus—led to increases in teachers self-reported knowledge and skills and to changes in classroom practice. We wanted to know whether these six features of high-quality professional development increased the effects of strategy-focused professional development.

Again we used "hierarchical linear modeling," a multivariate statistical technique, to determine the effect of these six features, independent of other factors that could have an effect on teachers' use of specific higher-order strategies in their classroom practice. In these analyses, described in Appendix D , we examined two aspects of professional development: (1) the extent to which it focused on specific higher-order strategies in 1997–98 and (2) the extent to which it was characterized by one of the six features of high quality.

The analyses described above in Exhibits 11, 12, and 13 showed that professional development activities that focused on a specific higher-order strategy in 1997–98 increased teachers' use of that strategy in their classroom in 1998–99. The new analyses described below enables us to measure whether participating in professional development that had features of high quality further increased teachers' use of specific teaching strategies in the following school year. Because our analyses controlled for prior use of the specific teaching strategies in 1996–97, and teachers' subject (mathematics or science) and grade level taught (elementary, middle, and secondary), we were able to see the effects of the quality of professional development on teaching practice independent of these other factors. We found the following:

- **Features of high quality—reform type, collective participation, active learning, and coherence—strengthen the effect of professional development activities that focus on specific higher-order teaching strategies.**

Exhibit 14 provides an overview of the relationships between teachers' use of specific higher-order strategies in their classroom practice and professional development that (1) focused on these or similar specific practices and (2) had features of high-quality professional development. Again, we examined higher-order teaching strategies in three areas—technology use, instructional methods, and assessment of student performance. The exhibit shows that for almost all of the analyses, the features of high quality are in the positive direction, which indicates that they increased the effect of professional development that focused on specific higher-order teaching strategies or sets of related strategies. Relatively few of these effects, however, are statistically significant (gray shading indicates that the effects are statistically significant). This may be due, in part, to the size of our sample, which was relatively small for this type of analysis.

EXHIBIT 14

Relationship Between Features of Professional Development and Activities Focused on Specific Pedagogical Strategies (Sign and Significance of Relationships)[44]

Independent Variable: Features of Quality of Professional Development	Technology		Instruction		Assessment	
	Effect of Specific Strategy[a]	Effect of Set of Related Strategies[b]	Effect of Specific Strategy	Effect of Set of Related Strategies	Effect of Specific Strategy	Effect of Set of Related Strategies
Reform type	-	+	+	+	-	+
Contact hours	+	-	+	+	-	+
Time span	+	+	+	+	-	+
Collective participation	+[c]	+	+	+	-	+
Active learning	+	+	-	+	-	+
Coherence	+	+	+	+	+	+

[a] This indicates the effect of professional development if it focused on only one high-order strategy in a particular area.
[b] This indicates the effect of professional development if it focused on all the high-order strategies in a particular area.
[c] Gray shading indicates that the effects are statistically significant.

How to read this exhibit: The "-" in the first row in the first column on the left shows that participating in a professional development activity that is a reform-type activity decreases the effect of professional development focused on higher-order technology use, but this relationship is not statistically significant. The "+" in the fourth row in the first column on the left shows that participating in a professional development activity that has collective participation increases the effect of professional development focused on higher-order technology use, and this relationship is statistically significant (indicated by the gray shading).

The next set of exhibits (Exhibits 15, 16, and 17) illustrates how a particular feature of the quality of professional development strengthens the effects of professional development that focuses on particular teaching strategies.

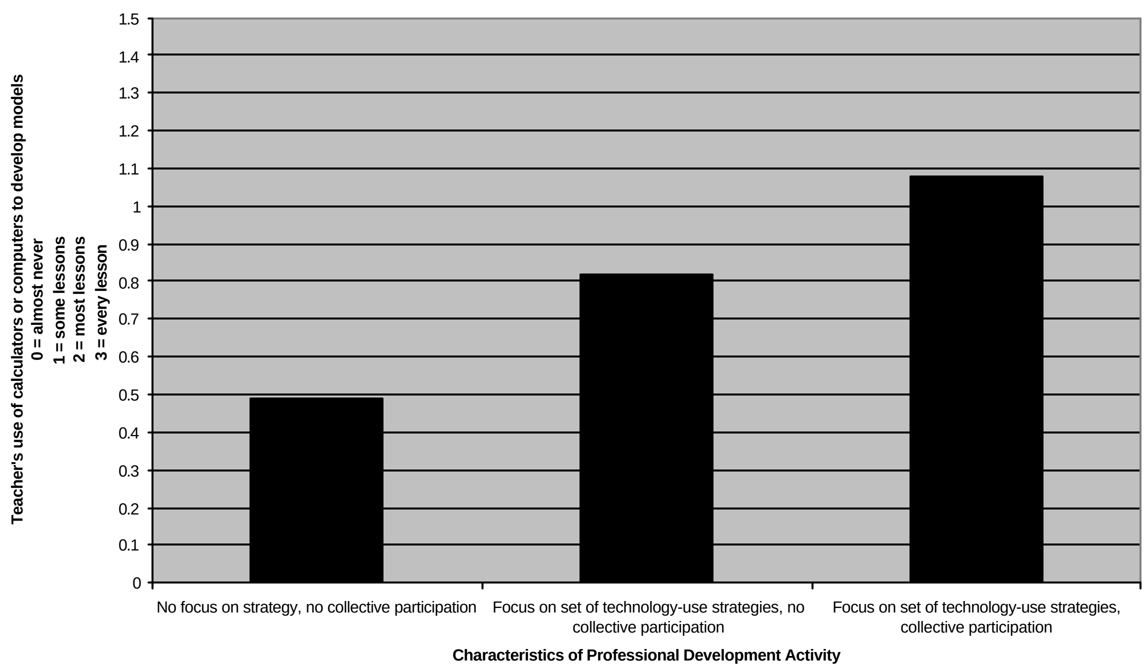
Exhibit 15 illustrates the effects of professional development activities that differ in focus and quality on the use of one particular teaching practice: the use of calculators and computers to develop models. The first bar on the left of the exhibit shows the effect of professional development that did not focus on the use of calculators or computers to develop models and was not characterized by collective participation. This bar indicates that if teachers' professional development neither focused on this strategy nor had collective participation, teachers report using this strategy between "almost never" (response category of 0) and "in some lessons" (response category of 1). The second bar illustrates the effect of professional development activities that focused on the use of calculators and computers to develop models as well as other technology-use strategies (a set of related strategies), but again were not characterized by collective participation. This second bar shows that professional development that focuses on a set of specific strategies of technology use boosted teachers' use of calculators and computers to develop models to over .8, indicating that more teachers are using this specific strategy in "some lessons." Finally, the third bar of Exhibit 15 shows the effect of professional development that is characterized both by a focus on specific strategies of higher-order technology use and by collective

[44] "Content focus" is not included in the list of features of quality because the measure of whether the activity focused on a particular teaching strategy is a proxy measure for content focus.

participation (i.e., teachers attended these professional development activities with other teachers from their own school, department, or grade level). This bar shows that professional development characterized by collective participation boosts even further the use of calculators and computers to develop models—to more than 1.0. This indicates that teachers who participated with their colleagues in professional development that focused on specific strategies for using technology used these strategies in "some lessons," on average.

EXHIBIT 15

Effects of Professional Development on the Use of Calculators and Computers to Develop Models, by the Activity's Focus on Specific Technology-Use Strategies and Collective Participation

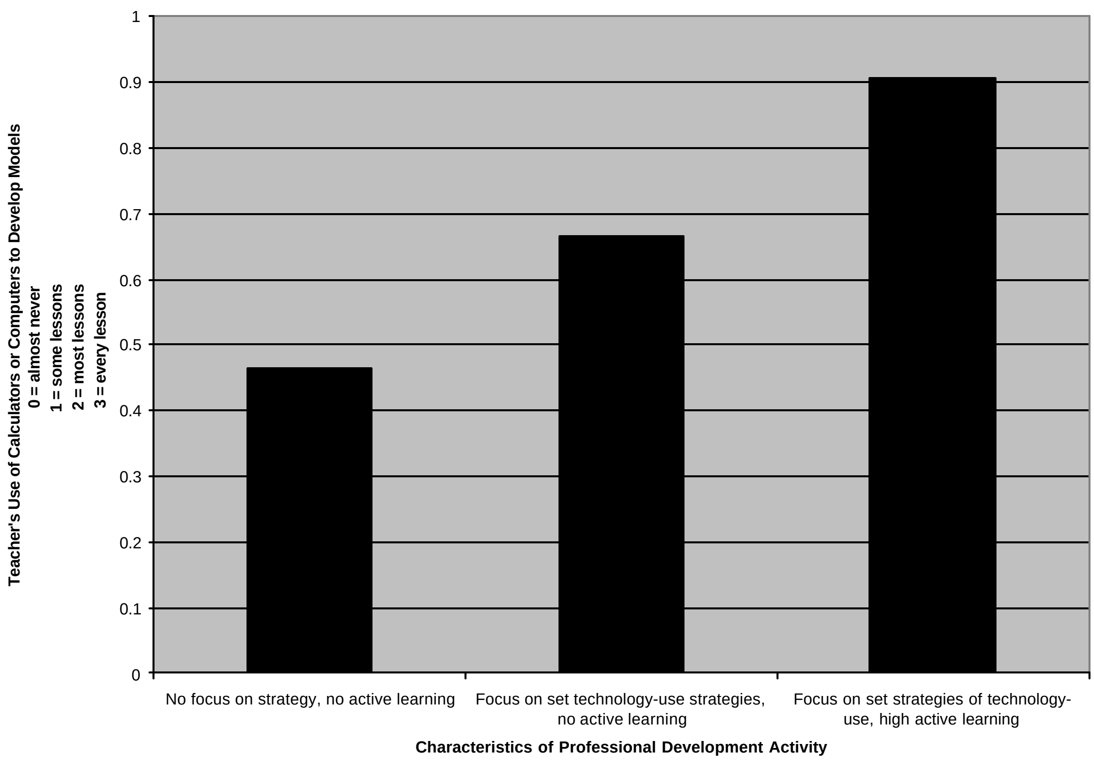


The results in Exhibit 15 suggest a substantial benefit when teachers from the same school, department, or grade level participate together in technology-related professional development. This is consistent with ideas about best practice and the way teachers learn and implement new knowledge, which suggest that teachers benefit from relying on one another in developing technological skills. Our findings are also consistent with the idea that professional development that is characterized by "active learning," where teachers are not passive "recipients" of information, also boosts the impact of professional development activities, as illustrated by Exhibit 16.

In Exhibit 16, the emphasis is on the effect of active learning rather than collective participation on the use of calculators and computers to develop models. The first bar on the left of the exhibit shows the effect of professional development that did not focus on this strategy and was not characterized by active learning. This bar indicates that if teachers' professional development neither focused on this strategy nor provided any opportunities for active learning, teachers report using this strategy between "almost never" (response category of 0) and "in some lessons" (response category of 1). The second bar illustrates the effect of professional development activities that focused on the use of calculators and computers to develop models as well as related technology-use strategies; but again, these activities were not characterized by active learning. This second bar indicates that professional development that focuses on specific strategies for using technology boosted teachers' use of calculators and computers to develop models to more than .6, indicating that more teachers are using this specific strategy in "some lessons." Finally, the third bar of Exhibit 16 shows the effect of professional development that is characterized both by a focus on specific strategies for technology use and substantial opportunities for active learning. This bar shows that professional development characterized by high active learning boosts even further the use of calculators and computers to develop models—to about .9. This indicates that, on average, teachers who had professional development that focused on specific technology-use strategies and had high opportunities for active learning used these technology strategies in "some lessons."

EXHIBIT 16

Effects of Professional Development on the Use of Calculators and Computers to Develop Models, by the Activity's Focus on Specific Technology-Use Strategies, and Active Learning



To sum up, Exhibits 15 and 16 illustrate the general patterns in our analyses, using specific examples of professional development focused on higher-order technology strategies. The exhibits show the effects on teachers' use of one higher-order technology strategy—the use of calculators and computers to develop models—of professional development activities that have two specific features. First, the activities focus on specific higher-order teaching strategies for using technology in the classroom. Second, the activities are characterized by one of the features of high-quality professional development—collective participation or active learning—that teachers in our national data reported having an effect on their knowledge and skills and their classroom practice. Our data analyses contain other examples that illustrate the general finding—that the features of high quality strengthen the effect of professional development focused on specific higher-order teaching strategies. Exhibit 17 is another illustration of this pattern, this time showing the effect of professional development on a specific higher-order instructional method, the use of problems for which there is no obvious solution.

EXHIBIT 17

Effects of Professional Development on the Use of Problems With No Obvious Solution, by the Activity's Focus on Problems with No Obvious Solution and Reform-type Participation in Professional Development

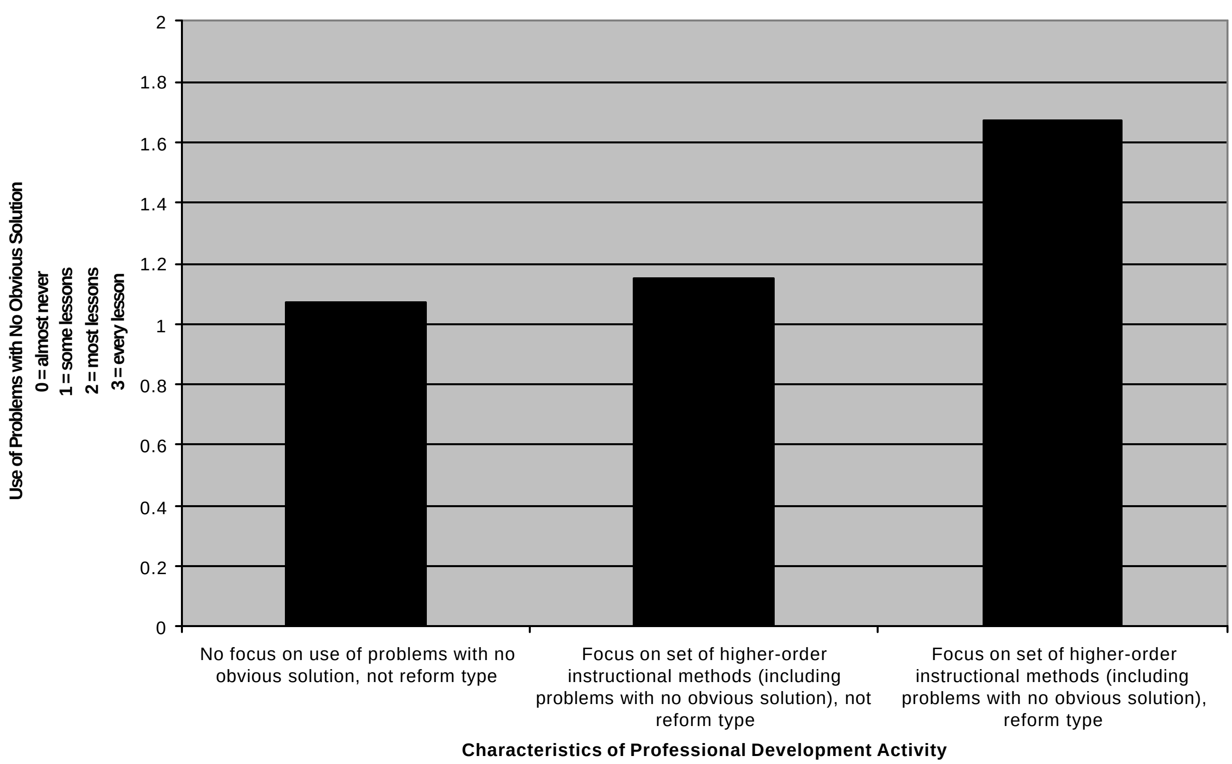


Exhibit 17 illustrates the effects of professional development of different focus and quality on teachers' use of a particular higher-order instructional method: the use of problems for which there is no obvious solution. The first bar on the left of the exhibit shows the effect of professional development that did not focus on this strategy and was not a reform type of activity (i.e., the activity was a traditional workshop, course, or conference rather than, for example, a study group, network, or mentoring relationship). This bar indicates that if teachers' professional development did not focus on the use of problems with no obvious solution and was not a reform type of professional development, teachers generally report using this strategy "in some lessons" (response category of 1). The second bar illustrates the effect of professional development activities that focused on the use of problems with no obvious solution and other related higher-order instructional strategies; but again the professional development activities were not reform types. This second bar indicates that professional development that focuses on the set of higher-order instructional methods boosted teachers' use of problems for which there is no obvious solution to almost 1.2, indicating that more teachers were using this strategy in "most lessons" (response category of 2). Finally, the third bar of Exhibit 17 shows the effect of professional development that is characterized by a focus on higher-order instructional strategies and is a reform type of professional development. This bar shows that reform types of professional development boost even further the use of problems with no obvious solution to about 1.6. This indicates that teachers who participated in reform types of professional development that focused on specific higher-order instructional methods used these methods in "most lessons."

The results in Exhibit 17 suggest a substantial benefit when teachers participate in reform types of professional development that focus on specific higher-order instructional methods. Again, this finding is consistent with ideas about best practice regarding how teachers learn and implement new knowledge. To use problems with no obvious solutions in their classrooms, teachers must have a deeper understanding of how children think and solve problems. Many researchers and reformers suggest that for teachers to have this deeper understanding, they must interact with their colleagues on a regular basis to discuss their work and their students' learning (Little, 1993; Loucks-Horsley et al., 1998). Reform types of professional development generally intend to provide more opportunities for such interaction.

Summary: The Effects of Professional Development on Teaching Practice in Mathematics and Science

We found that professional development that had features of high quality did significantly increase teachers' use of active, project-centered instruction. In particular, reform-type professional development and professional development characterized by teachers' active learning and coherence with teachers' goals and other aspects of teachers' environment (e.g., standards and assessments) were more likely to foster teachers' use of active, project-centered instruction than professional development that did not have these features. This is probably due to the link between specific features of professional development and active, project-centered instruction (e.g., reform-type activities are often characterized by the use of active instruction). However, our initial analyses found few effects of the quality features of professional development on the broad measures of teaching practice that we used in our national survey.

We also examined the effects of professional development that focused on specific teaching strategies that are intended to increase students' higher-order learning. We focused these analyses on teaching strategies in three areas: technology use, instructional methods, and approaches to assessing students' work. We found that professional development that focused on these strategies in 1997–98 did increase teachers' use of these strategies in1998–99.

Turning back to the features of high-quality professional development, we found that these features further increased the effect of professional development focused on specific teaching strategies. For example, professional development focusing on specific strategies for using technology for higher-order learning increases teachers' use of these strategies, independent of teachers' prior use of these strategies, the subject they teach, or the level of school in which they teach. If this professional development also is characterized by collective participation (i.e., teachers attending with colleagues from their school, grade, or department) or active learning (e.g., opportunities for meaningful analysis of student work), the impact on using the strategies in the classroom is even greater. Thus, our findings from the longitudinal sample of teachers support the findings from our national data, which identified the importance of the six features of high-quality professional development for enhancing teachers' knowledge and skills and classroom practice.

To examine our findings on the effects of professional development in the context of trends in teaching practice, the next section describes changes in teaching practice in our longitudinal sample over the three-year study period.

SECTION IV

Trends in Teaching Practice

In this section, we put our results on the effects of professional development in context by examining how much average change over time there is in different dimensions of teaching practice. In particular, we use our three years of data on teaching practice to describe trends and differences across teachers and schools in the three broad aspects of teaching practice that we discussed in Section III:

- **The overall alignment of teaching with national standards**, using the National Assessment of Education Progress (NAEP) as the national standard
- **Teachers' emphasis on the six performance goals** (i.e., memorize, understand concepts, perform procedures, generate hypotheses, collect, analyze/interpret, and make connections)
- **Teachers' use of the four pedagogical strategies** (i.e., teacher-centered instruction, individual seatwork, active instruction, and discussion-oriented instruction)

First we report patterns in teacher change over the three-year period of our study on alignment, performance goal emphasis, and pedagogical strategies. Then we analyze differences in teaching practice between subjects and school levels, schools, and teachers and for individual teachers over time.

Do Teachers Change Their Teaching Practice?

