NCEE 2010-4010 U.S. DEPARTMENT OF EDUCATION

Middle School Mathematics Professional Development Impact Study

Findings After the First Year of Implementation

Executive Summary

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

Middle School Mathematics Professional Development Impact Study

Findings After the First Year of Implementation

Executive Summary

April 2010

Michael S. Garet
Andrew J. Wayne
Fran Stancavage
James Taylor
Kirk Walters
Mengli Song
Seth Brown
Steven Hurlburt
American Institutes for Research

Pei Zhu
Susan Sepanik
Fred Doolittle
MDRC

Elizabeth Warner
Project Officer
Institute of Education Sciences

NCEE 2010-4010
U.S. DEPARTMENT OF EDUCATION

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
John Q. Easton
Acting Commissioner

April 2010

This report was prepared for the Institute of Education Sciences under Contract No. ED-04-CO-0025/0005. The project officer was Elizabeth Warner in the National Center for Education Evaluation and Regional Assistance.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 22207, Alexandria, VA 22304.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327. Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 703-605-6794 or order online at www.edpubs.gov.

This report also is available on the IES website at http://ncee.ed.gov.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-205-8113.

ACKNOWLEDGMENTS

This study represents a collaborative effort of school districts, schools, teachers, researchers, and professional development providers. We appreciate the willingness of the school districts, schools, and teachers to join the study, participate in the professional development, and respond to requests for data, feedback, and access to classrooms. We were also fortunate to have the advice of our Expert Advisory Panel: Sybilla Beckmann, University of Georgia; Julian Betts, University of California, San Diego; Doug Carnine, University of Oregon; Mark Dynarski, Mathematica Policy Research; Lynn Fuchs, Vanderbilt University; Russell Gersten, Instructional Research Group; Kenneth Koedinger, Carnegie Mellon University; Brian Rowan, University of Michigan; John Woodward, School of Education, University of Puget Sound; and Hung-Hsi Wu, University of California, Berkeley. We also appreciated the advice of Hyman Bass, University of Michigan, and others associated with the Learning Mathematics for Teaching project. We also appreciated the advice of W. James Lewis, University of Nebraska – Lincoln, and Andrew Porter, University of Pennsylvania. We also benefitted from the informed feedback on the study's statistical analyses and report from the following people at the American Institutes for Research (AIR) and MDRC: Howard Bloom, Gordon Berlin, George Bohrnstedt, Matthew Gushta, Rob Ivry, Pamela Morris, Marie-Andree Somers, Gary Phillips, and Shelley Rappaport.

We would like to thank all those who provided the professional development during the study, including the facilitators at America's Choice and Pearson Achievement Solutions, as well as the members of the treatment team who provided monitoring support—Steve Leinwand and Meredith Ludwig. We also thank those who served as site coordinators: Midori Hargrave, Jack Rickard, and several staff who served in these roles in the first year of implementation. We also thank Delphinia Brown, Suzannah Herrmann, and Amber Noel for coordinating the classroom observations and data processing, and Edith Tuazon and for her support of those efforts and assistance with project communications. We appreciated the excellent assistance of Jeanette Moses in multiple roles across the project. We also thank Lynne Blankenship and the conference staff for all their support in managing many of the study's professional development activities; Collin Payne for his excellent research assistance with the student records; all of the staff at REDA International, Inc., MDRC, Westat, and AIR who helped us collect and process data throughout the study; and the AIR and MDRC staff who helped us start the study up during the early years: Robert Ivry, Stephanie Safran, Kristin Porter, and Christian Geckeler. Finally, we would like to thank our report editors, Holly Baker, Lisa Knight, Patti Louthian, and Sharon Smith, who helped make the report useful and understandable.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST[1]

The research team for this study consisted of a prime contractor, American Institutes for Research (AIR), and three subcontractors, MDRC, REDA International, Inc., and Westat, Inc. None of these organizations or their key staff has financial interests that could be affected by findings from the Middle School Mathematics Professional Development Impact Study. No one on the 10-member Expert Advisory Panel, convened by the research team annually to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

[1] Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the particular tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

EXECUTIVE SUMMARY

Middle School Mathematics Professional Development Impact Study

This report presents interim results from the Middle School Mathematics Professional Development Impact Study, which is sponsored by the Institute of Education Sciences (IES). The report presents results immediately following 1 year of the study's professional development. A future report will present results following 2 years of professional development.