- **In our longitudinal sample, we find little change in overall teaching practice from 1996 to 1999. In particular, there was little change in teachers' average degree of alignment with the NAEP, in patterns of emphasis on performance goals, and in pedagogy.**

Exhibit 18 shows the mean, or average, levels of emphasis for all 11 measures of teaching practice for each of the three years of the study. These mean levels of emphasis reflect teaching practice across all teachers in our longitudinal sample. (See Appendix E for mean levels of emphasis by subject and school level.) Using hierarchical linear modeling, we tested whether there were significant changes in these mean levels of emphasis, controlling for subject, school level, and the interaction of subject and school level. We found that generally, teachers' classroom practice remains stable across the three years of the study. For example, the first row in Exhibit 18 indicates that in 1996–97, on average, 21 percent of teachers' content coverage was aligned with the NAEP. (This exhibit is identical to Exhibit 7 in Section III.) Although the alignment of content coverage with the NAEP drops to 20 percent in 1997–98 and 1998–99, this difference is not statistically significant. The exhibit also shows that teachers' average emphasis on higher-order performance goals and topics and on good pedagogical strategies also do not increase over time.[45,46]

[45] At the outset of the study we targeted several topics to monitor over time. These were subjects, such as probability, statistics, measurement, and geometry, that have been identified as special weaknesses for students in

EXHIBIT 18
(also displayed in Section III, Exhibit 7)
Average Characteristics of Teachers' Instruction for 1996–97, 1997–98, and 1998–99

	Means by Year		
Dimension of Teaching Practice	**1996–97**	**1997–98**	**1998–99**
Alignment of Content Coverage with the NAEP *Scale:* From 0 to 100% alignment with the NAEP	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Alignment Index	21% (9)	20% (9)	20% (8)
Performance Goals *Scale:* From 0 to 100% emphasis on the performance goal			
Memorize	15% (7)	16% (6)	15% (6)
Understand Concepts	23% (6)	22% (5)	22% (5)
Perform Procedures	20% (6)	20% (6)	20% (7)
Generate Hypotheses	11% (5)	12% (5)	12% (5)
Collect, Analyze & Interpret	13% (5)	13% (4)	13% (5)
Make Connections	17% (5)	17% (4)	17% (4)
Pedagogy *Scale:* Standardized scale where mean=50 and standard deviation=10, in 1996-97			
Didactic Instruction	50.1 (5.6)	50.6 (5.8)	49.9 (6.2)
Individual Seatwork	49.5 (5.4)	49.4 (5.5)	49.1 (5.6)
Active, Project-centered Instruction	49.7 (5.7)	49.5 (5.8)	49.1 (6.4)
Discussion-oriented Instruction	50.0 (10.8)	49.6 (9.3)	49.4 (8.7)

How to read this exhibit The percent alignment of teachers' content coverage with the NAEP decreased from 21 percent alignment in 1996–97 to 20 percent alignment in 1997–98 and 1998–99. These changes are not statistically significant. The percent emphasis that teachers give to the performance goal of memorization increased from 15 percent in 1996–97 to 16 percent in 1997–98; it decreased to 15 percent in 199899. These changes are not statistically significant. Teachers' use of the pedagogical strategy of didactic instruction increased from 50.1 to 50.6 in 1997–98; it decreases to 49.9 in 1998–99. These changes are not statistically significant. Pedagogy is on a standardized scale where 50 is the mean level of use of each of the four strategies; this number is not a percent, but rather a general metric designed to provide a point from which to show increases or decreases in teachers' use of a particular pedagogical strategy. The mean was standardized in 1996–97, but could vary in 1997–98 and 1998–99.

There may be several explanations for the lack of change in overall teaching practice. First, it is not unreasonable to think that teachers teach in ways that they believe to be effective and appropriate and do not make substantial changes in their practice from year to year, as a result of either professional development or other influences. We would not expect other professionals, in fields such as law or medicine, to make substantial changes in their behavior on the basis of one or

the United States (Beaton et al., 1996). We conducted analyses of these and an exhaustive list of topics by subject and school level, and the only topic for which there was even a marginally significant increase was advanced algebra in high school (p<.07, n=28). These analyses were done separately by school level and subject and had sample sizes of approximately 30 teachers, which limits the power to detect effects.

[46] The only significant change over time effect is an interaction between year and subject for teachers' emphasis on memorizing and understanding concepts. See Appendix E for the results of these analyses.

even several professional development experiences. Second, teachers may be getting too many signals to teach in conflicting ways, and the tensions may have the effect of canceling each other out. Systemic reforms foster multiple signals, and many of the signals may conflict, which could create a tension that causes teachers to be resistant to change (Cohen & Spillane, 1992).[47]

It also is possible that teachers may be changing, but slowly. We measured teaching at only three points in time over three years, and thus we may not have measured teachers long enough to capture substantial change. Further, teachers may be changing, but in ways not captured by our study. For example, teachers may be increasing their alignment with state standards, but not with the NAEP.[48] Finally, teachers may be shifting in a particular direction, but to increase emphasis on certain types of performance goals or pedagogy, teachers have to de-emphasize other strategies. This may result in the overall appearance of no change when the data are aggregated, when actually changes in one direction may offset changes in another direction.[49]

These results also highlight the measurement and theoretical complexities in the study of teacher change. There may be important differences between measures designed to capture teachers' instruction at one point in time and measures that are designed to detect teacher growth (see Rowan, 2000). For example, as seen in Section III of this report, we were able to detect the effects of professional development on teachers' use of specific strategies intended to foster higher-order student learning, but we found relatively few effects on global measures of teaching. We do not have any working paradigms for teacher growth—either theoretical or empirical frameworks for what types of changes one would expect, their magnitude, and when these changes would occur. This is an area where more theoretical and empirical work is needed. Despite these caveats, our findings provide strong evidence of the stability of teachers' classroom practice.

Variation in Teaching Practice Between Schools, Between Teachers Within Schools, and Across Time

Although we observe little change in classroom practice on average, individual teachers differ substantially in their classroom practice. We found the following:

- **Despite little *average* change over time in teaching practice in our longitudinal sample, individual teachers in our sample do vary in their classroom practices, and moderate variation does occur in the classroom practice of individual teachers from year to year.**

Although there is very little change across years in average teaching practice, there still are considerable differences between teachers, as illustrated by the standard deviations reported in Exhibit 18. For example, the exhibit indicates that on average, teachers gave 15 percent of their instructional emphasis to the performance goal of memorization in 1996–97. The standard deviation of 7 percent indicates that many teachers gave as little as 8 percent emphasis to this performance goal, whereas many others gave 22 percent emphasis to this goal. Similarly, if we look again at

[47] We observed during our site visits in 1996–97 that many of the schools were engaged in multiple reform efforts.

[48] AIR is currently conducting a study, *Moving Standards to the Classroom*, sponsored by the Planning and Evaluation Service, which is focused on examining the extent to which teachers' instruction in mathematics is aligned with state standards; in addition, the Wisconsin Center for Education Research, in collaboration with the Council of Chief State School Offices (CCSSO), has conducted a study of the alignment of teaching with state assessments and is planning a study of the alignment of teaching with state standards.

[49] This also might help explain the random variation in individual teachers' trajectories that we found.

alignment, the exhibit shows that in 1996–97, the average alignment of teachers' content coverage with the NAEP was 21 percent. The standard deviation of 9 percent, however, indicates that for many teachers, only 12 percent of content coverage was aligned with the NAEP, whereas for many other teachers, 30 percent of their content coverage was aligned with the NAEP. So, although we observe considerable stability in classroom practice on average, individual teachers differ substantially in their classroom practice.

Our longitudinal data enable us to answer many questions about the variation in teachers' classroom practice. For example, do teachers who teach different subjects or levels of school have different patterns of content alignment, performance goals, and pedagogy? Do teachers in some schools differ systematically in their classroom practice from teachers in other schools? Do teachers within the same school differ from one another in their classroom practice?

Our longitudinal data also enable us to answer questions about changes in individual teachers' instruction over time. Are teacher differences consistent across years? For example, do individual teachers who emphasize active, project-centered instruction in 1997–98 also emphasize this aspect of their classroom practice in 1998–99?

Using a statistical technique called "hierarchical linear modeling" we were able to use our longitudinal data to estimate what proportion of the "variation" in teachers' classroom practice (i.e., the 11 aspects of teaching that we measure) could be attributed to (1) average differences between teachers who teach different subjects or levels of school, (2) average differences between schools, (3) average differences between teachers in the same school, and (4) average differences between one year and the next. (See Appendix E for an explanation of the analyses undertaken for this section of this report.) For each dimension of teaching practice shown in the first column of Exhibit 19, we show the percent of the variation in teachers' professional development experiences that we can attribute to each of these differences.

The last column of Exhibit 19, labeled "unexplained year-to-year variation in individual teaching practice," shows the percent of the variation in individual teachers' practice between years that cannot be explained by any of the differences listed above. If teachers' classroom practice were totally consistent from one year to the next, all of the variation in their instruction would result from differences between teachers' subject and school levels; differences between teachers in the same school; differences between schools; and differences in the average teaching practice between one year and the next. None of the variation would be left unexplained.

EXHIBIT 19

Percent of Variation in Teaching Practice (1996–97, 1997–98, and 1998–99)

Dimensions of Teaching Practice	Percent of Variation in the Quality of Professional Development Explained by Differences Across				Unexplained Year-to-Year Variation in Individual Teaching Practice
	Subjects and School Level	Teachers in the Same School	Schools	Years	
Alignment					
Alignment Index *Scale:* From 0 to 100% alignment with the NAEP	42.4%	27.2%	0.0%	0.0%	30.4
Performance Goals *Scale:* From 0 to 100% emphasis on the performance goal					
Memorization	0.0%	39.1%	2.2%	0.0%	58.7
Understanding concepts	0.0	30.0	2.5	0.0	67.5
Performing procedures	27.9	29.7	0.0	0.0	44.2
Generating hypotheses	6.7	40.0	0.0	0.0	53.3
Collecting/analyzing/interpreting data	10.3	34.5	0.0	0.0	55.2
Making connections	3.7	29.6	0.0	0.0	66.7
Pedagogy *Scale:* Standardized scale where mean=50 and standard deviation=10, in 1996–97					
Traditional	25.3	40.9	6.2	0.0	27.6
Individual seatwork	10.1	40.6	9.6	0.4	39.3
Active project-centered learning	16.7	36.2	8.6	0.0	38.5
Discussion-oriented instruction	2.6	30.7	4.8	0.0	61.8

How to read this exhibit: Forty-two (42) percent of the variation in alignment is due to teachers' subject and school levels; none of the variance is due to differences between schools or average differences in teaching practice across years; 27 percent of the variation is between teachers within the same school, and 30 percent of the variation is unexplained year-to-year variation. Thus, teacher and subject/level differences account for most of the variance, and the remaining 30 percent is unexplained year-to-year variance.

The data reported in Exhibit 19 indicate that depending on the specific dimension of teaching practice, a substantial part of the variation in teaching practice can be explained by differences in teachers' subjects and school levels. For example, over 40 percent of the variation in alignment is due to subjects and levels. This may reflect the fact that, for example, elementary school mathematics instruction is more aligned with the NAEP than are other subjects and grades. (See Appendix E.) The exhibit also shows that almost 30 percent of the variation in emphasis on the performance goal "performing procedures" is due to subjects and levels. This supports our earlier finding from the analysis of our first year of longitudinal data, reported in Garet et al., 1999, that teachers of mathematics emphasize this performance goal significantly more than other teachers. Finally, about 25 percent of the variation in teachers' emphasis on traditional pedagogy is explained by teachers' subjects and school levels. This probably reflects the fact that high schools tend to have more traditional pedagogy than do middle or elementary schools. Apart from these differences between teachers' subjects and school levels, most of the variation in teaching practices is between teachers within the same school. (See Appendix E for a discussion of these results.)

There is practically no variation across schools in alignment and performance goal emphasis, and there is a small difference across schools in pedagogy. This difference may be due to the adoption of whole-school reform designs, which often focus on pedagogical strategies (e.g., some designs hinge on project-centered learning).

Finally, Exhibit 19 indicates that there is essentially no year-to-year variation in average teaching instruction. This reflects our earlier finding of little change in average overall teaching practice over time. Although there is substantial variation across teachers in their teaching practice, this variation can be attributed to differences across individual teachers in the same school to teachers who teach mathematics instead of science, or to teachers who teach at different levels of school.

A substantial amount of year-to-year variation in the teaching practice of individual teachers remains unexplained in our analysis. This unexplained year-to-year variance is higher for the six performance goals than for alignment or for the four measures of pedagogy. The high year-to-year variation in emphasis on performance goals may indicate that teachers adjust these goals to changes in the specific student composition of their classes each year.[50]

Summary

Although teachers may be changing on dimensions or qualities of practice that we did not measure, it is evident that on many central dimensions of classroom practice, teachers in the 30 schools we studied did not change from 1996–97 through 1998–99. However, despite the consistency of teaching practice over time, we found variation between individual teachers. We found differences in teaching practice by school level and subject, and we found that most of the variation in teaching practice is between individual teachers within the same schools, rather than between schools. Greater differences in teaching practice between schools might indicate consistent, systematic school-level instructional plans, but evidence of such planning was not found in our data.

Our findings highlight the value of conducting studies with increased power to measure change (for example, measuring teachers for more than three years). The findings also highlight the value of studies that incorporate models of teacher growth that would indicate what types of changes we would expect in teaching practice, as well as their timing and magnitude. For example, it would be helpful to have information on the type of dimensions on which we expect teachers to change and by how much we would expect them to change during a one-year period versus a two-year period.

Lastly, although teachers do not report changing their teaching practice in ways we might consider desirable, the fact that they are consistent over a three-year period in reporting their instructional practices lends strong support to our survey instrument as a reliable measure of teachers' instruction.

In the final section of this report, we summarize and synthesize our findings on teachers' professional development experiences and the effects of professional development on instruction and trends in teaching practice, and we suggest implications for designing and supporting professional development through the Eisenhower and other programs.

[50] The fact that only 30 percent of the variation in alignment is due to unexplained year-to-year variation indicates that the "test-retest" reliability of our measure of content coverage is relatively high. If all the year-to-year variation were due to measurement error, the implied reliability would be .70.

SECTION V

Summary and Implications

The Eisenhower Professional Development Program's primary goal is to support professional development for teachers that will improve classroom practice and, ultimately, student achievement. AIR's multiyear evaluation of the Eisenhower Program sought to provide information to help the Eisenhower Program, and other professional development efforts, achieve this important goal. Our evaluation was intended to help policymakers and program managers by (1) describing professional development activities supported by the Eisenhower program and the way they are managed and implemented and (2) evaluating the effects of the professional development activities on teaching practice.

To describe Eisenhower-funded activities and the effects of professional development on teaching practice, the national evaluation included both a nationally representative cross-sectional component and a more focused longitudinal component. Our analyses based on national samples of district Eisenhower coordinators, state agency for higher education (SAHE)-grantee project directors, and teachers, were reported in Garet et al., 1999. They describe the type and quality of Eisenhower activities, who participates in them, how they fit into other reform efforts, and how they are managed and implemented by districts and SAHEs. The Longitudinal Study of Teacher Change, described in this report, examines the effects of Eisenhower-assisted and other professional development on teachers in 30 schools. We analyzed the effects of professional development on broad and specific measures of teaching practice: the alignment of content coverage with national standards; an emphasis on performance goals for students; pedagogical strategies; and higher-order use of technology, instructional methods, and student assessments.

The findings and implications that we draw from both the national and longitudinal data, and summarize in the following section, extend beyond the Eisenhower Program. Although our national data on professional development focused only on Eisenhower-assisted activities, the longitudinal teacher survey included professional development funded by Eisenhower and other sources as well. This occurred because we asked teachers to describe the most helpful activity they participated in during the school year, and some of the activities that teachers chose were not funded by Eisenhower. Nevertheless, because our focus in all aspects of our study was on relationships between features of professional development and teaching practice, our findings apply to teachers' professional development in mathematics and science in general, whatever the funding source.

The Effects of Professional Development on Teaching Practice

On the basis of our national data, we concluded that six key features of professional development are effective in improving teaching practice: three structural features (characteristics of the structure of the activity)—reform type, duration, and collective participation—and three core features (characteristics of the substance of the activity)—active learning, coherence, and content focus (see Garet et al., 1999). These findings from our national data support other recent studies that highlight the importance of content focus in professional development (e.g., Cohen & Hill, 1998; Kennedy, 1998). The features of high-quality professional development identified in our national data also are consistent with ideas articulated in the Eisenhower legislation. Further, they deepen and extend the ideas in the Eisenhower legislation by providing details about what makes professional development effective. For example, the Eisenhower legislation promotes professional development that is linked to other reform efforts in a coherent, systematic way. The results from our national data show the effectiveness of specific dimensions of coherence, such as discussing professional development experiences with colleagues and participating in follow-up activities that build on previous activities.

The findings from our longitudinal data reinforce the importance of the six features of professional development identified in the national study. In addition, results from our longitudinal study extend our national findings by providing evidence of the link between focusing on specific teaching strategies in professional development and having teachers use those specific strategies in the classroom. Specifically, in our longitudinal study, we found the following:

- **Professional development focused on specific, higher-order teaching strategies increases teachers' use of those strategies in the classroom. This effect is even stronger when the professional development activity has features of high quality (e.g., reform type, active learning, coherence, and collective participation).**

These findings are especially strong because they are based on only one professional development experience per teacher per year. Teachers may experience many professional development activities in one year, so it is especially noteworthy that we found effects on teaching practice of the one experience that teachers chose to describe on our survey.

Our longitudinal data also indicate that professional development is more effective in changing teachers' classroom practice when it has specific features of high quality, such as the collective participation of teachers from the same school, department, or grade; active learning opportunities, such as reviewing student work or obtaining feedback on teaching; and coherence, for example, linking to other activities or building on teachers' previous knowledge. These findings are based on longitudinal data collected at three points in time. They validate the results from our national probability sample of teachers in Eisenhower-assisted activities, which indicated that features of quality were significantly related to teachers' self-reported outcomes (Garet et al., 1999).

Participation in Professional Development

Our findings on the effects of professional development should be considered in the context of the nature and quality of teachers' experiences in professional development. Our results suggest that change in teaching would occur if teachers experienced consistent, high-quality professional development. But we find that most teachers do not experience such activities. On average, the activities experienced by teachers in our Longitudinal Study of Teacher Change are about the same quality as those experienced by our national sample of teachers in Eisenhower-assisted activities. Our national data indicated the following about district-supported Eisenhower activities: an average of only 23 percent of teachers participating in Eisenhower-assisted professional development were in reform types of professional development; the average time span of a professional development activity was less than a week; the average number of contact hours was 25 and the median was 15 hours; most activities did not have collective participation or a major emphasis on content; and most activities had limited coherence and a small number of active learning opportunities (see Garet et al.,1999 for more details). In short, nationwide, the typical professional development experience was not high-quality. Nevertheless, our national data also documented great variation in the quality of teachers' professional development experiences, which indicates that at least some teachers participate in high-quality activities, at least some of the time.

Our longitudinal data indicates that the quality of professional development experiences varies considerably not only across teachers at a single point in time but also over time for the same teachers:

- **Teachers experience professional development that varies in quality from one year to the next. Further, teachers in the same school tend to have quite different professional development experiences.**

We find a substantial amount of year-to-year variation in the quality of the professional development of individual teachers. For example, 79 percent of the variation in the span and 62 percent of the variation in the content focus of a teacher's professional development experience are due to year-to-year variation. This finding indicates that the average teacher's professional development experiences do not add up to a long-term, coherent, high-quality program—the type of program that has the most potential for fostering significant and lasting teacher change.

We find some variation in participation in professional development between schools (e.g., 14 percent of the variation in collective participation and 7 percent of the variation in active learning is due to between-school variation), but most of the variation in the quality of the professional development in which teachers participate lies *within*, not *between*, schools. This finding supports the idea that professional development continues to be an individual teacher experience. Both our national and our longitudinal data indicate that professional development is more effective when teachers participate with others from their school, grade, or department. Thus, the variation in teachers' professional development experiences within the same school helps explain why professional development is not as effective as it could be.

Trends in Teaching Practice

Perhaps partly as a result of the uneven quality of professional development, we find the following:

- **In our longitudinal sample, we find little change in overall teaching practice from 1996 to 1999.**

Beyond the specific and targeted instructional practices, where we do observe change as a result of professional development, more generally we see little overall change in self-reported teaching practice. Given the usual low quality and inconsistent nature of professional development in which teachers participated, it is perhaps not surprising that we find little change in overall teaching practice over the period of the study. Our data show that teachers' alignment of content with national standards, the goals that teachers have for their students, and their basic pedagogical strategies appear to remain highly stable over time. It may be true that teachers changed on dimensions that we did not measure or that they changed the way they *implemented* certain practices instead of changing their relative emphasis on these practices. However, given the multiple and high-profile efforts of standards-based and school-based reforms to provide professional development to change teachers' practice in desirable ways, we are surprised that teachers, as a group, did not move in the directions in which reforms intend to push them.

This lack of results may be a function of weak and fragmented professional development. We find professional development with desirable features in short supply, and where it does occur, it does not occur systematically over time for particular teachers. Few teachers experience the kind of consistent, high-quality professional development that we have found changes teachers' instruction in desirable ways.

Measuring instruction at multiple points over a more extended period of time might increase our ability to capture change in average teaching practice. However, we are confident in our results that at least for the three years of our study, teachers changed little in terms of the content they teach, the pedagogy they use to teach it, and their emphasis on performance goals for students.

- **Despite little *average* change over time in teaching practice in our longitudinal sample, individual teachers in our sample do vary in their classroom practices, and moderate variation does occur in the classroom practice of individual teachers from year to year.**

Although in our longitudinal sample, teachers' practice did not change on average, individual teachers did make moderate changes in their teaching practice from one year to the next. For example, 30 percent of the variation in alignment and 28 percent of the variation in the use of traditional pedagogy is due to year-to-year variation. This year-to-year variation might be due to teachers' adapting to the ability levels of their students or to other influences related to their students or school.

Further, we find a great deal of variation across teachers in their classroom teaching practice. Most of this variation is between teachers in the same school, not between schools. For example, 40 percent of the variation in teachers' use of generating hypotheses and 31 percent of the variation in teachers' use of discussion-oriented instruction are due to variation between teachers in the same school. A substantial amount of variation between schools might suggest a coherent, organized school-fostered system of instruction. Instead, we find that individual teachers in the same school have very different teaching practices. This finding only adds support to the concept that both teaching and professional development are typically individual experiences.

Implications for Policy and Practice

In sum, we find that high-quality professional development that focuses on specific teaching strategies does affect teaching practice and that this effect is stronger if the professional development has the six dimensions of quality identified in the analysis of our national sample of teachers—the professional development is a reform rather than traditional type, is sustained over time, involves groups of teachers from the same school, provides opportunities for active learning, is coherent with other reforms and teachers' activities, and is focused on specific content and teaching strategies. However, teachers generally do not experience consistent, high-quality professional development. Professional development remains an experience that varies substantially from one teacher to the next, and even from one year to the next for a given teacher. Districts and schools face several challenges in providing high-quality professional development to all their teachers.

First, districts and schools often must choose between serving larger numbers of teachers with less focused and sustained professional development or providing higher-quality activities for fewer teachers. As we noted in Garet et al. (1999), good professional development requires substantial resources. Re-allocating resources and combining funding sources can be effective in increasing funds for professional development. However, in the absence of increased resources, the federal government, states, districts, and schools still have to make difficult choices whether to sponsor shorter, less in-depth professional development that serves a large number of teachers or to support more effective, focused, and sustained professional development for a smaller number of teachers. The Eisenhower legislation encourages the idea of sustained, intensive professional development, and the results of this study support the idea that districts and schools might have to focus professional development on fewer teachers in order to provide the type of high-quality activities that are effective in changing teaching practice.

Second, many districts and schools have limited capacity to translate into practice the knowledge about effective professional development. This evaluation has shown that professional development is most effective when it has the six features of quality that we identified earlier—reform type, duration, collective participation, active learning, coherence, and content focus (also Cohen and Hill, 1998; and Kennedy, 1998). As we stated in our last report, more information is needed on the characteristics and conditions that give some districts the capacity to provide this type of high-quality professional development. States and district could benefit from more detailed information and guidance from the federal government about how to use the Eisenhower program to design and provide professional development that has the specific high-quality features that make it effective for teachers.

Third, districts and schools often do not have the infrastructure to be able to manage and implement effective professional development. Improving the quality of professional development is an ambitious undertaking. The analysis of data from our national probability sample of district Eisenhower coordinators showed that planning that includes system alignment (e.g., the alignment of professional development with standards and assessments), funding coordination, and continuous improvement efforts significantly improves the quality of professional development activities that districts provide (Garet et al., 1999). Case data from our 10 districts and data from both our national and our longitudinal studies indicate that some of this planning exists but that it is not systematic or widespread. Our longitudinal study indicates that much of the variation in professional development and teaching practice is between individual teachers within schools, rather than between schools. This finding provides evidence that schools generally do not have a coherent, coordinated approach to professional development and instruction, at least not an approach that is effective in building consistency among their teachers. Participation in professional development is largely an individual teacher's decision; teachers often select the professional development in which they will participate from a number of options available from a highly disparate set of providers. An increased emphasis by the Eisenhower program on the importance of strategic, systematic planning for professional development may encourage both districts and schools to improve their efforts in this area.

In sum, our findings show that the most effective professional development is focused on specific higher-order teaching strategies and has features of high quality. Our national data, however, showed that on average, teachers do not experience high-quality professional development. Having a coherent, long-term plan would enable districts and schools to provide both the depth of professional development experiences needed for them to be effective and the breadth of coverage of specific content and teaching strategies that teachers should learn over time. The provision of high-quality programs of professional development by schools and districts may not completely solve the problem of the variation in the quality of professional development, since participation in professional development remains primarily the decision of individual teachers. Nevertheless, districts and schools could go a long way in developing high-quality professional development activities. To develop meaningful professional development plans, districts and schools would have to overcome challenges to focusing on and setting priorities for professional development activities over time, given limited resources; acquiring knowledge about the features of effective professional development; and building the infrastructure to design and implement the types of activities that teachers need to improve student learning. The Eisenhower Professional Development program and other sources of funding could continue to play an important role in helping districts and schools overcome these challenges and develop high-quality professional development experiences that will lead to better teaching and better learning.

REFERENCES

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

American Association for the Advancement of Science. (1989). *Project 2061: Science for all Americans*. Washington, DC: Author.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS International Study Center.

Birman, B. F., Kirshstein, R. J., Levin, D. A., Matheson, N., & Stephens, M. (1997). *The effectiveness of using technology in K-12 education: A preliminary framework and review*. Paper prepared for the U.S. Department of Education, Office of Educational Research and Improvement.

Birman, B. F., Reeve, A. L., & Sattler, C. L. (1998). *The Eisenhower Professional Development Program: Emerging themes from six districts*. Washington, DC: U.S. Department of Education.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Carey, N., & Frechtling, J. (1997, March). *Best practice in action: Follow-up survey on teacher enhancement programs*. Arlington, VA: National Science Foundation.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12*(3), 311–329.

Cohen, D. K., & Hill, H. C. (1998). *Instructional policy and classroom performance: The mathematics reform in California* (RR-39). Philadelphia: Consortium for Policy Research in Education.

Cohen, D. K., & Spillane, J. (Eds.). (1992). *Policy and practice: The relations between governance and instruction*. Washington, DC: American Educational Research Association.

Cognition and Technology Group at Vanderbilt University. (1994). Multimedia environments for developing literacy in at-risk students. In B. Means (Ed.), *Technology and education reform* (pp. 23–56). San Francisco: Jossey-Bass.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403–434.

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

Garet, M., Birman, B., Porter, A., Desimone, L., & Herman, R. with Suk Yoon, K. (1999). Designing effective professional development: Lessons from the Eisenhower Program. Washington, DC: U.S. Department of Education.

Hiebert, J. (1999). Relationships between research and the NCTM standards. *Journal for Research in Mathematics Education*, *30*(1), 3–19.

Kennedy, M. M. (1998). *Form and substance in in-service teacher education* (Research monograph no. 13). Arlington, VA: National Science Foundation.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont Portfolio Assessment Program: Findings and implications. *Issues and Practice*, *13*(3), 5–16.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis, 15*(2), 129–151.

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States.* Mahwah,NJ: Erlbaum.

Means, B. (Ed.). (1994). *Technology and education reform.* San Francisco: Jossey-Bass.

Means, B., Blando, J., Olson, K., Middleton, T., Morocco, C.C., Remz, A.R., and Zorfas, J. (1993). *Using technology to support education reform.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Means, B., & Olson, K. (1995). *Technology's role in education reform: Findings from a national study of innovating schools.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Mitchell, K. J. (1996). *Reforming and conforming: NASDC principals discuss school accountability systems.* Santa Monica, CA: RAND.

Mullens, J. (1998). *Validating teacher responses on instructional practice self-report surveys: Findings from a recent fieldtest.* Prepared for the Association for Public Policy Analysis and Management Research Conference, New York City, 1998.

Mullens, J.E., & Gayler, K. (1999). *Measuring classroom instructional processes: Using surveys and case study field test results to improve item construction.* Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement.

National Assessment Governing Board. (1997). *Science framework for the 1996 National Assessment of Educational Progress. A report of the Council of Chief State School Officers with the National Center for Improving Science Education and the American Institutes for Research.* Washington, DC: Author.

National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: Government Printing Office.

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

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1998). *Principles and standards for school mathematics: Discussion draft*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Science Teachers Association. (1992). *Volume 1: The content core, a guide for curriculum designers*. Washington, DC: Author.

Porter, A. C., Kirst, M. W., Osthoff, E. J., Smithson, J. L., & Schneider, S. A. (1993, October). *Reform up close: An analysis of high school mathematics and science classrooms*. Madison, WI: University of Wisconsin-Madison.

Raizen, S. A. (1998). Standards for science education. *Teachers College Record*, *100*(1), 66–121.

Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states: Findings from the National Assessment of Educational Progress.* Washington, DC: U.S. Department of Education.

Romberg, T.A. (1998). Comments: NCTM's curriculum and evaluation standards. *Teachers College Record*, 100(1), 8-21.

Richardson, V., & Placier, P. (in press). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.). New York: Macmillan.

Roitman, J. (1998). A mathematician looks at national standards. *Teachers College Record*, *100*(1), 22–44.

Rowan, B. (2000). *Assessing Teacher Quality: Insights from School Effectiveness Research*. Paper prepared for the U.S. Department of Education, Planning and Evaluation Service.

Shields, P. M., Marsh, J. A., & Adelman, N. E. (1998). *Evaluation of NSF's Statewide Systemic Initiatives (SSI) Program: The SSIs' impacts on classroom practice*. Menlo Park, CA: SRI.

Sivin-Kachala, J., & Bialo, E. (1996). *Report on the effectiveness of technology in schools, '95–'96*. Washington, DC: Software Publishers Association.

Smith, M.S., & O'Day, J. (1991). Systemic school reform. In S.H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing: The 1990 yearbook of the Politics of Education Association* (pp. 233–267). Bristol, PA: Falmer Press.

Snyder, T.D., Hoffman, C.M., & Geddes, C.M. (1999). *Digest of education statistics 1998*. Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics. (1999a). *Teacher quality: A report on the preparation and qualifications of public school teachers* (NCES 1999–080). Washington, DC: Author.

U.S. Department of Education, Office of the Under Secretary, Planning and Evaluation Service. (1999b). *Federal legislation enacted in 1994: An evaluation of implementation and impact.* Washington, DC: Author.

U.S. Department of Education. (1999c). *FY 2000 Annual plan: Volume 2. Program performance plans*. Washington, DC: Author.

Weiss, I. R., Montgomery, D. L., Ridgway, C. J., & Bond, S. L. (1998). *Local systemic change through teacher enhancement: Year three cross-site report.* Chapel Hill, NC: Horizon Research, Inc.

Wilson, S. M., & Lowenberg, D. (1991). *Changing visions and changing practices: Patchworks in learning to teach mathematics for understanding*. Research Report 91-2. East Lansing, MI: The National Center for Research on Teacher Education.

APPENDIX A

STUDENT ACHIEVEMENT TRENDS IN THE 30 SCHOOLS

We compiled the achievement data to examine the extent to which schools with unusually strong professional development exhibited positive trends in achievement. But the data proved to be inconclusive, for two reasons. First, as discussed in Section II, we found few school differences in the quality of professional development. Second, there were very few patterns of any note in the trends in achievement across schools. Exhibit A.1 displays the available achievement data for each of the schools in our study.

EXHIBIT A.1

Student Achievement Trends in the 30 Schools

State	District	School	Grade Level	1996	1997	1998	1999	Metric
KY	Boonetown	E	3	-	48.4	57.3	-	CTBS/5 Mean Score (range = 1–99)
			6	-	57.6	55.2	-	CTBS/5 Mean Score (range = 1–99)
			4/5	-	-	-	65.0	Kentucky Core Content Test 1998–1999 Academic Indexes
		J	7/8	-	-	-	63.7	Kentucky Core Content Test 1998–1999 Academic Indexes
		H	10–12	-	-	-	64.3	Kentucky Core Content Test 1998–1999 Academic Indexes
	Weller	E1	3	-	52.9	46.2	-	CTBS/5 Mean Score (range = 1–99)
			6	-	40.4	47.4	-	CTBS/5 Mean Score (range = 1–99)
			4/5	-	-	-	60.7	Kentucky Core Content Test 1998–1999 Academic Indexes
			7/8	-	-	-	58.6	Kentucky Core Content Test 1998–1999 Academic Indexes
		E2	3	-	53.5	59.0	-	CTBS/5 Mean Score (range = 1–99)
			6	-	51.9	42.2	-	CTBS/5 Mean Score (range = 1–99)
			4/5	-	-	-	55.4	Kentucky Core Content Test 1998–1999 Academic Indexes
			7/8	-	-	-	60.9	Kentucky Core Content Test 1998–1999 Academic Indexes
		H	10–12	-	-	-	58.1	Kentucky Core Content Test 1998–1999 Academic Indexes

E=elementary school, M=middle school, J=junior high school, H=high school
“-”=data not available

EXHIBIT A.1 (Continued)

Student Achievement Trends in the 30 Schools

State	District	School	Grade Level	1996	1997	1998	1999	Metric
NY	East City	E1	3	92.5	90.8	90.8	-	Percentage At/Above State Minimum Level
			6	-	-	38.0	-	Percentage At/Above State Minimum Level
		E2	6	-	-	-	-	Percentage At/Above State Minimum Level
		J	6	126	161	159	-	Percentage At/Above State Minimum Level
			8 accel.	-	64.8	96.8	-	Percent of tested passing Regents Test for Sequential Math I Course
	Richmond	E	3	100.0	100.0	97.3	-	Percentage At/Above State Minimum Level
			6	99	98.8	100.0	-	Percentage At/Above State Minimum Level
		M	-	0	0	0	-	Percentage of Average Grade Enrollment (AGE) Passing Regents Math Course I
		H	9–12	34.0	26.0	26.0	-	Percentage of Average Grade Enrollment (AGE) Passing Regents Math Course I
			9–12	25.0	27.0	24.0	-	Percentage of Average Grade Enrollment (AGE) Passing Regents Math Course III
			-	-	-	-	-	
OH	Maple City	E	4	-	20.3	18.2	7.6	Passing percentage (Ohio Proficiency Tests)
		M	6	-	17.9	10.4	6.7	Passing percentage (Ohio Proficiency Tests)
		H	12	-	42.7	50.5	45.1	Passing percentage (Ohio Proficiency Tests)
	Buckeye	E	4	-	79.7	75.3	75.4	Passing percentage (Ohio Proficiency Tests)
		M	9	-		28.0	18.0	Passing percentage (Ohio Proficiency Tests)
		H	12	-	68.9	66.6	66.1	Passing percentage (Ohio Proficiency Tests)
TX	Lone Star	E	4	-	95.0	89.7	-	Percent passing (TAAS)
		M	8	-	64.3	64.7	-	Percent passing (TAAS)
		H	10	-	96.8	99.4	-	Percent passing (TAAS)
	Rhinestone	E	4	-	85.1	89.7	-	Percent passing (TAAS)
		M	8	-	65.2	73.5	-	Percent passing (TAAS)
		H	10	-	71.5	81.3	-	Percent passing (TAAS)
WA	Rainforest	E	4	-	-	21.7	12.8	Percent students meeting standard (WA Assessment of Student Learning)
		M	7	-	-	7.3	16.2	Percent students meeting standard (WA Assessment of Student Learning)
		H	10	-	-	-	25.2	Percent students meeting standard (WA Assessment of Student Learning)
	Riverside	E	4	-	38.9	36.2	58.1	Percent students meeting standard (WA Assessment of Student Learning)
		J	7	-	-	29.9	30.9	Percent students meeting standard (WA Assessment of Student Learning)
		H	10	-	-	-	50.4	Percent students meeting standard (WA Assessment of Student Learning)

E=elementary school, M=middle school, J=junior high school, H=high school

APPENDIX B

SAMPLE DESIGN AND RESPONSE RATES FOR THE LONGITUDINAL TEACHER SURVEY

In the *Longitudinal Study of Teacher Change*, we collected data from teachers in 30 schools—three schools in each of our 10 in-depth case districts. We surveyed teachers at three time points: the fall of 1997, the spring of 1998, and the spring of 1999.

In this appendix, we describe the sample design for the three waves of the longitudinal survey, we summarize the response rates for the three waves of the survey, we describe the demographic characteristics of responding teachers, and we discuss the mathematics and science courses on which the teachers reported in the classroom practice section of the survey.

Sample Design for the Three Waves of the Survey

Sample Design: Schools

In selecting the overall sample of 30 schools, we balanced several objectives. First, we sought schools in which teachers were likely to participate in Eisenhower-assisted activities over the 1997–98 year, the year in which we conducted classroom observations, teacher interviews, and focus groups in the case schools. In addition, we desired a mix of elementary, middle, and high schools. Finally, we sought schools that varied in demographic composition, including percent of students in poverty, as measured by eligibility for free lunch, and percent minority. In particular, we planned to oversample high-poverty schools, those with more than 50 percent of students eligible for free lunch, because the Eisenhower program intends to encourage the participation of teachers in such schools.

To achieve these objectives, we asked the Eisenhower coordinators in each case district to help us identify schools that met our criteria and that were willing to participate in our site visits, longitudinal survey, and classroom observations. On the basis of information provided by the coordinators as well as demographic information from the Common Core of Data, we selected one elementary, one middle, and one high school in each in-depth case district.[51]

Of the 30 schools in our final sample, 17 are high poverty and 12 have more than 50 percent minority students. In the fall of 1997, student enrollment in the schools ranged from 247 students in one rural elementary school to 1,554 students in a suburban high school, with an average size of 818 students.

[51] One district, East City, includes only elementary and middle schools, so we chose two elementary and one middle. In a second district, Weller, two K–8 schools were selected rather than an elementary and a middle school.

Sample Design: Teachers

Baseline wave. We conducted the baseline wave of the data collection (which asked teachers about their classroom instruction during the 1996–97 school year) in the fall of 1997. We defined the population for the first wave of the survey to include all teachers who had taught mathematics or science in the 30 case schools in 1996–97 and were still teaching in the same schools in 1997–98. We did not attempt to locate teachers who taught mathematics or science in the sample schools in 1996–97 but were no longer teaching in the same school in 1997–98.

To identify teachers meeting our criteria, we asked the principals in each of the 30 schools to provide a roster of all teachers, indicating whether each teacher taught mathematics, science, or both. Altogether, 575 teachers met the criteria to be included in the baseline wave of the survey.

In most cases, the elementary schools in our sample organized instruction in self-contained classrooms. Thus, most elementary teachers in the schools were general elementary teachers who taught all subjects, including both science and mathematics. In schools with self-contained classrooms, we included all elementary teachers in our sample. We randomly selected half of the elementary teachers to receive a mathematics version of the survey and half to receive a science version. Some elementary schools had specialist teachers who taught only mathematics or science, and we included these teachers in our sample.

Almost all of the middle and high schools in our sample were departmentalized, with teachers who were subject-matter specialists. We included all teachers identified as mathematics or science teachers in departmentalized schools, along with any other teachers the schools identified as teaching some mathematics or science, such as the special education teacher.

Second wave. We conducted the second wave of the survey (which asked teachers about their classroom instruction and professional development experiences during the 1997–98 school year) in the spring of 1998. We administered the survey to the same 575 teachers included in the first wave of the survey.

Third wave. We conducted the third wave of the survey (which asked teachers about their classroom instruction and professional development experiences during the 1998–99 school year) in the spring of 1999. We defined the population for the third wave to include all teachers who taught mathematics or science in the 30 case schools during the 1998–99 school year, including teachers who were new to the schools that year. We did not survey teachers who had taught in the sample schools in 1997–98 but had left prior to the spring of 1999. Altogether, 604 teachers met the criteria to be included in the third wave of the survey. Of these, 482 teachers had been included in the sample for the first and second waves of the survey, and 122 were teachers who were new to the case schools. Of the 575 teachers who were included in the first and second wave surveys, 93 had left the case schools prior to the spring of 1999. This represents a turnover rate of about 16 percent.

Response Rates for Teachers

Response rates for each wave of the survey are shown in Exhibits B.1, B.2, and B.3. The response rate for the first wave was 75 percent; for the second wave, it was 74 percent; and for the final wave, 75 percent. There were differences in response rates by school; in many schools, response rates were nearly 100 percent in each wave, and in others they were below 70 percent.

EXHIBIT B.1

Response Rates for the Baseline Wave of the Longitudinal Teacher Survey, Fall 1997

	Mathematics			Science			
	Elementary School	Middle School	High School	Elementary School	Middle School	High School	TOTAL
A. Number of teachers sent surveys	137	65	89	133	64	87	575
B. Number of surveys returned	91	43	73	101	48	74	430
C. Survey response rate (row B divided by row A)	66%	66%	82%	76%	75%	85%	75%
D. Number of responding teachers who did not teach mathematics or science during 1996–97 year	12	3	0	16	5	6	42
E. Number of responding teachers who taught mathematics or science during 1996–97 year (row B minus row D)	79	40	73	85	43	68	388
F. Number of teachers who completed survey but provided inadequate data for analysis of content taught	5	2	4	16	2	4	33
G. Percent of teachers completing content items (row E minus row F divided by row E)	94%	95%	95%	81%	95%	94%	91%
H. Number of teachers who completed survey but provided inadequate data for analysis of pedagogy	2	0	1	1	0	1	5
I. Percent of teachers completing pedagogy items (row E minus row H divided by row E)	97%	100%	99%	99%	100%	99%	99%
J. Percent of teachers completing both content and pedagogy items	91%	95%	93%	80%	95%	93%	90%

EXHIBIT B.2

Response Rates for the Second Wave of the Longitudinal Teacher Survey, Spring 1998

	Mathematics			Science			
	Elementary School	Middle School	High School	Elementary School	Middle School	High School	TOTAL
A. Number of teachers sent surveys	143	62	86	139	62	83	575
B. Number of surveys returned	103	44	66	102	51	62	428
C. Survey response rate (row B divided by row A)	72%	71%	77%	73%	82%	75%	74%
D. Number of responding teachers who did not teach mathematics or science during 1997–98 year	6	0	0	18	1	0	25
E. Number of responding teachers who taught mathematics or science during 1997–98 year (row B minus row D)	97	44	66	84	50	62	403
F. Number of teachers who completed survey but provided inadequate data for analysis of content taught	16	2	8	37	8	4	75
G. Percent of teachers completing content items (row E minus row F divided by row E)	84%	95%	88%	56%	84%	94%	81%
H. Number of teachers who completed survey but provided inadequate data for analysis of pedagogy	6	0	1	18	1	1	27
I. Percent of teachers completing pedagogy items (row E minus row H divided by row E)	94%	100%	98%	79%	98%	98%	93%
J. Percent of teachers completing both content and pedagogy items	90%	95%	86%	77%	86%	92%	87%

EXHIBIT B.3

Response Rates for the Third Wave of the Longitudinal Teacher Survey, Spring 1999

	Mathematics			Science			TOTAL
	Elementary School	Middle School	High School	Elementary School	Middle School	High School	
A. Number of teachers sent surveys	149	67	89	150	67	82	604
B. Number of surveys returned	115	48	76	114	42	57	452
C. Survey response rate (row B divided by row A)	77%	72%	85%	76%	63%	70%	75%
D. Number of responding teachers who did not teach mathematics or science during 1998–99 year	5	1	0	20	0	1	27
E. Number of responding teachers who taught mathematics or science during 1998–99 year (row B minus row D)	110	47	76	94	42	56	425
F. Number of teachers who completed survey but provided inadequate data for analysis of content taught	12	4	11	39	2	7	75
G. Percent of teachers completing content items (row E minus row F divided by row E)	89%	91%	86%	59%	95%	88%	82%
H. Number of teachers who completed survey but provided inadequate data for analysis of pedagogy	5	2	0	20	1	2	30
I. Percent of teachers completing pedagogy items (row E minus row H divided by row E)	95%	96%	100%	79%	98%	96%	93%
J. Percent of teachers completing both content and pedagogy items	94%	91%	86%	80%	93%	89%	88%

As indicated above, 482 teachers taught in the sample schools for all three years, and thus these teachers were included in the intended sample for all three waves of the survey. Of these 482 teachers, 287 (60 percent) responded to all three waves; 95 (20 percent) responded to two waves; 60 (12 percent) responded to one wave; and 40 (8 percent) did not respond to any of the three waves.