Student achievement in mathematics has been a focal concern in the United States for many years. The National Research Council's 2001 report and the recent report of the National Mathematics Advisory Panel (2008) both called attention to student achievement in mathematics, and both called for all students to learn algebra by the end of eighth grade. Reports have argued, further, that achieving this goal requires that students first successfully learn several topics in rational numbers—fractions, decimals, ratio, rate, proportion, and percent. These topics are typically covered in grades 4 through 7, yet many students continue to struggle with them beyond the seventh grade. The National Mathematics Advisory Panel wrote that "difficulty with fractions (including decimals and percent) is pervasive and is a major obstacle to further progress in mathematics, including algebra" (p. xix). The panel also specified that by the end of seventh grade, "students should be able to solve problems involving percent, ratio, and rate, and extend this work to proportionality" (p. 20).

The U.S. Department of Education's National Center for Educational Evaluation and Regional Assistance (NCEE)—within the Institute of Education Sciences—initiated the Middle School Mathematics Professional Development Impact Study to test the impact of a professional development (PD) program for teachers that was designed to address the problem of low student achievement in topics in rational numbers.[2] The study focuses on seventh grade, the culminating year for teaching those topics. The study is being conducted by the American Institutes for Research (AIR) and MDRC together with their evaluation partners REDA International and Westat.

Currently, through the Elementary and Secondary Education Act, the federal government provides significant resources for PD, but little rigorous evidence is available on the impact of PD on teacher and student outcomes.[3] Hundreds of studies have addressed the topic of teacher learning and PD (for reviews, see Borko 2004; Clewell, Campbell, and Perlman 2004; Kennedy 1998; Richardson and Placier 2001; Supovitz 2001; Yoon, Duncan, Lee, Scarloss, and Shapley 2007).[4] The most recent review of studies of the impact of teacher PD on student achievement revealed a total of nine studies that have rigorous designs—randomized control trials (RCTs) or certain quasi-experimental designs (QEDs)—that allow causal inferences to be made (Yoon et al. 2007). Four of

[2] The professional development focused on positive rational numbers. The decision to restrict the focus to positive rational numbers was based on advice from the study's external advisors, who suggested that including negative rational numbers would broaden the scope of the content beyond what could be addressed in the allotted time for the PD program.

[3] In the 2001 reauthorization of the Elementary and Secondary Education Act of 1965 (ESEA), the Congress expanded the federal resources available for teacher professional development by establishing—under Title II, Part A—the Improving Teacher Quality State Grants program. The grants program provides support for activities designed to ensure an adequate supply of knowledgeable teachers, and states and school districts spent $529 million of Title II, Part A funds on teacher professional development, according to an analysis of spending for the 2004-2005 school year. An even more widely used source of funds for teacher professional development is Title I, through which states and districts spent $988 million for teacher professional development in 2004-2005 (Birman et al 2007, p. 69). ESEA requires that schools that have been identified for improvement spend at least 10 percent of their Title I allocations on professional development (Title I, Part A, Section 1116(b)(3)(A)(i)).

[4] For example, Yoon et al. (2007) alone identified 1,343 studies of PD.

the nine studies focused on the effect of a PD program on mathematics achievement, and none focused on mathematics at the middle school level.

The Middle School Mathematics PD Impact Study is the first rigorous test of the impact of a PD program focused on teachers of middle school mathematics. Within 12 participating school districts, the study randomly assigned 77 mid- and high-poverty schools to treatment and control conditions and collected outcome data on teachers and students. The PD was delivered by two provider organizations, each of which served the treatment schools in six of the 12 participating districts. Seventh-grade teachers in the treatment schools had the opportunity to receive the PD program offered by the study and could also continue to participate in the PD activities that they would have received in the absence of the study. Seventh-grade teachers in the control schools received only the PD that they would have received in the absence of the study.

The study has three central research questions:

1. What impact did the PD program provided in this study have on teacher knowledge of rational number topics?
2. What impact did the PD program provided in this study have on teacher instructional practices?
3. What impact did the PD program provided in this study have on student achievement in rational number topics?