In each wave, some of the responding teachers reported that they did not teach mathematics or science in the survey year. Teachers who did not teach mathematics or science were instructed to skip the subject-specific sections of their surveys. (See Exhibits B.1, B.2, and B.3 for the number of such teachers for each wave of the survey.)

Of the teachers who taught mathematics or science during the survey years, some did not complete a sufficiently high proportion of the classroom content items on the survey for us to include their responses in the analyses of content, and some did not complete the necessary items on pedagogy.[52] For example, in the base wave, 9 percent of the teachers who taught mathematics or science in 1996–97 failed to provide complete responses on the content items, and 1 percent did not provide complete responses to the pedagogy items. (See Exhibit B.1.) A somewhat higher percent of responding teachers failed to provide data on the content items in the second and third waves. (See Exhibits B.2 and B.3.)

Characteristics of Respondents

The teachers who responded to the survey are fairly representative of the general teaching population (see Exhibit B.4). Nationally, 73 percent of teachers are female (Snyder et al., 1999); in the respondents to the baseline survey, 74 percent are female. Fourteen percent of all teachers are minorities (Snyder et al., 1999), and 18 percent of the baseline sample are minorities. Almost 10 percent of all teachers have less than three years of teaching experience (Snyder et al., 1999); in the baseline sample, 9 percent have less than three years of teaching experience in total, and 11 percent of teachers have less than three years of teaching experience in the surveyed subject.

We examined potential differences, in demographic background, classroom instructional practices, and professional development experiences, between the teachers who responded to all three waves of the survey and the full sample who responded to the baseline wave, and we found almost no significant differences. The only significant difference we identified is that teachers in high-poverty schools were somewhat less well represented in the sample responding in all three waves than in the baseline sample.

[52] Almost half the teachers who did not fully complete the content items were elementary teachers who received the science version of the survey. It is possible that these teachers did not believe they taught science with sufficient frequency or with sufficient depth to complete the content section.

EXHIBIT B.4

Demographic Characteristics of Respondents to the Baseline Wave of Longitudinal Teacher Survey, Fall 1997 Percent (Number[53])

	Mathematics			Science			TOTAL
	Elementary School	Middle School	High School	Elementary School	Middle School	High School	
Gender							
Female	95 (70)	77 (27)	57 (39)	96 (65)	65 (26)	51 (32)	74 (259)
Male	5 (4)	23 (8)	43 (29)	4 (3)	35 (14)	49 (31)	26 (89)
Total	100 (74)	100 (35)	100 (68)	100 (68)	100 (40)	100 (63)	100 (348)
Ethnicity/Race							
Asian or Pacific Islander	-	-	2 (1)	-	-	2 (1)	1 (2)
African American	8 (6)	9 (3)	3 (2)	3 (2)	13 (5)	10 (6)	7 (24)
White	75 (54)	77 (27)	93 (63)	78 (53)	83 (33)	83 (52)	82 (282)
Hispanic	17 (12)	14 (5)	2 (1)	16 (11)	5 (2)	5 (3)	10 (34)
Other	-	-	2 (1)	3 (2)	-	2 (1)	1 (4)
Total	100 (72)	100 (35)	100 (68)	100 (68)	100 (40)	100 (63)	100 (346)
Novice Teachers*							
Across subjects	13 (9)	7 (2)	3 (2)	11 (7)	6 (2)	12 (7)	9 (29)
In surveyed subject	11 (6)	10 (3)	5 (3)	16 (8)	12 (4)	14 (8)	11 (32)

* Three or fewer years of teaching.

Sample of Courses Described in the Longitudinal Teacher Survey

In the baseline wave of the longitudinal teacher survey, we asked teachers to select a mathematics or science course to describe, choosing, if possible, a yearlong course they had taught in 1996–97, were continuing to teach in 1997–98, and expected to teach in 1998–99. If there were several courses they could choose, teachers were asked to select a course that enrolled students performing at mixed achievement levels. In the second wave, we asked teachers to describe a course they were teaching in 1997–98, choosing the same course taught in 1996–97 if possible. In third wave, we asked teachers to describe a course taught during the 1997–98 year, choosing, if possible, the same course taught in 1997–98.

In general, teachers described yearlong, mixed achievement courses. (See Exhibit B.3 for a description of the courses that teachers described in the baseline wave.). Although most teachers described yearlong courses, some teachers, especially science and high school teachers, focused on semester courses. Most teachers chose courses enrolling students of mixed achievement levels, but some teachers, especially mathematics teachers and high school teachers, described classes enrolling students of homogeneous low or high achievement.

[53] Because some teachers did not complete demographic information on the surveys, the numbers of teachers reported here are slightly lower than the numbers reported in the analyses.

EXHIBIT B.5

Sample of Courses Described in the Baseline Longitudinal Teacher Survey, Fall 1997 (n=355*)

	Mathematics			Science			
	Elementary School	Middle School	High School	Elementary School	Middle School	High School	TOTAL
Duration							
Year	92%	97%	74%	73%	87%	73%	81%
Semester	6%	0%	26%	15%	3%	24%	14%
Other	2%	3%	0%	12%	10%	3%	5%
Total	100%	100%	100%	100%	100%	100%	100%
Composition							
Homogeneous high	1%	11%	16%	0%	2%	14%	7%
Homogeneous middle	9%	18%	32%	3%	15%	23%	17%
Homogeneous low	9%	16%	17%	6%	%	8%	10%
Mixed	80%	53%	35%	91%	78%	55%	66%
Total	100%	100%	100%	100%	100%	100%	100%

* In some cases, the n is slightly lower because of missing data.

We asked teachers to report the title of the course they described, as well as the typical grade levels of students enrolled. Not surprisingly, elementary school teachers generally described mathematics and science instruction in self-contained classrooms enrolling students at one or two grade levels (e.g., "4thgrade mathematics"). Most middle school teachers described courses titled "7thgrade mathematics" or "8thgrade mathematics," but 10 of the middle school mathematics teachers and eight of the middle school science teachers described courses with more specific titles indicating the focus of the course (i.e., pre-algebra and algebra for mathematics; life, earth, and physical science).

Most of the high school teachers described courses with specific titles. Of the high school mathematics teachers surveyed, 32 described algebra courses; 13 described geometry courses; and a few each described calculus, integrated math, or trigonometry. Of the high school science teachers, 18 described biology courses (with a few honors biology); 13 described chemistry courses; 11 described physics courses; and one to three teachers described earth science, physical science, or astronomy. Four of the science courses and five of the mathematics courses were honors, advanced, or advanced placement courses.

As a group, the courses that teachers described in the second and third waves of the survey were similar to the courses described in the first wave. For some analyses, we restricted the sample to teachers who completed all three waves of the survey and reported on a course with the same title in all three waves. There were 207 such teachers. We compared the demographic background of teachers who reported on the same course in each wave with the full sample who responded in the baseline wave and found no differences.

APPENDIX C

ANALYSIS METHODS AND RESULTS FOR VARIATION IN TEACHERS' PROFESSIONAL DEVELOPMENT EXPERIENCES

In Section II, we summarized the features of professional development activities in which teachers participated in 1997–98 and 1998–99. In this appendix, we describe the methods we used for these analyses in more detail. The appendix includes three sections. First, we provide a brief overview of the approach we took to analyze teachers' professional development experiences; second, we describe the results we obtained; and third, we discuss the methods we used to derive the percent of variation estimates shown in Exhibit 5 of Section II.

Overview of Analysis Strategy

To characterize teachers' experiences in professional development, we examined three structural features of the professional development activities in which teachers participated—reform versus traditional type, duration (time span and hours), and collective participation. We also examined three core features—content focus, active learning, and coherence—as well as the degree to which teachers reported that the professional development in which they participated enhanced their knowledge and skills. Finally, we assessed the proportion of teachers who participated in professional development in mathematics or science in 1997–98 and 1998–99

We collected data on professional development experiences from teachers who taught in 30 schools—three schools in each of 10 districts. (See Sections I and II.) We asked teachers about two professional development experiences—one in 1997–98 and the other in 1998–99. Thus, the data we collected have a multilevel structure, with activities nested within teachers, teachers nested within schools, and schools nested within districts.

We drew on these data to examine differences in teachers' professional development experiences by school level and subject, as well as by district, school, teacher, year, and Eisenhower sponsorship. We conducted these analyses within a hierarchical linear modeling (HLM) framework. We treated the school, the teacher, and the professional development activity as levels in a three-level model, with activities nested within teachers within schools. This approach treats schools, teachers, and activities as "random effects," and we used our data to estimate the variance across schools, teachers, and activities in teachers' professional development experiences. (See Exhibit C.1 for model equations.)

We treated the 10 districts, the school level (elementary, middle, and high), the subject (mathematics and science), and the year (1997–98 and 1998–99) as fixed effects—that is, as conventional dummy-coded independent variables. We treated district and school levels as school-level variables, subject as a teacher-level variable, and year and sponsorship as activity-level variables.[54] (See Exhibit C.1.)

[54] We initially included all possible interactions of school level, subject, and year, but, with one exception, none were significant for any of our dependent variables, so we dropped them from the final models. The one exception

EXHIBIT C.1

Models for Features of Professional Development

The dependent variable, y_{tij}, is a feature of professional development (e.g., time span) for an activity that took place at time *t*, for teacher *i*, in school *j.*

Model 1 (Three-level unconditional hierarchical model):

Level 1 (activity): $y_{tij} = p_{0ij} + e_{tij}$.

Level 2 (teacher): $p_{0ij} = b_{00j} + n_{0ij}$.

Level 3 (school): $b_{00j} = g_{000} + m_{00j}$.

Model 2 (Three-level hierarchical model, conditional on school level and subject):

Level 1 (activity): $y_{tij} = p_{0ij} + e_{tij}$.

Level 2 (teacher): $p_{0ij} = b_{00j} + b_{01}s_{ij} + n_{0ij}$, where $s_{ij} =$ subject is coded math=1, science=0.

Level 3 (school): $b_{00j} = g_{000} + g_{001}e_j + g_{002}h_j + m_{00j}$, where $e_j =$ elementary and $h_j =$ high school are coded 0/1 and middle is excluded.

Model 3 (Three-level hierarchical model, conditional on school level, subject, and time):

Level 1 (activity): $y_{tij} = p_{0ij} + p_1 time_{tij} + e_{tij}$, where $time_{tij}$ is coded 0 for activities in 1997–98 and 1 in 1998—99.

Level 2 (teacher): $p_{0ij} = b_{00j} + b_{01}s_{ij} + n_{0ij}$, where $s_{ij} =$ subject is coded math=0, science=1.

Level 3 (school): $b_{00j} = g_{000} + g_{001}e_j + g_{002}h_j + m_{00j}$, where $e_j =$ elementary and $h_j =$ high school are coded 0/1 and middle is excluded.

Model 4 (Three-level hierarchical model, conditional on school level, subject, time, and district):

Level 1 (activity): $y_{tij} = p_{0ij} + p_1 time_{tij} + e_{tij}$, where $time_{tij}$ is coded 0 for activities in 1997–98 and 1 in 1998—99.

Level 2 (teacher): $p_{0ij} = b_{00j} + b_{01}s_{ij} + n_{0ij}$, where $s_{ij} =$ subject is coded math=0, science=1.

Level 3 (school): $b_{00j} = g_{000} + g_{001}e_j + g_{002}h_j + g_{003}districtA_j + ... + g_{00,10}districtI_j + m_{00j}$, where $e_j =$ elementary and $h_j =$ high school are coded 0/1 and middle is excluded; and $districtA_j ... districtI_j$ are indicator variables representing the district in which school *j* is located.

EXHIBIT C.1 (Continued)

Models for Features of Professional Development

Model 5 (Three-level hierarchical model, conditional on school level, subject, time, district, and sponsorship):

Level 1 (activity): $y_{tij} = p_{0ij} + p_1 time_{tij} + p_2 sponsor_{tij} + e_{tij}$, where $time_{tij}$ is coded 0 for activities in 1997–98,and 1 in 1998–99; and $sponsor_{tij}$ is coded 0 for 1997–98 activities that are Eisenhower-funded and 0 otherwise.

Level 2 (teacher): $p_{0ij} = b_{00j} + b_{01} s_{ij} + n_{0ij}$, where $s_{ij} =$ subject is coded math=0, science=1.

Level 3 (school): $b_{00j} = g_{000} + g_{001} e_j + g_{002} h_j + g_{003} districtA_j + ... + g_{00,10} districtI_j + m_{00j}$, where $e_j =$ elementary and $h_j =$ high school are coded 0/1 and middle is excluded; and $districtA_j ... districtI_j$ are indicator variables representing the district in which school *j* is located.

We considered treating the district as a level in our HLM models, but the number of districts in our sample (10) is too small to provide a good estimate of the between-district variance component. Furthermore, the statistical methods we employed to estimate the hierarchical models assume that the random effects are normally distributed. Although this assumption is reasonable for schools, teachers, and activities, it seems less reasonable for districts, which were purposefully selected to include considerable variation in size, urbanicity, and district professional development policies.

For each dependent variable (reform vs. traditional type, time span, hours, collective participation, content focus, active learning, coherence, enhanced knowledge and skills, and participation in professional development), we estimated a set of five models. (See Exhibit C.1.) Model 1, a "fully unconditional" HLM model, includes no fixed effects. Thus, it estimates the variation across schools, teachers, and activities with no controls. In Model 2, we added school subject and level; in Model 3, we added year; in Model 4, we added district; and in Model 5, we added Eisenhower sponsorship.

Overview of Analysis Results

The results of our analyses are reported in Exhibits C.2 (reform type), C.3 (time span), C.4 (hours), C.5 (collective participation), C.6 (active learning), C.7 (content focus), C.8 (coherence), C.9 (enhanced knowledge and skills), and C.10 (participation).

To illustrate the meaning of the parameter estimates, we discuss the results for content focus, displayed in Exhibit C.7. (The results displayed in the other exhibits are similar in form.) Content focus is measured by a single item, scaled 0=no emphasis on mathematics or science content; 1=minor emphasis; 2=major emphasis.

The fixed-effect parameter estimates are displayed in three groups, containing the school-level, teacher-level, and activity-level parameters. The results for Model 1 are displayed in the first column of the exhibit. Model 1 contains only one fixed effect, the intercept, which has a value of 1.098. The Model 1 intercept represents the value of content focus for the typical school in our sample.[55] Thus, on average, teachers in our sample reported that the professional development in which they participated placed a "minor emphasis" on content.

The variance components shown at the bottom of the exhibit describe the between-school, between-teacher within schools, and residual variation (variation between activities within schools). The between-school variance component of 0.052 indicates that there is significant between-school variation in content focus ($p<.05$). One way to interpret the magnitude of the between-school variance is to compute its square root, which is the between-school standard deviation. The square root of 0.052 is about 0.23, which indicates that teachers in some schools have an average content focus about 0.23 point above the overall mean of 1.098, whereas teachers in other schools have an average content focus 0.23 point below the overall mean.

The variation among teachers within schools (0.180) is larger than the between-school variance component. The square root of the variance is 0.42, which indicates that teachers within the same school may experience professional development for which the content focus is 0.42 point above or below the school mean.

The residual variation—or variation among the different activities attended by the same teacher—is 0.375. This indicates that the two activities each teacher in our sample reported on differed substantially in content focus. (The square root of the variance is 0.61.)

The results for Model 2 are shown in the second column of the exhibit. The intercept of 1.024 is the average or typical value for schools in our sample, controlling for school level and subject taught. The coefficient for elementary school (0.286) is statistically significant, which indicates that elementary teachers report higher levels of content focus than middle school teachers (the reference category) do. We also tested the joint significance of the set of coefficients for school level (the coefficients for elementary and high school), and the results displayed below the table indicate that school level is significant ($p<.01$). The coefficient for the teacher's subject taught is very small (-0.017), which indicates that there is little difference between mathematics and science teachers in the content focus of professional development.

The variance components for Model 2 are shown at the bottom of the exhibit. The variance component for school (0.025) is smaller than the between-school variance in Model 1 and is no longer significant. This indicates that much of the between-school variance observed in Model 1 is due to school level. Otherwise, the variance components are unchanged. In particular, including the subject (mathematics and science) has not reduced the between-teacher variance.

The results for Model 3 are shown in the third column of the exhibit. The coefficient for year (-0.014) is very close to zero, which indicates that there is no difference between the average content focus that teachers experienced in 1997–98 and 1998–99.

---

[55] As described above, we treated school as a random effect. Thus, the intercept represents the average or typical value of content focus among schools in our sample; the variance components shown at the bottom of the exhibit indicate the variation among schools around this value.

The results for Model 4 are shown in the fourth column. The coefficients for many of the individual districts are significant, which indicates that the average content focus in these districts differs from the content focus in the 10th district (the reference category). For example, District D, with a coefficient of 0.827, has a substantially higher than average content focus. We tested the joint significance of the set of district coefficients, and the results indicate a significant district effect ($p<.10$). The between-school variance component for Model 4 is zero, which indicates that there is no remaining between-school variance when school level and district are included in the model.

The results for Model 5 are shown in the final column. The coefficient for Eisenhower sponsorship is very small (0.012), which indicates that there is no difference in content focus between Eisenhower-supported and non-Eisenhower-supported activities when all other variables in the model are controlled.[56]

Methods Used in Section II to Compute Percent Variation

In Exhibit 5 in Section II, we report the percent of variation in the quality of professional development explained by differences across subjects and school levels, schools, teachers in the same school, and years, as well as residual unexplained variance. These variance estimates are derived from the results for Models 1, 2, and 3. The total variance to be explained is computed as the total variance for Model 1—that is, the sum of the variance between schools, the variance between teachers, and the residual variance for Model 1 ($m_{00j} + n_{0ij} + e_{tij}$). The variance from subject and level is computed as the difference between the total variance for Model 1 ($m_{00j} + n_{0ij} + e_{tij}$) and the total variance for Model 2 ($m_{00j} + n_{0ij} + e_{tij}$), which controls for subject and level.[57] The variance from years is computed as the difference between the total variance ($m_{00j} + n_{0ij} + e_{tij}$) for Model 2 and the total variance for Model 3, which controls for subject, level, and year (1997–98 and 1998–99). The variance from school and teacher and the residual unexplained variance are the variance component estimates for Model 3 (m_{00j}, n_{0ij}, and e_{tij}).[58,59,60]

---

[56] We had data on Eisenhower-sponsorship only for 1997–98. Thus, we coded Eisenhower sponsorship 1 for 1997–98 activities that were Eisenhower sponsored; 0 for 1997–98 activities that were not Eisenhower sponsored; and 0 for 1998–99 activities. When both year and Eisenhower sponsorship are included in the model, Eisenhower-sponsored represents the difference between Eisenhower-sponsored and non-sponsored activities in 1997–98, and year represents the difference between *all* activities in 1998–99 and non-Eisenhower-sponsored activities in 1997–98.

[57] This is analogous to the conventional percent of variation explained in an OLS regression model.

[58] The variance explained by the fixed effects is a function of the order in which the variance estimates are derived. We computed the variance due to school level and subject prior to computing the variance due to year; if we had chosen the reverse order, the variance due to year might be slightly higher and the variance due to school level and subject, slightly lower.

[59] Because we treated district as a fixed effect, we did not obtain direct estimates of the between-district variance component. It would be possible to derive an estimate of the percent of variance between districts from the variance for Models 3 and 4. We elected not to report these estimates, however, because, as we discussed above, the between-district variance is not very meaningful or reliable, given the small number of districts in the sample.

[60] We expected that the variance between schools and teachers and the residual variance would decline from Model 1 to Model 5, as additional control variables were added to the model. In a few cases, however, the variance increased with the addition of controls. This increase tended to occur when none of the added controls was

To illustrate these computations, we turn again to content focus (Exhibit C.7). The total variance in content focus to be explained is the sum of the three variance components for Model 1 (0.05 + 0.18 + 0.37 = 0.60).[61] The variance explained by school and level is the difference between 0.60 and the total variance for Model 2 (0.02 + 0.18 + 0.37 = 0.57), or 0.03. In percentage terms, the variance explained by school level and subject is 0.03 divided by 0.60, or 5 percent. Similarly, the variance explained by year is the difference between the total variance for Model 2 (0.57) and the total variance for Model 3 (0.02 + 0.18 + 0.37) = 0.57, or 0.

Finally, the percent variance explained by schools is the variance component for schools in Model 3, divided by the total variance for Model 1 (0.02 divided by 0.06), or 3.3 percent. The percent variance explained by teachers is the variance component for teachers in Model 3, divided by the total variance for Model 1 (0.18 divided by 0.60), or 30 percent. The percent variance unexplained is the residual variance in Model 3 divided by the total variance for Model 1 (0.37 divided by 0.60), or 61.7 percent.

significant, suggesting that any added explanatory power because due to the controls was more than offset by the loss of degrees of freedom. Given the method we used to compute the variance explained by fixed effects, an increase in variance due to the addition of fixed effects would result in a negative variance explained. We treated the few negative estimates as zero.

[61] The calculations of the percent variance were carried out using one less significant digit than appears in the Exhibits. If the parameters shown in the Exhibit are rounded to two digits, the results may on occasion differ slightly from those used in the calculations because the parameters in the exhibit have been rounded to three digits.

EXHIBIT C.2

Reform Type: Effects of Level, Subject, Year, District, and Sponsorship

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District	Model 5: Sponsor-ship
School-level					
Intercept	0.169***	0.195***	0.180**	0.148+	0.193*
Elementary		-.019	-.019	-.033	-.032
High school		-.043	-.043	-.017	-.016
District A				-.037	-.063
District B				.021	-.005
District C				.056	.045
District D				.198*	.172+
District E				.021	-.005
District F				-.048	-.064
District G				.241*	.211*
District H				-.057	-.071
District I				.036	.011
Teacher-level					
Math		-.003	-.003	.010	-.011
Activity-level					
Year (1998–99)			.031	.030	-.003
Eisenhower-sponsored					-.086
Variance components					
Between-school	.001	.001	.001	.000	.000
Between-teacher	.009	.010	.010	.009	.010
Residual	.131	.131	.131	.130	.130

+ p<.10, * p<.05, ** p<.01, *** p<.001
N=401

EXHIBIT C.3

Span: Effects of Level, Subject, Year, District, and Sponsorship

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District	Model 5: Sponsor-ship
School-level					
Intercept	4.028***	4.496***	4.662***	4.818***	4.98***
Elementary		-.218	-.215	-.087	-.090
High school		-.239	-.238	-.110	-.094
District A				-.726	-.831
District B				-.894	-.937
District C				-.316	-.395
District D				-.698	-.799
District E				-.140	-.250
District F				.064	-.009
District G				-.668	-.776
District H				-.850	-.909
District I				.458	.367
Teacher-level					
Math		-.521*	-.518*	-.377	-.383
Activity-level					
Year (1998–99)			-.347+	-.342	-.433+
Eisenhower-sponsored					-.292
Variance components					
Between-school	.031	.000	.000	.121	.137
Between-teacher	1.055*	1.031**	1.060**	.975*	.963*
Residual	4.239	4.257	4.211	4.219	4.287

+ p<.10, * p<.05, ** p<.01, *** p<.001
N=390

EXHIBIT C.4

Hours: Effects of Level, Subject, Year, District, and Sponsorship

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District	Model 5: Sponsor-ship
School-level					
Intercept	19.89***	22.00***	23.48***	27.59***	26.45***
Elementary		-4.68	-4.70	-4.94	-4.81
High school		3.87	3.85	4.97	5.03
District A				-8.35	-7.60
District B				-14.36*	-13.47*
District C				-0.87	-0.53
District D				-6.01	-5.26
District E				-6.96	-6.22
District F				-0.14	0.39
District G				-1.18	-0.33
District H				-12.16*	-11.66*
District I				-1.14	-0.64
Teacher-level					
Math		-3.17	-3.13	-2.08	-2.21
Activity-level					
Year (1998–99)			-3.02	-3.11	-2.55
Eisenhower-sponsored					1.86
Variance components					
Between-school	30.46	13.53	13.46	3.67	3.69
Between-teacher	77.58*	79.53*	80.94*	80.21*	81.29*
Residual	417.44	417.56	414.98	415.14	420.10

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level; Model 2 (p<.05), Model 4 (p<.01), Model 5 (p<.01)
N=393

EXHIBIT C.5

Collective Participation: Effects of Level, Subject, Year, District, and Sponsorship

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District	Model 5: Sponsor-ship
School-level					
Intercept	0.366***	0.311**	0.322**	.367*	0.375+
Elementary		.076	.076	.045	.045
High school		-.047	-.047	-.021	-.017
District A				-.121	-.124
District B				-.020	-.013
District C				-.215	-.216
District D				.304	.311
District E				-.077	-.083
District F				.029	.024
District G				.075	.073
District H				-.274	-.278
District I				.093	.088
Teacher-level					
Math		.074	.074	.071	.064
Activity-level					
Year (1998–99)			-.024	-.023	-.026
Eisenhower-sponsored					.010
Variance components					
Between-school	.043*	.040*	.041*	.054*	.058+
Between-teacher	.032+	.033+	.033+	.029	.024
Residual	.220	.220	.221	.222	.228

+ p<.10, * p<.05, ** p<.01, *** p<.001
N=396

EXHIBIT C.6

Active Learning: Effects of Level, Subject, Year, District, and Sponsorship

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District	Model 5: Sponsor-ship
School-level					
Intercept	3.429***	3.268***	3.467***	3.323***	3.661***
Elementary		.122	.140	-.326	-.320
High school		.086	.102	.091	.095
District A				-1.485	-1.680
District B				.977	.915
District C				-.128	-.250
District D				4.043***	3.848**
District E				-.138	-.331
District F				-.069	-.194
District G				-.065	-.274
District H				-.122	-.220
District I				1.632	1.432
Teacher-level					
Math		.143	.138	.051	.016
Activity-level					
Year (1998–99)			-.436	-.415	-.610*
Eisenhower-sponsored					-.573
Variance components					
Between-school	.990	1.097	1.096	.000	.000
Between-teacher	5.999***	6.087	6.129***	6.051***	6.008***
Residual	6.139	6.139	6.072	6.076	6.155

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of district; Model 4 (p < .05), Model 5 (p < .05)
N=376

EXHIBIT C.7

Content Focus: Effects of Level, Subject, Year, District, and Sponsorship

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District	Model 5: Sponsor-ship
School-level					
Intercept	1.098***	1.024***	1.030***	0.707***	0.636**
Elementary		.286*	.286*	.226+	.265*
High school		-.103	-.103	-.132	-.104
District A				.418+	.463*
District B				.330+	.367+
District C				.284	.370*
District D				.827***	.856***
District E				.406*	.448*
District F				.414*	.450*
District G				.637**	.678**
District H				.335+	.370+
District I				.328	.372+
Teacher-level					
Math		-.018	-.017	-.047	-.042+
Activity-level					
Year (1998–99)			-.014	-.012	-.011
Eisenhower-sponsored					.012
Variance components					
Between-school	.052*	.025	.025	.000	.000
Between-teacher	.180***	.183***	.182***	.185***	.191***
Residual	.375	.374	.376	.376	.367

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect for level; Model 2 (p<.01), Model 3 (p<.01), Model 4 (p<.01), Model 5 (p<.01)
Significant effect for district; Model 4 (p<.10), Model 5 (p<.10)
N=394

EXHIBIT C.8

Coherence: Effects of Level, Subject, Year, District, and Sponsorship

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District	Model 5: Sponsor-ship
School-level					
Intercept	5.576***	5.997***	6.065***	6.056***	6.001***
Elementary		-.475	-.366	-.505	-.509
High school		-1.038**	-.914*	-1.058**	-1.038**
District A				.157	.194
District B				.695	.811
District C				-.305	-.286
District D				.769	.822
District E				.221	.255
District F				.053	.076
District G				-.287	-.237
District H				.044	.073
District I				.478	.524
Teacher-level					
Math		.234	.238	.196	.173
Activity-level					
Year (1998–99)			-.148	-.325	-.298+
Eisenhower-sponsored					.147
Variance components					
Between-school	.225	.110	.107	.247	.266
Between-teacher	1.757***	1.753***	1.775***	1.759***	1.692***
Residual	1.840	1.838	1.806	1.795	1.838

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect for level; Model 2 (p<.05), Model 3 (p<.10), Model 4 (p<.05), Model 5 (p<.05)
N=374

EXHIBIT C.9

Enhanced Knowledge and Skills: Effects of Level, Subject, Year, District, and Sponsorship

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District	Model 5: Sponsor-ship
School-level					
Intercept	2.939***	2.946***	3.007***	2.593***	2.444***
Elementary		.092	.092	.052	.080
High school		-.324+	-.324+	-.314*	-.286*
District A				.469+	.557*
District B				.791*	.870***
District C				.348+	.434*
District D				.864**	.947***
District E				.527*	.614**
District F				.400+	.463+
District G				.692*	.788**
District H				.249	.376
District I				.359	.450+
Teacher-level					
Math		.133	.135	.106	.113
Activity-level					
Year (1998–99)			-.125*	-.124*	-.071
Eisenhower-sponsored					.174
Variance components					
Between-school	.064+	.045	.045	.000	.000
Between-teacher	.337***	.329***	.333***	.340***	.328***
Residual	.359	.359	.353	.353	.357

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect for level; Model 2 (p<.05), Model 3 (p<.05), Model 4 (p<.01), Model 5(p<.01)
Significant effect for district; Model 4 (p<.05), Model 5 (p<.05)
N=400

EXHIBIT C.10

Participation in Professional Development: Effects of Level, Subject, Year, District, and Sponsorship

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	0.843***	0.866***	.899***	.985***
Elementary		-.185***	-.181***	-.193***
High school		.070	.070	.077
District A				-.244**
District B				-.148**
District C				-.097
District D				-.093
District E				-.013
District F				-.030
District G				-.079
District H				-.122+
District I				.002
Teacher-level				
Math		-.032	-.026	-.023
Math*Elementary		.215***	.210**	.207**
Math*High School		.003	.002	.002
Activity-level				
Year (1998–99)			-.071**	-.071**
Variance components				
Between-school	.009*	.005+	.005+	.001
Between-teacher	.000	.000	.000	.000
Residual	.135	.132	.131	.131

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant level effect: Model 2 (p<.01), Model 3 (p<.01), Model 4 (p<.001)
Significant level*subject effect: Model 2 (p<.001), Model 3 (p<.001), Model 4 (p<.001)
Significant district effect: Model 4 (p<.05)
N=864

APPENDIX D

ANALYSIS METHODS AND RESULTS FOR THE EFFECTS OF PROFESSIONAL DEVELOPMENT ON TEACHING PRACTICE

In Section III, we summarized the conclusions of a set of analyses we conducted to examine the extent to which professional development that focuses on specific teaching strategies increases teachers' use of these strategies in the classroom. In this appendix, we describe these analyses in more detail.

The appendix is organized in three parts. First, we provide an overview of our analysis strategy. Second, we describe the results we obtained. Finally, we discuss the methods we used to develop Exhibits 14 through 19 in Section III on the basis of the analysis results.

Overview of Analysis Strategy

As described in Section III, we conducted the analysis on the basis of data from all three waves of the Longitudinal Teacher Survey. We sought to explain teaching practice in 1998–99 (the third wave of the survey) on the basis of teachers' professional development experiences during 1997–98 (the second wave of the survey), controlling for teachers' classroom teaching practice in 1996–97 (the first wave of the survey). Given this analysis strategy, the sample for the analysis is restricted to teachers who returned all three waves of the survey, who participated in professional development in 1997–98, and who continued to teach the same course over all three waves of the survey. The last restriction is necessary because changes in the course taught might introduce changes in teaching practice apart from the effects of professional development experiences. Finally, the sample is restricted to teachers who provided complete data on all of the necessary items. The number of teachers meeting these conditions ranges from about 125 to 135, depending on the specific analysis.

We conducted three parallel sets of analyses, each focusing on a different area of teaching practice. First, we examined the effects of professional development on teaching strategies involving the use of technology for higher-order learning; then, we examined instructional methods for higher-order learning; and finally, we examined assessment strategies. (See Section III for a description of each of these areas and the specific survey items on which we focused.)

To clarify the approach we used, we describe the data, measures, and statistical model we used for technology in detail below. The methods we used for instruction and assessment are exactly parallel.

Data and Measures

For each teacher in our data set, we collected data on the extent to which the teacher used each of four technology strategies: calculators or computers to develop models; calculators or computers for data collection; computers to write reports; and computers to access the Internet. For each teacher, we collected data on the extent to which the teacher used each of the four strategies in the classroom in 1996–97, measured on a scale from 0 to 3, where 0=almost never, 1=some lessons, 2=most lessons, and 3=every lesson. Thus, we have four measures of each teacher's technology use in 1996–97, one for each strategy. We collected parallel data on the teacher's classroom use of each strategy in 1998–99. Finally, we collected data indicating whether the professional development activity in which the teacher participated in 1997–98 focused on each of the four strategies. For each of the four strategies, we assigned a code of 1 if the activity focused on the strategy and 0 if it did not.

In addition to these data about each of the four technology strategies, we also collected data about the teacher and the quality of the professional development activity the teacher attended in 1997–98. In particular, we collected data on each teacher's subject area and school grade level, and we collected data on six measures of the quality of the professional development activity the teacher attended in 1997–98: the activity type (reform vs. traditional), time span, contact hours, collective participation, active learning, and coherence. (For the scales used in measuring each of these features of professional development, see Section II.)

We used these data to address three main issues about the effects of professional development on teaching practice. First, we used the data to examine whether teachers who participated in professional development that focused on a particular teaching strategy (e.g., the use of calculators or computers to develop models) increased their classroom use of that strategy over the period from 1996–97 to 1998–99 more than did similar teachers who did not participate in professional development that focused on the strategy. Second, we used the data to examine whether teachers who participated in professional development that focused on several related strategies (e.g., the use of calculators or computers to develop models and to collect and analyze data) increased their use of calculators and computers to develop models more than did teachers who focused only on that strategy during their professional development. Finally, we used the data to examine whether the benefits of participating in professional development that focused on particular teaching strategies were strengthened if a teacher's professional development had features of high quality (i.e., reform type, appropriate time span, sufficient contact hours, collective participation, active learning, and coherence).

To estimate the magnitude of participating in professional development focused on particular technology teaching strategies, we created two new variables to characterize each professional development activity: the *mean focus* the activity gave to the set of four technology strategies and the *relative focus* the activity gave to each of the four specific technology strategies.

> **Mean focus.** To assess the extent to which the professional development activity that a teacher attended focused on multiple, related strategies, we calculated the average or mean focus given to the teaching strategies we measured. For higher-order technology use, the mean focus is the average emphasis placed on the four higher-order technology-use strategies: use of calculators or computers to develop models, use of calculators or computers to collect data, use of computers to write reports, and use of computers to access the Internet. Since each strategy is coded 1 if it was given attention as part of the teacher's professional development activity and 0 if it was not, the mean focus ranges from 0, if no technology strategies were covered in the activity, to 0.5 if two of the four strategies were covered, to 1 if all four strategies were covered. The more strategies the activity focused on, the higher the mean focus. (See Exhibits D.1 and D.2 for more information on the derivation of mean focus.)
>
> **Relative focus.** To measure the effects of focusing on one strategy rather than another within a professional development activity, we used a measure of relative focus. For example, if an activity focused on two of the four higher-order technology strategies, including the use of calculators and computers to develop models, the relative focus for the use of calculators and computers to develop

models would have a value of 0.5—calculated as the difference between the value of 1 for the use of calculators or computers to develop models and the mean focus of 0.5. (See Exhibits D.1 and D.2 for more information on the derivation of relative focus.)

EXHIBIT D.1

Calculation of Mean Focus and Relative Focus

Suppose the professional development activity attended by a teacher focused on two technology strategies (calculators or computers to develop models and calculators or computers for data collection), but it did not focus on the other two strategies (computers to write reports and computers to access the Internet). As described in the text, the *mean focus* for the teacher is 2/4, or 0.5.

The *relative focus* for each of the four strategies is computed by subtracting the mean focus from the focus for each strategy. The results are shown in the table below.

Higher-order technology strategies	Focus	Relative focus
Strategy 1: Calculators or computers to develop models	1	1-.5= +0.5
Strategy 2: Calculators or computers for data collection	1	1-.5= +0.5
Strategy 3: Computers to write reports	0	0-.5= -0.5
Strategy 4: Computers to access the internet	0	0-.5= -0.5

As shown in the table, for this example, the relative focus for each strategy has a value of plus or minus 0.5, depending on whether or not the activity focused on the strategy.

The profile of values of relative focus for the four topics (labeled 1, 2, 3, and 4) are shown in the graph below.

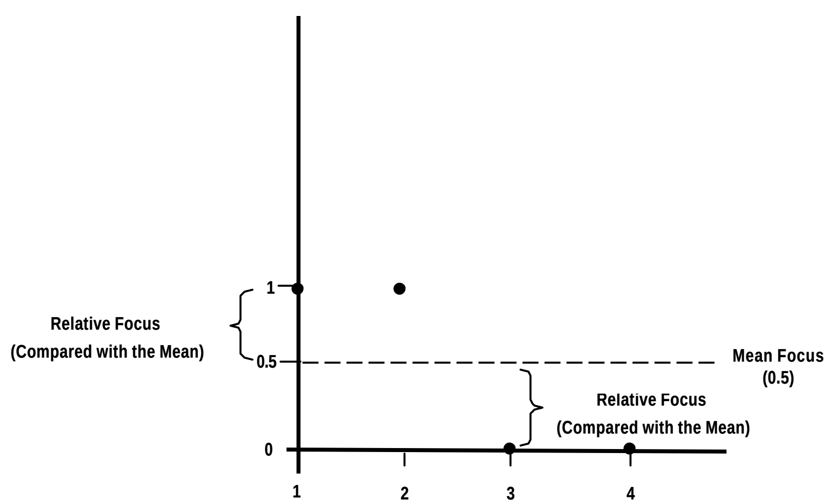


EXHIBIT D.2

Effects of Professional Development on the Use of Higher-Order Teaching Strategies: Model

Variables in the model:

$y_{pi} =$ extent of teacher *i* ‘s 1998–99 classroom use of teaching strategy *p* (*p*=1 to 4 for technology, *p*=1 to 5 for instruction, and *p*=1 to 6 for assessment).

$x_{pi} =$ extent of teacher *i* ‘s 1996–97 classroom use of teaching strategy *p.*

$t_{pi} =$ 0/1 variable indicating whether or not teacher *i* ‘s 1997–98 professional development focused on teaching strategy *p.*

$m_i =$ mean focus of teacher *i* ‘s professional development on the set of higher-order teaching strategies (for technology, $m_i = \frac{t_{1i} + t_{2i} + t_{3i} + t_{4i}}{4}$).

$d_{pi} =$ relative focus of teacher *i* ‘s professional development on strategy *p* (e.g., $d_{pi} = t_{pi} - m_i$).

$q_i =$ quality of teacher *i* ‘s professional development (e.g., teacher score on active learning scale).

$s1_{pi}, s2_{pi}, etc =$ set of 0/1 variables specifying the teaching strategy being modeled (e.g., for technology, $s1_{pi} = 1$ indicates the use of calculators or computers to develop models; $s2_{pi} = 1$ indicates the use of calculators or computers for data collection and analysis).

$subject =$ 0/1 variable specifying the teacher's subject (mathematics=1/science=0).

$elem, high =$ 0/1 variables specifying the school level, with middle school used as the reference category.

Level 1 model: Use of specific teaching strategies

$$y_{pi} = \mathbf{p}_{0i} + \mathbf{p}_{1i} d_{pi} + \mathbf{p}_2 x_{pi} + \mathbf{p}_3 s1_{pi} + \mathbf{p}_4 s2_{pi} + \mathbf{p}_5 s3_{pi} + \mathbf{e}_{pi}$$

Level 2 model: Teacher/activity-level effects on use of specific teaching strategies

$\mathbf{p}_{0i} = \mathbf{b}_{00} + \mathbf{b}_{01} m_i + \mathbf{b}_{02} subject + \mathbf{b}_{03} elem + \mathbf{b}_{04} high + \mathbf{b}_{05} q_i + \mathbf{b}_{06} m_i q_i + \mathbf{n}_{0i}$, equation for the intercept in the level 1 model

$\mathbf{p}_{1i} = \mathbf{b}_{20} + \mathbf{b}_{21} q_i + \mathbf{n}_{2i}$, equation for the slope for d_{pi} (relative focus on strategy *p*) in the level 1 model

We chose to use mean focus and relative focus to characterize professional development activities because the variables clearly distinguish between the benefits of focusing on one strategy rather than another within a professional development activity (captured by the relative focus) and the benefits of professional development activities that focus on many or few strategies (captured by the mean focus).[62] The effects of focusing on a set of strategies in a professional development activity can be examined by comparing the magnitude of the coefficients for mean focus and relative focus. If the coefficient for mean focus is higher than the coefficient for relative focus, there is a "spillover" effect in which focusing on a set of related strategies has an effect over and above the effect of focusing on an individual strategy alone. If the coefficients for the two variables are equal, focusing on multiple strategies neither helps nor hurts. If the coefficient for mean focus is lower than the coefficient for relative focus, it indicates that focusing on multiple strategies is harmful—that is, activities focusing on a single strategy are more effective in boosting the use of the strategy than are activities that focus on several related strategies.[63]

Statistical Methods

Technically, our data have a two-level structure, with the four technology teaching strategies nested within teachers. In the discussion that follows, we refer to the two levels at which we have data as the "strategy" and the "teacher/activity" levels. We use the term teacher/activity for the teacher level because our data at that level include both teacher characteristics (e.g., subject taught) and characteristics of the quality of the professional development activity the teacher attended in 1997–98.

Given the two-level (strategy-level and teacher/activity-level) structure of the data, we estimated the effects of professional development by using a hierarchical linear model. (See Exhibit D.2 for the model equations.) The model for the effects of professional development on the use of higher-order teaching strategies in technology includes the following teacher/activity-level and strategy-level variables:

> **Teacher/activity-level variables.** At the teacher/activity level, we included the following variables in the model: the *mean focus* given to the four higher-order technology strategies during the professional development activity the teacher attended in 1997–98, the *quality* of the professional development (e.g., the time span or degree of collective participation), and controls for the teacher's subject (mathematics or science) and grade level (elementary, middle, or high school).[64]

---

[62] The approach we followed is similar to the approach used by Bryk and Raudenbush (1992) to distinguish individual and contextual effects in models involving students nested within schools. In such models, Bryk and Raudenbush propose centering measures of student background on the school mean and entering both the centered student values and the school means in the analysis.

[63] The conclusions can be derived from the variable definitions in Exhibit D.2. If t_{pi} is the focus given to strategy p by teacher i (coded 1/0), m_i is the mean focus on the four technology strategies for teacher i, $d_{pi} = (t_{pi} - m_i)$ is the relative focus on strategy p for teacher i, b_m is the coefficient for mean focus, and b_r is the coefficient for relative focus, then the overall effect on the use of strategy p by teacher i can be written as follows:

$$b_m m_i + b_r d_{pi} = b_m m_i + b_r (t_{pi} - m_i) = (b_m - b_r) m_i + b_r t_{pi}$$

Thus, mean focus (m_i) has a positive effect if $b_m - b_r > 0$; no effect if $b_m = b_r$, and a negative effect if $b_m < b_r$.

[64] Mean focus is a teacher/activity-level variable because it characterizes the activity the teacher attended as a whole (the average emphasis the professional development activity placed on the four technology strategies); it does not characterize each strategy separately.

Strategy-level variables. For each of the four technology teaching strategies, we included two variables in the model: the teacher's 1996–97 use of the strategy and the *relative focus* given to the strategy during the professional development the teacher attended in 1997–98. We also included a set of indicator variables specifying the particular strategy. These variables represent the fact that on average, teachers may have increased their use of some strategies more than others over the period under study.

We assumed that two key parameters in the strategy-level model would vary among teachers: the strategy-level *intercept*, which represents the average use of the four technology teaching strategies in 1998–99, controlling for their use in 1996–97 and for the teacher's 1997–98 participation in professional development; and the strategy-level *slope*, which represents the effects of focusing on a particular technology-use strategy during professional development on classroom use of the strategy in 1998–99.[65] (See Exhibit D.2 for the equations for the strategy-level and teacher/activity-level equations.) These assumptions reflect the idea that teachers may differ in the degree to which they changed practice over the period from 1996–97 through 1998–99 and in their responsiveness to professional development. One key analysis question concerns the extent to which a teacher's strategy-level slope and intercept are affected by characteristics of the activities in which the teacher participated—in particular, the mean focus on the set of technology-use strategies and the quality features of the activity.

We conducted separate analyses for each of the three areas under study (technology, instructional practice, and assessment). For each area, we estimated seven models, one including only the strategy variables (mean focus and relative focus) and controls, and the others adding each of the six professional development quality feature, one at a time.[66]

At the teacher level, the sample size for our analyses is about 125.[67] Since, for each teacher, we have data on four technology-use strategies, the sample size available to estimate the effects of professional development on classroom use is about 4 * 125 = 500. The sample of strategies for the analysis of higher-order instruction is about 5 * 125 = 625; and the sample for higher-order assessment is about 6 * 125 = 750.

Overview of Analysis Results

The results of our analysis for technology are presented in Exhibit D.3, for instruction in D.4, and for assessment in D.5. Each exhibit contains the results for seven models. Model 1 examines the effects of focusing on a strategy during professional development on use of the strategy in 1998–99, but does not include the effects of the professional development quality features. Models 2 through 7 examine each quality feature, one at a time (reform type, time span, hours, collective participation, active learning, and coherence).

To illustrate the meaning of the parameters displayed in the tables, we discuss two technology models presented in Exhibit D.3: Model 1 (the baseline model) and Model 6 (the model examining the effects of active learning).

[65] In technical terms, we modeled these two parameters as random effects. We modeled all other parameters as fixed effects.
[66] Given the relatively small overall sample size, we estimated separate models for each quality feature instead of including all quality features in a single model.
[67] The exact sample size depends on the number of teachers with complete data on the variables included in the analysis.

Exhibit D.3: Model 1 (baseline model). The parameter estimates for Model 1 are presented in two main groups. The first group contains the parameters for the strategy-level parameters that do not vary among teachers, i.e., the parameter for the effects of 1996–97 use of each strategy on 1998–99 use, and parameters representing the average 1998–99 use for each specific strategy relative to the others, controlling for 1996–97 use. The subscripted Greek letters in parentheses on each row of the table refer to the coefficients in the equations in Exhibit D.2.

The first strategy-level parameter shown in the Exhibit (π_2 = 0.462***) indicates that as we would expect, 1996–97 use of each strategy has a positive, highly significant effect on 1998–99 use. The remaining strategy-level coefficients represent the average 1998–99 use for each strategy relative to the others. The results for these coefficients indicate that in 1998–99, teachers tended to use computers to write reports more than they used the other strategies, controlling for their 1996–97 level of use (π_5 = 0.179***). Recall that technology use is measured on a scale from 0 to 3, where 0=almost never, 1=some lessons, 2=most lessons, and 3=every lesson. Thus, the coefficient estimate of 0.179 for using the computer to write reports indicates that this strategy was used somewhat more frequently than others, controlling for prior use.[68]

The second group of parameters presented for Model 1 contains the parameters representing the effects of teacher/activity variables on each teacher's intercept and slope in the strategy-level model. The first coefficient shown (β_{00} = 0.384**) represents the baseline level of use for the typical teacher in 1998–99, controlling for 1996–97 use. The coefficient indicates that a teacher who did not use a technology strategy at all in 1996–97 would be expected to have a use of 0.384 in 1998–99.[69]

The next coefficient, which represents the effects of the mean focus given to technology-use strategies in professional development (β_{01} = 0.342**), indicates that teachers who participated in professional development that covered more strategies tended to make more use of each strategy in their classroom practice in 1998–99, controlling for prior use. In particular, a teacher who was in an activity that focused on all four technology-use strategies (mean focus=1) would have a predicted 1998–99 use 0.342 points higher than a teacher whose professional development did not focus on any of the four technology strategies (mean focus=0).

The three teacher/activity-level coefficients that follow represent the effects of subject taught and school level. Among these coefficients, the only significant effect is for elementary school (β_{03} = -0.338***). This coefficient indicates that elementary teachers were less likely to use higher-order technology strategies in 1998–99 than were middle- and high-school teachers, controlling for prior use.

The final teacher/activity-level coefficient represents the baseline effect of the relative focus on a particular technology strategy on the use of the strategy in the classroom (i.e., the effect for a

[68] We also found differences in the 1998–99 use of specific strategies in the areas of instruction and assessment. For instruction, teachers tended to have students debate ideas and explain their reasoning or to have students develop their technical writing skills more frequently than they had students work on independent long-term projects, work on problems with no obvious solution, or work on interdisciplinary lessons, controlling for use of these strategies in 1996–97. In assessment, teachers tended to place more importance on performance tasks and systematic observation of students than on essay tests, math/science reports, projects, or portfolios, controlling for the use of these strategies in 1996–97.

[69] The expected use of 0.384 pertains to the use of computers to access the Internet, by a science teacher at the middle school level. To determine the expected use of other strategies, the specific strategy coefficients must be added (e.g., 0.179 for computers to write reports). To determine the expected use for a mathematics teacher or a teacher at the elementary or high school level, the appropriate coefficients must be added.

typical teacher).[70] The estimated coefficient is positive and significant (β_{20} = 0.310**), which indicates that professional development that focuses on a particular technology-use strategy (relative to the other technology strategies) increases the use of the strategy in the classroom.

The coefficient for the effect of the mean focus on the set of four technology strategies (β_{01} = 0.342) is slightly larger than the coefficient for the relative focus on a particular strategy (β_{20} = 0.310). As described above, the difference between the coefficients represents the added effect of focusing on multiple strategies on the use of a particular strategy, over and above the effect of focusing on the particular strategy alone. The positive difference we observe suggests a positive spillover effect, but a test of the difference between the two coefficients indicates that the spillover effect is not significant.[71]

The variance components shown near the bottom of the exhibit indicate that there is significant between-teacher variation in the strategy-level intercept and slope (0.076*** and 0.167*), after controlling for the variables in the model.[72] This indicates that teachers differ in their use of technology strategies in 1998–99, after controlling for the variables in the model; in addition, they differ in their responsiveness to professional development. Other characteristics, beyond those included in the model, may help explain this variation.

Exhibit D.3: Model 6 (active learning). Model 6 differs from Model 1 in that it includes a variable representing the extent to which active learning opportunities were provided as part of the activities in which teachers participated. Otherwise, the model is identical to Model 1.

The first group of parameters shown contains strategy-level parameters. The results are almost identical to those for Model 1. The 1996–97 use of each strategy has a positive, highly significant effect on 1998–99 use (π_2 = 0.423***). In addition, teachers tended to report more use of computers to write reports than the other strategies, controlling for 1996–97 use (π_5 = 0.183***).

The next group of coefficients contains the parameters for the effects of teacher/activity-level variables on the strategy-level intercept. The baseline coefficient (β_{00} = 0.368**) represents the average level of use of each technology strategy for the typical teacher in 1998–99, controlling for prior use. The next coefficient (β_{01} = 0.203), represents the effects of the mean focus on the set of four technology strategies. The coefficient is positive, as hypothesized, but not significant. The next three coefficients represent the effects of subject taught and school level. As in Model 1, the results indicate that elementary teachers tended to use technology strategies less than did other teachers in 1998–99, controlling for prior use.

The coefficient for the effect of active learning on the strategy-level intercept (β_{05} = 0.014) is positive but not significant. The coefficient represents the effect of the number of active learning opportunities provided as part of a teacher's professional development activity (which, as described

---

[70] As we indicated above, we assumed that the strategy-level intercept might vary among teachers. Thus, the baseline coefficient shown is the average or typical value among teachers in the sample; values for individual teachers vary around this average. The variance components shown at the bottom of Exhibit D.3 indicate the extent of this variation.

[71] The standard error for the difference in the coefficients (not shown in the exhibit) is 0.14, and the significance level is $p<.83$. The estimated spillover effect is positive but negligible for instruction. It is positive and larger for assessment, although still not significant ($p<.23$).

[72] The other two variance components shown in the exhibit include the covariation between the intercept and slope and the residual variance. The first of these indicates the extent to which teachers who have unusually high intercepts also have unusually high slopes; the second indicates the remaining variance in the use of each technology strategy in 1998–99, after all other measured variables and variance components are taken into account.

in Section II, ranges from 0 to 20) on the average use of technology in 1998–99, controlling for prior use.[73]

The coefficient for the interaction of active learning and mean focus (β_{06} = 0.019) is also positive but is not significant. The coefficient represents the extent to which the effect of focusing on a set of technology-use strategies as part of a professional development activity is strengthened if the activity incorporates opportunities for active learning.

The final group of coefficients concerns the effects of teacher/activity-level variables on the strategy-level slope. The first coefficient in this group represents the effect of the relative focus on a particular technology strategy on the use of the strategy in the classroom (i.e., the effect for a typical teacher). The estimated coefficient is positive but not significant (β_{20} = 0.137). The second coefficient in this group represents the degree to which the effect of focusing on a particular technology-use strategy during professional development is strengthened if the activity provides opportunities for active learning. The estimated coefficient (β_{21} = 0.041+) indicates that the effect is positive and significant ($p<.10$). The magnitude of the effect can be assessed by combining the baseline slope estimate (β_{20} = 0.137) and the coefficient for active learning (β_{21} = 0.041+) for activities that differ in the active learning opportunities provided. For an activity that provided no opportunities for active learning, the effects of the relative focus on a particular strategy on the use of the strategy in the classroom would be 0.137 (β_{20} only). For an activity that provided 10 opportunities for active learning, the effects of relative focus on a particular strategy would be 0.137 + 10 * 0.041 = 0.137 + 0.410 = 0.547, which is a substantial increase.

The variance components near the bottom of the exhibit indicate that significant variation remains between teachers in the strategy-level intercept, which indicates that teachers differ in their 1998–99 use of specific technology-use strategies, after controlling for the variables in the model. The variation among teachers in the strategy-level slope is substantially lower than in the baseline model and is no longer significant. This suggests that by including active learning opportunities in the model, we have explained a good deal of the variation among teachers in the effectiveness of the professional development they experienced.

Methods Used to Derive Exhibits in Section III

The models we estimated involve a substantial number of coefficients at the teacher/activity and strategy levels, and drawing substantive conclusions from the models entails combining the results of a number of coefficients at both levels. To clarify the discussion in Section III, we presented simulated results indicating the predicted use of specific higher-order teaching strategies in 1998-99.These were based on alternative assumptions about teachers' experiences in professional development.

In the following paragraphs, we briefly summarize the assumptions underlying each exhibit in Section III.

Exhibit 11. Effects of professional development on the use of calculators or computers to develop models. The exhibit is based on parameter estimates for Model 1 in Exhibit D.3. The simulated 1998–99 use pertains to a middle school science teacher whose 1996–97 use was at the overall 1996–97 mean (0.49). The first bar shown in the exhibit represents teachers whose

[73] The positive coefficient suggests that activities that provide active learning opportunities may increase the use of technology in the classroom, regardless of the focus of the activity. We did not hypothesize such an effect; we included the additive effect of active learning opportunities as a control.

professional development focused on no technology strategies (mean focus=0, relative focus=0), and the second represents teachers who focused only on the use of calculators or computers to develop models (mean focus=0.25, relative focus=0.75).

Exhibit 12. Effects of professional development on the use of problems with no immediate solution. The exhibit is based on parameter estimates for Model 1 in Exhibit D.4. The simulated 1998–99 use pertains to a middle school science teacher whose 1996–97 use was at the overall 1996–97 mean (0.88). The assumptions about mean focus and relative focus are analogous to those described in Exhibit 14.

Exhibit 13. Effects of professional development on the use of mathematics and science projects to determine student grades. The exhibit is based on parameter estimates for Model 1 in Exhibit D.5. The simulated 1998–99 use pertains to a middle school science teacher whose 1996–97 use was at the overall 1996–97 mean (1.34). The assumptions about mean focus and relative focus are analogous to those described in Exhibit 14.

Exhibit 15. Effects of professional development on the use of calculators and computers to develop models (collective participation). The exhibit is based on parameter estimates for Model 5 in Exhibit D.3. The simulated 1998–99 use pertains to a middle school science teacher whose 1996–97 use of calculators or computers to develop models was at the overall 1996–97 mean (0.49). The first bar shown in the exhibit represents teachers whose professional development focused on no technology-use strategies (mean focus=0, relative focus=0) and involved no collective participation (collection participation=0). The second bar shown represents teachers whose professional development focused on all four technology-use strategies (mean focus=1, relative focus=0) and involved no collective participation (collective participation=0). The third bar represents teachers whose professional development focused on all four technology use strategies (mean focus=1, relative focus=0) and involved high collective participation (collective participation=2).

Exhibit 16. Effects of professional development on the use of calculators and computers to develop models (active learning). The exhibit is based on parameter estimates for Model 6 in Exhibit D.3. The simulated 1998–99 use pertains to a middle school science teacher whose 1996–97 use of calculators or computers to develop models was at the overall 1996–97 mean (0.49). The first bar shown in the exhibit represents teachers whose professional development focused on no technology-use strategies (mean focus=0, relative focus=0) and involved no active learning (active learning=0). The second bar shown represents teachers whose professional development focused on all four technology-use strategies (mean focus=1, relative focus=0) and involved no active learning (active learning=0). The third bar represents teachers whose professional development focused on all four technology use strategies (mean focus=1, relative focus=0) and involved high active learning (active learning=8).

Exhibit 17. Effects of professional development on the use of problems for which there is no obvious solution (reform type). The exhibit is based on parameter estimates for Model 2 in Exhibit D.4. The simulated 1998–99 use pertains to a middle school science teacher whose 1996–97 use of problems for which there is no obvious solution was at the overall 1996–97 mean (0.88). The first bar shown in the exhibit represents teachers whose professional development focused on no higher-order instruction strategies (mean focus=0, relative focus=0) and was not organized as a reform-type activity (reform type=0). The second bar shown represents teachers whose professional development focused on all five higher-order strategies (mean focus=1, relative focus=0) and was not a reform-type activity (reform type=0). The third bar represents teachers whose professional

development focused on all five higher-order instruction strategies (mean focus=1, relative focus=0) and was a reform-type activity (reform type=1).

EXHIBIT D.3

Effects of Professional Development on the Use of Higher-Order Teaching Strategies: Technology

Coefficient	Model 1: Base	Model 2: Reform type	Model 3: Time span	Model 4: Hours
Level 1 model: Use of teaching strategies				
1996–97 Extent of classroom use of strategy, π_2	.462***	.465***	.452***	.463***
Calculators or computers to develop models (0/1), π_3	.076	.071	.093	.099
Calculators or computers for data collection (0/1), π_4	.071	.066	.090	.094
Computers to write reports (0/1), π_5	.179**	.165*	.176**	.185**
(reference category: computers to access the Internet)				
Level 2 model: Teacher/activity-level effects				
Effects on intercept in strategy model (p_{0i})				
Baseline, β_{00}	.384***	.449***	.338**	.356***
Mean focus on set of strategies, β_{01}	.342**	-.028	.188	.444*
Subject taught: (Mathematics), β_{02}	-.068	-.060	-.078	-.080
Elementary School, β_{03}	-.338***	-.343***	-.352***	-.351***
High School, β_{04}	-.071	-.059	-.055	-.075
Reform type, β_{05}		-.072		
Time span, β_{05}			.010	
Hours, β_{05}				.001
Collective participation, β_{05}				
Active learning, β_{05}				
Coherence, β_{05}				
Reform type*mean focus, β_{06}		.170		
Time span*mean focus, β_{06}			.039	
Hours*mean focus, β_{06}				-.004
Collective participation*mean focus, β_{06}				
Active learning*mean focus, β_{06}				
Coherence*mean focus, β_{06}				
Effects on d_{pi} slope in strategy model (p_{1i})				
Baseline, β_{20}	.310***	.235	.139	.299*
Reform type, β_{21}		-.049		
Time span, β_{21}			.042	
Hours, β_{21}				.000
Collective participation, β_{21}				
Active learning, β_{21}				
Coherence, β_{21}				
Variance components				
Between-teacher variance in intercept	.076***	.081***	.080***	.079***
Between-teacher variance in slope	.167*	.182*	.184*	.190*
Covariation in intercept/slope	.048	.048	.043	.046
Residual	.222	.222	.207	.207
Degrees of freedom				
Strategy level	351	341	341	347
Teacher/activity level	114	109	109	111

+ p<.10, * p<.05, **p<.01, ***p<.001

EXHIBIT D.3 (Continued)

Effects of Professional Development on the Use of Higher-Order Teaching Strategies: Technology

Coefficient	Model 5: Collective participation	Model 6: Active learning	Model 7: Coherence
Level 1 model: Use of teaching strategies			
1996–97 Extent of classroom use of strategy, π_2	.454***	.423***	.436***
Calculators or computers to develop models (0/1), π_3	.082	.083	.069
Calculators or computers for data collection (0/1), π_4	.071	.081	.078
Computers to write reports (0/1), π_5	.177**	.183**	.184**
(reference category: computers to access the Internet)			
Level 2 model: Teacher/activity-level effects			
Effects on intercept in strategy model (p_{0i})			
Baseline, β_{00}	.350***	.368**	.206
Mean focus on set of strategies, β_{01}	.328*	.203	.055
Subject taught: (Mathematics), β_{02}	-.082	-.095	-.080
Elementary School, β_{03}	-.348***	-.336***	-.315**
High School, β_{04}	-.055	-.028	-.046
Reform type, β_{05}			
Time span, β_{05}			
Hours, β_{05}			
Collective participation, β_{05}	.101		
Active learning, β_{05}		.014	
Coherence, β_{05}			.033
Reform type*mean focus, β_{06}			
Time span*mean focus, β_{06}			
Hours*mean focus, β_{06}			
Collective participation*mean focus, β_{06}	.027		
Active learning*mean focus, β_{06}		.019	
Coherence*mean focus, β_{06}			.045
Effects on d_{pi} slope in strategy model (p_{1i})			
Baseline, β_{20}	.190+	.137	.048
Reform type, β_{21}			
Time span, β_{21}			
Hours, β_{21}			
Collective participation, β_{21}	.326*		
Active learning, β_{21}		.041+	
Coherence, β_{21}			.047
Variance components			
Between-teacher variance in strategy intercept	.074***	.075***	.078***
Between-teacher variance in strategy slope	.148+	.080	.166+
Covariation in intercept/slope	.040	.030	.045
Residual	.221	.228	.229
Degrees of freedom			
Strategy level	350	332	323
Teacher/activity level	112	106	103

+ p<.10, * p<.05, **p<.01, ***p<.001

EXHIBIT D.4

Effects of Professional Development on the Use of Higher-Order Teaching Strategies: Instruction

Coefficient	Model 1: Base	Model 2: Reform type	Model 3: Time span	Model 4: Hours
Level 1 model: Use of teaching strategies				
1996–97 Extent of classroom use of strategy, π_2	0.387***	.390***	.383***	.384***
Work on independent, long-term projects (0/1), π_3	-.269***	-.282***	-.268***	-.255***
Work on problems with no obvious solution (0/1), π_4	-.205**	-.227**	-.197**	-.196**
Develop technical writing skills (0/1), π_5	-.065	-.100	-.07	-.048
Work on interdisciplinary lessons (0/1), π_6	-.125+	-.142+	-.133+	-.117
(reference category: debate ideas, explain reasoning)				
Level 2 model: Teacher/activity-level effects				
Effects on intercept in strategy model (p_{0i})				
Baseline, β_{00}	.962***	.925***	.842***	.858***
Mean focus on set of strategies, β_{01}	.234*	.082	-.098	.114
Subject taught: (Mathematics), β_{02}	-.058	.037	.055	.044
Elementary School, β_{03}	-.225**	-.219**	-.238**	-.213*
High School, β_{04}	-.219*	-.245**	-.197*	-.206*
Reform type, β_{05}		-.351*		
Time span, β_{05}			.004	
Hours, β_{05}				-.001
Collective participation, β_{05}				
Active learning, β_{05}				
Coherence, β_{05}				
Reform type*mean focus, β_{06}		.872**		
Time span*mean focus, β_{06}			.065	
Hours*mean focus, β_{06}				.004
Collective participation*mean focus, β_{06}				
Active learning*mean focus, β_{06}				
Coherence*mean focus, β_{06}				
Effects on d_{pi} slope in strategy model (p_{1i})				
Baseline, β_{20}	.223***	.212**	.069	.169+
Reform type, β_{21}		.018		
Time span, β_{21}			.035	
Hours, β_{21}				.002
Collective participation, β_{21}				
Active learning, β_{21}				
Coherence, β_{21}				
Variance components				
Between-teacher variance in intercept	.062***	.058***	.059***	.061***
Between-teacher variance in slope	.047	.059	.050	.048
Covariation in intercept/slope	.029	.034	.023	.024
Residual	.306	.304	.306	.308
Degrees of freedom				
Strategy level	513	496	500	504
Teacher/activity level	126	120	121	122

+ p<.10, * p<.05, **p<.01, ***p<.001

EXHIBIT D.4 (Continued)

Effects of Professional Development on the Use of Higher-Order Teaching Strategies: Instruction

Coefficient	Model 5: Collective participation	Model 6: Active learning	Model 7: Coherence
Level 1 model: Use of teaching strategies			
1996–97 Extent of classroom use of strategy, π_2	.389***	.375***	.384***
Work on independent, long-term projects (0/1), π_3	-.267***	-.257***	-.254**
Work on problems with no obvious solution (0/1), π_4	-.204**	-.203**	-.200**
Develop technical writing skills (0/1), π_5	-.065	-.067	-.067
Work on interdisciplinary lessons (0/1), π_6	-.124	-.148*	-.134+
(reference category: debate ideas, explain reasoning)			
Level 2 model: Teacher/activity-level effects			
Effects on intercept in strategy model (π_{0i})			
Baseline, β_{00}	.837***	.918***	.826***
Mean focus on set of strategies, β_{01}	.121	-.048	-.320
Subject taught: (Mathematics), β_{02}	.037	.028	.036
Elementary School, β_{03}	-.218**	-.203*	-.202*
High School, β_{04}	-.191*	-.182*	-.183*
Reform type, β_{05}			
Time span, β_{05}			
Hours, β_{05}			
Collective participation, β_{05}	.032		
Active learning, β_{05}		-.017	
Coherence, β_{05}			.005
Reform type*mean focus, β_{06}			
Time span*mean focus, β_{06}			
Hours*mean focus, β_{06}			
Collective participation*mean focus, β_{06}	.193		
Active learning*mean focus, β_{06}		.057+	
Coherence*mean focus, β_{06}			.086
Effects on d_{pi} slope in strategy model (π_{1i})			
Baseline, β_{20}	.218**	.227*	.166
Reform type, β_{21}			
Time span, β_{21}			
Hours, β_{21}			
Collective participation, β_{21}	.011		
Active learning, β_{21}		.001	
Coherence, β_{21}			.011
Variance components			
Between-teacher variance in strategy intercept	.058***	.058***	.061***
Between-teacher variance in strategy slope	.050	.051	.050
Covariation in intercept/slope	.032	.027	.027
Residual	.306	.308	.308
Degrees of freedom			
Strategy level	512	488	476
Teacher/activity level	124	118	115

+ $p<.10$, * $p<.05$, ** $p<.01$, *** $p<.001$

EXHIBIT D.5

Effects of Professional Development on the Use of Higher-Order Teaching Strategies: Assessment

Coefficient	Model 1: Base	Model 2: Reform type	Model 3: Time span	Model 4: Hours
Level 1 model: Use of teaching strategies				
1996–97 Extent of classroom use of strategy, π_2	.445***	.440***	.442***	.443***
Essay tests (0/1), π_3	.130	.137	.152+	.138
Performance tasks (0/1), π_4	.494***	.507***	.518***	.504***
Systematic observation of students (0/1), π_5	.640***	.662***	.648***	.642***
Math/science reports (0/1), π_6	.225**	.239**	.252**	.230**
Math/science project (0/1), π_7	.289***	.308***	.319***	.299***
(reference category: portfolios)				
Level 2 model: Teacher/activity-level effects				
Effects on intercept in strategy model (p_{0i})				
Baseline, β_{00}	.488***	.574***	.591***	.582***
Mean focus on set of strategies, β_{01}	.494**	.342*	.108	.340+
Subject taught: (Mathematics), β_{02}	-.131	-.137+	-.128	-.137+
Elementary School, β_{03}	-.235*	-.259*	-.257*	-.258*
High School, β_{04}	-.143	-.184	-.144	-.133
Reform type, β_{05}		-.411*		
Time span, β_{05}			-.031	
Hours, β_{05}				-.005
Collective participation, β_{05}				
Active learning, β_{05}				
Coherence, β_{05}				
Reform type*mean focus, β_{06}		.932*		
Time span*mean focus, β_{06}			.095	
Hours*mean focus, β_{06}				.008
Collective participation*mean focus, β_{06}				
Active learning*mean focus, β_{06}				
Coherence*mean focus, β_{06}				
Effects on d_{pi} slope in strategy model (p_{1i})				
Baseline, β_{20}	.297***	.331***	.355*	.320**
Reform type, β_{21}		-.225		
Time span, β_{21}			-.014	
Hours, β_{21}				-.001
Collective participation, β_{21}				
Active learning, β_{21}				
Coherence, β_{21}				
Variance components				
Between-teacher variance in intercept	.125***	.120***	.128***	.125***
Between-teacher variance in slope	.057	.059	.068	.064
Covariation in intercept/slope	.001	.010	.003	.005
Residual	.438	.443	.440	.438
Degrees of freedom				
Strategy level	616	600	600	610
Teacher/activity level	120	115	115	117

+ p<.10, * p<.05, **p<.01, ***p<.001

EXHIBIT D.5 (Continued)
Effects of Professional Development on the Use of Higher-Order Teaching Strategies: Assessment

Coefficient	Model 5: Collective participation	Model 6: Active learning	Model 7: Coherence
Level 1 model: Use of teaching strategies			
1996–97 Extent of classroom use of strategy, π_2	.441***	.434***	.438***
Essay tests (0/1), π_3	.137	.087	.071
Performance tasks (0/1), π_4	.507***	.457***	.472***
Systematic observation of students (0/1), π_5	.653***	.617***	.643***
Math/science reports (0/1), π_6	.239***	.192*	.188*
Math/science project (0/1), π_7	.298**	.267**	.255**
(reference category: portfolios)			
Level 2 model: Teacher/activity-level effects			
Effects on intercept in strategy model (p_{0i})			
Baseline, β_{00}	.514***	.489***	.572*
Mean focus on set of strategies, β_{01}	.344+	.458*	-.252
Subject taught: (Mathematics), β_{02}	-.143+	-.115	-.122
Elementary School, β_{03}	-.223*	-.185+	-.178
High School, β_{04}	-.130	-.085	-.078
Reform type, β_{05}			
Time span, β_{05}			
Hours, β_{05}			
Collective participation, β_{05}	-.076		
Active learning, β_{05}		-.025	
Coherence, β_{05}			-.029
Reform type*mean focus, β_{06}			
Time span*mean focus, β_{06}			
Hours*mean focus, β_{06}			
Collective participation*mean focus, β_{06}	.313		
Active learning*mean focus, β_{06}		.042	
Coherence*mean focus, β_{06}			.141+
Effects on d_{pi} slope in strategy model (p_{1i})			
Baseline, β_{20}	.376***	.409***	.012
Reform type, β_{21}			
Time span, β_{21}			
Hours, β_{21}			
Collective participation, β_{21}	-.183		
Active learning, β_{21}		-.023	
Coherence, β_{21}			.046
Variance components			
Between-teacher variance in strategy intercept	.125***	.110***	.111***
Between-teacher variance in strategy slope	.053	.068	.068
Covariation in intercept/slope	.009	-.001	-.001
Residual		.431	.431
Degrees of freedom			
Strategy level	615	585	570
Teacher/activity level	118	112	109

+ p<.10, * p<.05, **p<.01, ***p<.001

APPENDIX E

ANALYSIS METHODS AND RESULTS FOR VARIATION AND TRENDS IN TEACHING PRACTICE

In Sections III and IV, we described the classroom practices of teachers in our sample of 30 schools. As discussed in Section III, we examined 11 measures of classroom teaching practice: the alignment between the content that teachers cover and the items on the National Assessment of Educational Progress (NAEP); the emphasis placed on six performance goals (memorization, understanding, procedural skills, hypothesis generation, data collection/analysis, and making connections); and the extent to which teachers rely on four pedagogical approaches (didactic instruction, individual seatwork, active project-centered instruction, and discussion-oriented instruction).

In this appendix, we describe the methods we used to analyze these 11 measures of classroom teaching practice. First, we discuss the pedagogy measures we employed. (The measures of alignment and emphasis on perrformance goals are discussed in detail in Section III.) Next, we provide an overview of our analytic approach. Then, we discuss the results we obtained. Finally, we discus the methods we used to compute the percent variation in teaching practices across school levels, subjects, schools, and years, displayed in Exhibit 19 in Section IV.

Pedagogy Measures

To create scales measuring pedagogical approaches, we drew on a series of items in the longitudinal survey about teachers' pedagogical strategies. Exhibit E.1 lists the relevant survey items.

To identify patterns in teachers' pedagogical activities, we conducted factor analyses on the full set of items identified in Exhibit E.1. The following four factors, consistent with research on pedagogy, emerged.

Didactic instruction. (alpha reliability=.75).

- Students working on interdisciplinary lessons (reverse coded)
- Students using concrete models or manipulatives (reverse coded)
- Teacher lecturing to class
- Teacher working with students in small groups (reverse coded)
- Students listening/taking notes/observing demonstration
- Students reading (reverse coded)
- Students completing a performance task, writing (reverse coded)
- Students presenting material to the class (reverse coded)

Individual seatwork. (alpha reliability=.69).

- Students working on or reviewing homework in class
- Students working on paper-and-pencil exercises related to the topic
- Students reciting or drilling orally
- Students completing a short test or quiz to review previous lesson
- Students work individually without teacher's ongoing assistance
- Students work individually with teacher's ongoing assistance
- Students work in pairs or small groups without teacher's ongoing assistance

Active, project-centered instruction. (alpha reliability=.67).

- Students working on independent, long-term projects
- Students working on problems for which there is no immediately obvious method or solution
- Students developing technical or mathematical/scientific writing skills
- Teacher working with students individually (reverse coded)
- Teacher helping students with experiments, projects, or other hands-on experiences
- Students doing lab or field work
- Students completing exercises/taking a test or quiz (reverse coded)
- Students doing other (reverse coded)

Discussion-oriented instruction. (alpha reliability=.67).

- Teacher leading whole class discussion
- Students engaging in discussion

For more information on the development and scaling of the four composite measures of pedagogy, see Garet, et al. (1999).

EXHIBIT E.1

Pedagogy Questions Drawn from Middle School Mathematics Survey

1. Indicate the percentage of time in the target class you spent on math instruction in each of the following activities. (Note: Total should sum to 100%.)

Teacher Activities	**Percent of Instr. Time**
a. Lecturing to the class	________
b. Providing demonstrations to the class (including lab demonstrations	________
c. Leading whole-class discussions	________
d. Working with students in small groups	________
e. Working with students individually	________
f. Performing routine administrative tasks (e.g., taking attendance, making announcements, etc.)	________
g. Helping students with experiments, projects, or other hands-on experiences	________
h. Other: (please specify) ____________________	________
TOTAL	100%

2. Indicate the percentage of class time spent on math instruction that the typical student is engaged in each of the following activities. (Note: Total should sum to 100%.)

Student Activities	**Percent of Instr. Time**
a. Listening/taking notes/observing demonstrations	________
b. Engaged in discussion	________
c. Doing lab or field work	________
d. Completing exercises/taking a test or quiz	________
e. Reading	________
f. Completing a performance task, writing	________
g. Presenting material to the class	________
h. Other: (please specify) ______________________	________
TOTAL	100%

EXHIBIT E.1 (Continued)

Pedagogy Questions Drawn from Middle School Mathematics Survey

3. How often did you have students (during math): (Circle one for each line.)

	Almost Never	Some Lessons	Most Lessons	Every Lesson
a. Work on or review homework in class.............................	0	1	2	3
b. Work on paper-and-pencil exercises related to the topic	0	1	2	3
c. Work on independent, long-term (at least one week) projects	0	1	2	3
d. Work on problems for which there is no immediately obvious method or solution	0	1	2	3
e. Develop technical or mathematical writing skills, including using equations, graphs, tables, and text together	0	1	2	3
f. Work on interdisciplinary lessons (e.g., writing journals in class)	0	1	2	3
g. Recite or drill orally..	0	1	2	3
h. Debate ideas or otherwise explain their reasoning..........	0	1	2	3
i. Complete a short test or quiz to review previous lesson	0	1	2	3
j. Use concrete models or manipulatives	0	1	2	3

4. About how often did students use the following as part of math instruction: (Circle one for each line.)

	Almost Never	Some Lessons	Most Lessons	Every Lesson
a. Standard calculators to solve basic exercises or problems..	0	1	2	3
b. Programmable calculators to solve advanced exercises or problems...	0	1	2	3
c. Graphing calculators to graph equations or data...	0	1	2	3
d. Calculators or computers to develop models or simulations..	0	1	2	3
e. Calculators or computers for data collection and analysis..	0	1	2	3
f. Computers for drill and practice on skill acquisition				
g. Computers to write reports...	0	1	2	3
h. Computers to access the Internet..................................	0	1	2	3

EXHIBIT E.1 (Continued)

Pedagogy Questions Drawn from Middle School Mathematics Survey

6. About how often did you interact with students in the targeted class in the fo llowing ways: (Circle one for each line.)

	Almost Never	Some Lessons	Most Lessons	Every Lesson
a. Students work individually without your ongoing assistance	0	1	2	3
b. Students work individually with your ongoing assistance	0	1	2	3
c. Work together as a class with students responding to one another	0	1	2	3
d. Work in pairs or small groups without your ongoing assistance	0	1	2	3
e. Work in pairs or small groups with your ongoing assistance	0	1	2	3

7. How important were the following assessment strategies in determining students' grades in this math class: (Circle one for each line.)

	Not Used	Minor Importance	Moderate Importance	Very Important
a. Objective tests (e.g., multiple choice) choice, true/false, short answer)	0	1	2	3
b. Essay tests........	0	1	2	3
c. Performance tasks or events	0	1	2	3
d. Systematic observation of students........	0	1	2	3
e. Math reports	0	1	2	3
f. Math projects	0	1	2	3
g. Homework assignments	0	1	2	3
h. Portfolios........	0	1	2	3

Overview of Analysis Strategy

We collected data on classroom teaching practice for the 1996–97, 1997–98, and 1998–99 school years, for all teachers who taught mathematics or science in the 30 case schools. In each wave of the survey, we asked each teacher to report on a specific class the teacher taught that school year.

We drew on these data to examine differences in teaching practice by school level and subject, as well as by district, school, and year. We conducted these analyses within a hierarchical linear modeling (HLM) framework. We treated the school, the teacher, and the class as levels in a three-level model, with classes nested within teachers within schools. This approach treats schools, teachers, and classes as "random effects," and we used our data to estimate the variance across schools, teachers, and classes in teaching practice. (See Exhibit E.2 for the equations.)

We treated the 10 districts, the school level (elementary, middle, and high), and the subject (mathematics and science) as fixed effects—that is, as conventional dummy-coded independent variables. We also treated year as a fixed effect, coded 0 for 1996–97, 1 for 1997–98, and 2 for 1998–99.[74] We included all possible interactions between school level, subject, and time.[75] We treated district and school level as school-level variables, subject as a teacher-level variable, and year and sponsorship as activity-level variables. (See Exhibit C.1.)

We considered treating district as a level in our HLM models, but the number of districts in our sample (10) is too small to provide a good estimate of the between-district variance component. Furthermore, the statistical methods we employed to estimate the hierarchical models assume that the random effects are normally distributed. Although this assumption is reasonable for schools, teachers, and activities, it seems less reasonable for districts, which were purposefully selected to include considerable variation in size, urbanicity, and district professional development policies.

[74] This assumes that the effect of year is captured by a linear trend. We estimated models including dummy variables for year, and the results differed only slightly from the linear models, which are easier to interpret and require fewer parameters.

[75] The treatment of subject and time as fixed effects implies that the effects of subject and time operate in the same way across schools. It would be possible to treat subject and time as random effects—that is, as effects that vary across schools—but given the sample size, this added complexity seemed unwarranted.

EXHIBIT E.2

Models for Aspects of Teaching Practice

The dependent variable, y_{tij}, is a characteristic of instruction (e.g., alignment) for a classroom at time *t*, for teacher *i*, in school *j.*

Model 1 (Three-level unconditional hierarchical model):

Level 1 (classroom): $y_{tij} = p_{0ij} + e_{tij}$.

Level 2 (teacher): $p_{0ij} = b_{00j} + n_{ij}$.

Level 3 (school): $b_{00j} = g_{000} + m_j$.

Model 2 (Three-level hierarchical model, conditional on school level and subject):

Level 1 (classroom): $y_{tij} = p_{0ij} + e_{tij}$.

Level 2 (teacher): $p_{0ij} = b_{00j} + b_{01j}s_{ij} + n_{ij}$,

where $s_{ij} =$ subject is coded math=0, science=1.

Level 3 (school): $b_{00j} = g_{000} + g_{001}e_j + g_{002}h_j + m_{0j}$;

$b_{01j} = g_{010} + g_{011}e_j + g_{012}h_j$,

where $e_j =$ elementary and $h_j =$ high school are coded 0/1 and middle is excluded.

Model 3 (Three-level hierarchical model, conditional on school level, subject, and time):

Level 1 (classroom): $y_{tij} = p_{0ij} + p_{1ij}time_{tij} + e_{tij}$,

where $time_{tij}$ is coded 0 for classroom teaching in 1996–97, 1 for 1997–98, and 2 for 1998—99.

Level 2 (teacher): $p_{0ij} = b_{00j} + b_{01j}s_{ij} + n_{0ij}$;

$p_{1ij} = b_{10j} + b_{11j}s_{ij}$

where $s_{ij} =$ subject is coded math=1, science=0.

Level 3 (school): $b_{00j} = g_{000} + g_{001}e_j + g_{002}h_j + m_{00j}$;

$b_{01j} = g_{010} + g_{011}e_j + g_{012}h_j$;

$b_{10j} = g_{100} + g_{101}e_j + g_{102}h_j$;

$b_{11j} = g_{110} + g_{111}e_j + g_{112}h_j$,

where $e_j =$ elementary and $h_j =$ high school are coded 0/1 and middle is excluded.

EXHIBIT E.2 (Continued)

Models for Aspects of Teaching Practice

Model 4 (Three-level hierarchical model, conditional on school level, subject, time, and district):

Level 1 (classroom): $y_{tij} = p_{0ij} + p_{1ij} time_{tij} + e_{tij}$,

where $time_{tij}$ is coded 0 for classroom teaching in 1996–97, 1 for 1997–98, and 2 for 1998—99.

Level 2 (teacher): $p_{0ij} = b_{00j} + b_{01j} s_{ij} + n_{0ij}$;

$p_{1ij} = b_{10j} + b_{11j} s_{ij}$,

where $s_{ij} =$ subject is coded math=1, science=0.

Level 3 (school): $b_{00j} = g_{000} + g_{001} e_j + g_{002} h_j + g_{003} districtA_j + ... + g_{00,10} districtI_j + m_{00j}$

$b_{01j} = g_{010} + g_{011} e_j + g_{012} h_j$

$b_{10j} = g_{100} + g_{101} e_j + g_{102} h_j$

$b_{11j} = g_{110} + g_{111} e_j + g_{112} h_j$

where $e_j =$ elementary and $h_j =$ high school are coded 0/1 and middle is excluded; and $districtA_j ... districtI_j$, representing the district in which school *j* is located, are coded 0/1.

The main purpose of our analysis of teaching practice was to examine changes that might be due to professional development or other policies and supports. To eliminate possible change in teaching practice because of changes in teachers' teaching assignments over the three years of the study, we restricted the analysis of the descriptive statistics reported in Exhibit 7 in Section III and Exhibit 18 in Section IV to teachers who reported on the *same course* in all *three* waves of the survey; we restricted our HLM analyses to teachers who reported on the same course for at least *two* waves of the survey.[76,77,78] Elementary teachers generally teach only one "course" each year—i.e., 3rd grade mathematics or science; thus, elementary teachers who did not teach at the same grade level for all three years were elimnated from the tables of descriptive statistics, and those who did not

[76] In the first wave of the survey, we asked each teacher to choose a course to report on that the teacher was likely to continue to teach each year. In the second and third waves of the survey, we printed the name of the course the teacher had reported on in the prior wave, to help the teacher select the same course, if possible.

[77] For teachers who taught the same course in two of the three years, we restricted the data to include only the two years containing the common course. We computed inter-year correlations on our teaching measures for teachers who taught the same course both years and those who taught different courses. As anticipated, the inter-year correlations were somewhat lower for teachers who changed courses than for those who did not.

[78] The HLM models we estimated, as specified in Exhibit E.2, do not require the same number of observations for each teacher. Thus, to maximize the power of our estimates, we included teachers with either two or three data points.

teach at the same level for at least two of the three years of the study were eliminated from the HLM analysis. Similarly, secondary teachers who did not teach a common course over at least three years (e.g., first-year algebra) were eliminated from the descriptive analysis, and those who did not teach a common course for at least two years were eliminated from the HLM analysis.

Overview of Analysis Results

The results of our analyses are reported in Exhibits E.3 (alignment), E.4 (memorization), E.5 (understanding), E.6 (performing procedures), E.7 (generating hypotheses), E.8 (collecting/analyzing data), E.9 (making connections), E.10 (didactic instruction), E.11 (individual seatwork), E.12 (active project-centered instruction), and E.13 (discussion-oriented instruction).

To illustrate the meaning of the parameter estimates, we discuss the results for alignment, displayed in Exhibit E.3. (The results displayed in the other exhibits are similar in form.) As described in Section III, alignment is a continuous variable measured on a scale from 0 to 1, where 0 indicates no alignment of a teacher's content coverage with the National Assessment of Educational Progress (NAEP) and 1 indicates perfect alignment. When multiplied by 100, the scale can be interpreted as the "percent alignment with the NAEP."

The fixed-effect parameters estimates are displayed in three groups containing the school-level, teacher-level, and classroom-level parameters. The results for Model 1 are displayed in the first column of the exhibit. Model 1 contains only one fixed effect, the intercept, which has a value of 0.191. The Model 1 intercept represents the value of alignment in the typical school in our sample.[79] Thus, on average, the alignment of teacher practice with the NAEP was about 19.1 percent.

The variance components shown at the bottom of the exhibit describe the between-school, between-teacher, and residual variation (variation between classes taught by the same teacher in different years). The between-school variance component of 0.0009 indicates that there is significant between-school variation in alignment ($p<.05$). As we discussed in Appendix C, one way to interpret the magnitude of the between-school variance is to compute its square root, which is the between-school standard deviation. The square root of 0.0009 is 0.03, which indicates that teachers in some schools are about 3 percent more aligned than the overall average of 19.1 percent, and teachers in other schools are about 3 percent less aligned.

The variation among teachers within schools (0.0056) is considerably larger than the between-school variance component. The square root of the variance is 0.075, which indicates that within a single school, some teachers may be 7.5 percent more aligned than the average for the school, whereas others are 7.5 percent less aligned.

The residual variation—or variation among the classes the same teacher teaches in different years—is 0.0028. The square root of this variance is 5.3 percent, which indicates that the courses a teacher teaches in different years may differ a good deal in their alignment with the NAEP.

---

[79] As described above, we treated school as a random effect. Thus, the intercept represents the average or typical value of content focus among schools in our sample; the variance components shown at the bottom of the exhibit indicate the variation among schools around this value.

The results for Model 2 are shown in the second column of the exhibit. The intercept of 0.149 is the average or typical value for schools in our sample, controlling for school level and subject taught. The coefficient for high school is negative and significant (-.037), which indicates that the content covered by high school teachers is on average less aligned with the NAEP than the content covered by middle school teachers (the reference category). We believe that this lower level of alignment for high school teachers is due to the nature of the 12th grade NAEP, which is intended as a measure of general high school mathematics and science, not of high-standards advanced academic coursework, such as pre-calculus or physics.

The coefficient for mathematics is positive and significant (.128), which indicates that the content covered by mathematics teachers is 12.8 percent more aligned with the NAEP than the content covered by science teachers. This undoubtedly reflects the fact that the the elementary and the secondary mathematics curricula are more consistent from school to school than the elementary science curriculum. There is also a significant interaction of subject and level (-.061 for high school mathematics), which indicates that although mathematics overall is more highly aligned than science, the alignment gap between mathematics and science is somewhat lower for high school science than for middle and elementary school science.

The variance components for Model 2 are shown at the bottom of the exhibit. The variance component for school has fallen from 0.0009 in Model 1 to 0.0000 in Model 2, which indicates that all of the between-school variation in Model 1 is due to school level; once the differences between elementary, middle, and high schools are accounted for, there is no remaining variation among schools. The between-teacher variation for Model 2 (0.0025) is also lower than the between-teacher variation for Model 1, but it remains significant, which indicates that there is significant variation in alignment among teachers who teach the same grade level and subject. The residual year-to-year variation in Model 2 is 0.0028, the same as the variation observed in Model 1.

The results for Model 3 are shown in the third column of the exhibit. Neither the coefficient for time nor any of the interactions of time with school level or subject are significant, which indicates that the average level of alignment did not change over the three years of the study. The variance components remain essentially as in Model 2.

The results for Model 4 are shown in the final column of the exhibit. The coefficient for one district (District B) is significant, which indicates that on average, teachers in District B are more aligned with the NAEP than teachers in District J, the reference category (.043). However, the overall effect of district was not significant in the model for alignment, which indicates that as a group, the 10 districts do not differ appreciably. (Significant effects of district, when they occur, are shown at the bottom of the table. See, for example, Exhibit E.5.)

Methods Used in Section II to Compute Percent Variation

In Exhibit 19 in Section IV, we report the percent of variation in teaching practices explained by differences across subjects and school levels, schools, and teachers in the same school, as well as residual unexplained variance. These variance estimates are derived from the results for Models 1, 2, and 3. The total variance to be explained is computed as the total variance for Model 1—that is, the sum of the variance between schools, the variance between teachers, and the residual variance for Model 1 ($m_{00j} + n_{0ij} + e_{tij}$). The variance from subject and level is computed as the difference between the total variance for Model 1 ($m_{00j} + n_{0ij} + e_{tij}$) and the total variance for Model 2 ($m_{00j} + n_{0ij} + e_{tij}$), which controls for subject and level.[80] The variance from years is computed as the difference between the total variance ($m_{00j} + n_{0ij} + e_{tij}$) for Model 2 and the total variance for Model 3, which controls for subject, level, and year (1997–98 and 1998–99). The variance from school and teacher and the residual unexplained variance are the variance component estimates for Model 3 (m_{00j}, n_{0ij}, and e_{tij}).[81,82,83]

To illustrate these computations, we turn again to alignment (Exhibit E.3). The total variance in alignment to be explained is the sum of the three variance comonents for Model 1 (0.0009 + 0.0056 + 0.0028 = 0.0093). The variance explained by school and level is the difference between 0.0093 and the total variance for Model 2 (0.0000 + 0.0025 + 0.0028 = 0.0053), or 0.0040. In percentage terms, the variance explained by school level and subject is 0.0040 divided by 0.0093, or 43 percent. Similarly, the variance explaind by year is the difference between the total variance for Model 2 and the total variance for Model 3. In this case, the total variance rose slightly with the addition of year, from 0.0053 for Model 2 to 0.0054 for Model 3. As discussed in footnote 10, this increase is a result of the fact that the addition of year to the Model involved the estimation of six new parameters for year and its interactions, all of which are very close to zero. We therefor treat the variance from year as zero.

Finally, the percent variance explained by schools is the variance component for schools in Model 3 divided by the total variance for Model 1 (0.0000 divided by 0.0093), or zero percent. The variance component for teachers is 0.0025 divided by the total variance for Model 1 (0.0025 divided by 0.0093), or 26.9 percent. The percent variation unexplained is the residual variance for Model 3 divided by the total variance for Model 1 (0.0025 divided by 0.0093), or 30.1 percent.

---

[80] This is analogous to the conventional percent of variation explained in an OLS regression model.

[81] The variance explained by the fixed effects is a function of the order in which the variance estimates are derived. We computed the variance from school level and subject prior to computing the variance from year; if we had chosen the reverse order, the variance from year might be slightly higher, and the variance from school level and subject slightly lower.

[82] Because we treated district as a fixed effect, we did not obtain direct estimates of the between-district variance component. It would be possible to derive an estimate of the percent of variance between districts from the variance for Models 3 and 4. We elected not to report these estimates, however, because, as we discussed above, the between-district variance is not very meaningful or reliable, given the small number of districts in the sample.

[83] We expected that the variance between schools and teachers and the residual variance would decline from Model 1 to Model 5, as additional control variables were added to the model. In a few cases, however, the variance increased with the addition of controls. This increase tended to occur when none of the added controls was significant, which suggests that any added explanatory power from the controls was more than offset by the loss of degrees of freedom. Given the method we used to compute the variance explained by fixed effects, an increase in variance from the addition of fixed effects would result in a negative variance explained. We treated the few negative estimates as zero.

EXHIBIT E.3

Alignment: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	0.191***	0.149***	0.157***	0.138***
Elementary		.002	-.011	-.007
High school		-.037**	-.047**	-.049**
District A				.026
District B				.043**
District C				.008
District D				.002
District E				.020
District F				.013
District G				.022
District H				.024
District I				.021
Teacher-level				
Math		.128***	.130***	.130***
Math*Elementary		.010	-.001	-.002
Math*High school		-.061***	-.056*	-.056*
Classroom-level				
Time			-.009	-.008
Time*Elementary			.014	.014
Time*High school			.011	.010
Time*Math			-.003	-.003
Time * Elementary*Math			-.008	-.008
Time*High school*Math			-.005	-.005
Variance components				
Between-school	.0009*	.0000	.0000	.0000
Between-teacher	.0056***	.0025***	.0026***	.0025***
Residual	.0028	.0028	.0028	.0028

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level: Model 2 (p<.001), Model 3 (p<.001), Model 4 (p<.001)
Significant effect of level*subject: Model 2 (p<.001), Model 3 (p<.01), Model 4 (p<.01)
N=845

EXHIBIT E.4

Emphasis Given to Memorization: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	.155***	.149***	.165***	.137***
Elementary		.003	.002	.003
High school		.012	-.004	-.008
District A				.045**
District B				.034*
District C				.025
District D				.025
District E				.024
District F				.032*
District G				.027
District H				.036*
District I				.037*
Teacher-level				
Math		.001	-.014	-.014
Math*Elementary		.007	.007	.006
Math*High school		-.014	.004	.004
Classroom-level				
Time			-.016**	-.016**
Time*Elementary			.002	.002
Time*High school			.019**	.012**
Time*Math			.016	.016
Time * Elementary*Math			.000	-.000
Time*High school*Math			-.019	-.019
Variance components				
Between-school	.0001*	.0001	.0001	.0002
Between-teacher	.0018***	.0018***	.0018***	.0018***
Residual	.0027	.0027	.0027	.0027

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of time*subject: Model 3 (p<.01), Model 4 (p< .01)
N=846

EXHIBIT E.5

Emphasis Given to Understanding: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	.223***	.210***	.211***	.218***
Elementary		.011	.013	.015
High school		.016	.022	.018
District A				.005
District B				-.013
District C				-.007
District D				-.042***
District E				.006
District F				.014
District G				-.018
District H				.011
District I				.001
Teacher-level				
Math		.011	.016	.014
Math*Elementary		-.011	-.017	-.017
Math*High school		-.000	-.001	.001
Classroom-level				
Time			-.002	-.002
Time*Elementary			-.002	-.002
Time*High school			-.006	-.001
Time*Math			-.006	-.005
Time * Elementary*Math			.007	.008
Time*High school*Math			.001	.001
Variance components				
Between-school	.0001	.0001	.0001	.0000
Between-teacher	.0012***	.0012***	.0013***	.0013***
Residual	.0027	.0027	.0027	.0027

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level: Model 3 (p<.05)
Significant effect of district: Model 4 (p<.01)
N=843

EXHIBIT E.6

Emphasis Given to Performing Procedures: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	.191***	.154***	.143***	.151***
Elementary		.003	.011	.013
High school		.011	.026**	.027**
District A				-.001
District B				-.008
District C				-.010
District D				-.004
District E				-.008
District F				-.020
District G				-.008
District H				-.020
District I				.003
Teacher-level				
Math		.074***	.088***	.087***
Math*Elementary		-.030**	-.044***	-.044***
Math*High school		.010	-.007	-.006
Classroom-level				
Time			.011*	.011*
Time*Elementary			-.009	-.008
Time*High school			-.015**	-.015**
Time*Math			-.015*	-.015*
Time * Elementary*Math			.014+	.014
Time*High school*Math			.018*	.018*
Variance components				
Between-school	.0000	.0000	.0000	.0000
Between-teacher	.0024***	.0012***	.0012***	.0012***
Residual	.0018	.0019	.0019	.0019

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level: Model 2(p<.001), Model 3 (p<.001), Model 4 (p<.001)
Significant effect of level*subject Model 2 (p<.001), Model 3 (p<.001), Model 4 (p<.001)
N=845

EXHIBIT E.7

Emphasis Given to Generating Hypotheses: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	.119***	.140***	.140***	.139***
Elementary		-.009	-.014	-.018*
High school		-.025***	-.030***	-.023**
District A				-.010
District B				.004
District C				.004
District D				.025**
District E				-.005
District F				-.001
District G				.010
District H				.009
District I				-.021*
Teacher-level				
Math		-.032***	-.039***	-.038***
Math*Elementary		.018	.030**	.030**
Math*High school		.019	.021	.019
Classroom-level				
Time			-.000	.000
Time*Elementary			.006	.005
Time*High school			.006	.005
Time*Math			.008	.008
Time * Elementary*Math			-.013	-.013
Time*High school*Math			-.003	-.002
Variance components				
Between-school	.0000	.0000	.0000	.0000
Between-teacher	.0013***	.0012***	.0012***	.0011***
Residual	.0016	.0016	.0016	.0016

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level: Model 2 (p<.01), Model 3 (p<.01), Model 4 (p<.05)
N=846

EXHIBIT E.8

Emphasis Given to Collecting Data: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	.133***	.154***	.157***	.156***
Elementary		-.009	-.022**	-.023**
High school		-.009	-.010	-.006
District A				-.013
District B				.008
District C				.004
District D				.023**
District E				-.009
District F				-.011
District G				.016
District H				.004
District I				-.012
Teacher-level				
Math		-.032***	.042***	-.041***
Math*Elementary		.013	.036**	.036**
Math*High school		.008	-.005	-.006
Classroom-level				
Time			-.003	-.002
Time*Elementary			.013**	.013**
Time*High school			.001	.001
Time*Math			.011	.010
Time * Elementary*Math			-.025***	-.025***
Time*High school*Math			-.004	-.003
Variance components				
Between-school	.0000	.0000	.0000	.0000
Between-teacher	.0013***	.0010***	.0010***	.0010***
Residual	.0016	.0016	.0016	.0016

Significant effect of level*subject: Model 3 (p<.001), Model 4 (p<.001)
Significant effect of time*level*subject: Model 3 (p<.001), Model 4 (p<.001)
Significant effect of district: Model 4 (p<.01)
N=845

EXHIBIT E.9

Emphasis Given to Making Connections: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	.173***	.191***	.188***	.204***
Elementary		-.012	-.012	-.015
High school		-.015*	-.019*	-.018
District A				-.023**
District B				-.019*
District C				-.010
District D				-.011
District E				-.008
District F				-.020*
District G				-.016
District H				-.032***
District I				.000
Teacher-level				
Math		-.021**	-.015	-.014
Math*Elementary		.013	.010	-.010
Math*High school		.001	.005	.004
Classroom-level				
Time			.003	.003
Time*Elementary			.000	.000
Time*High school			.005	.005
Time*Math			-.007	-.008
Time * Elementary*Math			.004	.004
Time*High school*Math			-.004	-.004
Variance components				
Between-school	.0000	.0000	.0000	.0000
Between-teacher	.0001***	.0001***	.0008***	.0008***
Residual	.0018	.0018	.0018	.0018

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level: Model 2 (p<.05)
Significant effect of time*subject: Model 3 (p<.05), Model 4 (p<.05)
N=844

EXHIBIT E.10

Emphasis Given to Didactic Instruction: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	50.074***	48.227***	43.389***	48.427***
Elementary		.138	.400	.413
High school		4.156***	3.582***	3.368***
District A				1.263
District B				-.613
District C				2.635*
District D				-.877
District E				.803
District F				-.410
District G				-1.862
District H				-1.589
District I				.372
Teacher-level				
Math		3.187***	3.507***	-.332
Math*Elementary		-5.121***	-5.701***	.633
Math*High school		-.719	-.843	.124
Classroom-level				
Time			-.168	-.184
Time*Elementary			-.280	-.263
Time*High school			.614	.621
Time*Math			-.347	-.332
Time * Elementary*Math			.643	.633
Time*High school*Math			.130	.124
Variance components				
Between-school	9.884***	2.178**	2.201**	1.328
Between-teacher	15.616***	14.409***	14.438***	14.403***
Residual	9.698	9.696	9.674	9.67

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level: Model 2 (p<.001), Model 3 (p<.001), Model 4 (p<.001)
Significant effect of level*subject: Model 2 (p<.001), Model 3 (p<.001), Model 4 (p<.001)
Significant effect of time*level: Model 3 (p<.05), Model 4 (p<.05)
N=942

EXHIBIT E.11

Emphasis Given to Individual Seatwork: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	50.061***	49.015***	49.581***	49.093***
Elementary		-2.12**	-2.156**	-2.320**
High school		.736	.190	.452
District A				.990
District B				1.054
District C				-1.227
District D				3.672**
District E				-.418
District F				-1.816
District G				1.156
District H				2.012
District I				-.474
Teacher-level				
Math		2.837***	1.383	1.410
Math*Elementary		1.323	2.524*	2.482*
Math*High school		-.268	1.505	1.451
Classroom-level				
Time				-.583
Time*Elementary				.017
Time*High school				.552
Time*Math				1.539**
Time * Elementary*Math				-1.247
Time*High school*Math				-1.858**
Variance components				
Between-school	2.785**	2.936**	2.919**	1.817
Between-teacher	15.639***	12.384***	12.335***	12.360***
Residual	11.994	12.019***	11.951***	11.939

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level: Model 2 (p<.01), Model 4 (p<.01)
Significant effect of time*level*subject: Model 3 (p<05), Model 4 (p<.05)
Significant effect of district: Model 4 (p<.05)
N=934

EXHIBIT E.12

Emphasis Given to Active, Project-Centered Instruction: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	50.068***	53.924***	54.088***	52.467***
Elementary		-2.819***	-3.093***	-3.179***
High school		-1.215	-1.287	-.865
District A				-.367
District B				1.572
District C				.568
District D				4.755**
District E				1.931
District F				.927
District G				2.835
District H				1.185
District I				2.556
Teacher-level				
Math		-7.290***	-8.145***	-8.086***
Math*Elementary		4.981***	6.472***	6.428***
Math*High school		1.491	2.366	2.284
Classroom-level				
Time			-.166	-.157
Time*Elementary			.274	.265
Time*High school			.078	.076
Time*Math			.916	.912
Time *Elementary*Math			-1.527**	-1.547*
Time*High school*Math			-.932	-.937
Variance components				
Between-school	1.708	2.784**	2.919**	2.957
Between-teacher	18.129***	11.657***	11.582***	11.445***
Residual	12.406	12.425	12.445	12.445

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level*subject: Model 2 (p<.001), Model 3 (p<.001), Model 4 (p<.001)
N=942

EXHIBIT E.13

Emphasis Given to Discussion-Oriented Instruction: Effects of Level, Subject, Year, and District

Coefficient	Model 1: Intercept only	Model 2: School level and subject	Model 3: Year	Model 4: District
School-level				
Intercept	49.090***	48.481***	47.841***	46.868***
Elementary		3.686**	3.806**	4.030*
High school		*1.799	-1.544	-1.856
District A				1.353
District B				1.390
District C				.613
District D				.081
District E				3.916*
District F				-.881
District G				-1.715
District H				1.602
District I				3.105
Teacher-level				
Math		.444	.133	.189
Math*Elementary		-3.349*	-1.846	-1.926
Math*High school		1.273	3.049	2.884
Classroom-level				
Time			.669	.663
Time*Elementary			-.109	-.092
Time*High school			-.261	-.281
Time*Math			.363	.352
Time * Elementary*Math			-1.537	-1.524
Time*High school*Math			-1.870	-1.830
Variance components				
Between-school	4.583*	3.772*	3.781*	3.169
Between-teacher	25.505***	24.183***	24.414***	24.453***
Residual	48.594	48.661	48.509	48.540

+ p<.10, * p<.05, ** p<.01, *** p<.001
Significant effect of level: Model 2 (p<.01), Model 3 (p<.01), Model 4 (p<.01)
Significant effect of level*subject: Model 2 (p<.01), Model 3 (p<.05), Model 4 (p<.05)
N=916