The study produced the following results:

- **The study's PD program was implemented as intended.** The PD providers delivered an average of 67.6 hours of PD per site, compared to 68 hours intended, and the treatment group teachers attended an average of 83 percent of the PD that was delivered. In surveys given to treatment and control group teachers, treatment group teachers reported participating in 55.4 hours more mathematics-related PD than the control group teachers.
- **The PD program did not produce a statistically significant impact on teacher knowledge of rational numbers (effect size = 0.19, p-value = 0.15).** On average, 54.7 percent of teachers in the treatment group answered test items of average difficulty correctly, compared with 50.1 percent for teachers in the control group.
- **The PD program had a statistically significant impact on the frequency with which teachers engaged in activities that elicited student thinking, one of the three measures of instructional practice used in the study (effect size = 0.48).** This measure encompasses such behaviors as asking other students whether they agree or disagree with a particular student's response and also includes behaviors elicited from the students such as offering additional justifications or strategies. Treatment teachers on average engaged in 1.03 more activities per hour that elicited student thinking. The PD program did not produce a statistically significant impact on the other two measures of instructional practice: *Teacher uses representations* (effect size = 0.30; p-value = 0.0539) and *Teacher focuses on mathematical reasoning* (effect size = 0.19; p-value = 0.32).
- **The PD program did not produce a statistically significant impact on student achievement (effect size = 0.04, p-value = 0.37).**

Overview of the PD Program

The PD program delivered in this study was designed to develop teachers' capability to teach positive rational number topics effectively. The PD program consisted of 68 contact hours, all addressing rational number topics, which is more PD in mathematics than most mathematics teachers typically receive in a single year.[5,6] The PD included a 3-day summer institute and a series of 1-day follow-up seminars held during the school year, with in-school coaching following each seminar day. Within that structure, the specification of the PD program was guided by the literature, which is largely based on correlational research and practitioner experience.[7]

Within each topic in rational numbers, the PD program focused on two aspects of teachers' content knowledge. The first, common knowledge of mathematics (CK), is the knowledge of topics in rational numbers that students should ideally have after completing the seventh grade. This knowledge includes computational or procedural skills, conceptual understanding, and problem-solving skills in rational number topics.

The second aspect of teachers' content knowledge emphasized in the PD, specialized knowledge of mathematics for teaching (SK), is additional knowledge of rational numbers that may be useful for teaching rational number topics. For example, SK includes identifying the key mathematical understanding within a topic or problem, identifying common errors that occur in student work, and choosing useful representations and explanations when teaching rational numbers.

The summer institute and seminars blended activities intended to develop specialized knowledge of mathematics for teaching and to strengthen common knowledge of mathematics. The institutes and seminars were designed to use multiple delivery formats to provide teachers a variety of learning opportunities. The planned PD activities included opportunities for teachers to solve mathematics problems individually and in groups, make short oral presentations to explain how they solved problems, receive feedback on how they solved and presented their solutions, engage in discussions about the most common student misconceptions associated with topics in rational numbers, and plan lessons that they would teach during the follow-up coaching visits.

The primary purpose of the coaching component of the PD program was to help teachers apply material covered in the institutes and seminars to their classroom instruction. The coaching component was designed to consist of 10 days of coaching provided through five 2-day visits to each school. During the coaching visits at each school, the facilitators focused their activities on the school's seventh-grade mathematics teachers. Each 2-day coaching visit was designed to occur immediately after one of the 5 seminar days and to link to the preceding seminar, using both individual and group activities.

---

5 Sixty-eight hours is the number of contact hours provided during the first year of the PD program, which is the focus of this report. Additional contact hours were provided in the second year of the PD program.

6 A national survey of teachers completed in 2005–2006 found that 11 percent of elementary teachers and 22 percent of secondary teachers assigned to teach mathematics participated in professional development in mathematics lasting more than 24 hours (U.S. Department of Education 2009, p. 95).

7 In the nine rigorous studies identified by Yoon et al. (2007), the variation in the features of the PD programs that were tested was not sufficient to draw conclusions about the characteristics of the PD programs that were effective. For example, across the nine studies, all PD programs were delivered in the form of a workshop or a summer institute, along with some form of follow-up support.

Using the common structure, content, and other parameters described above, two providers selected through a competitive process delivered the PD program: America's Choice and Pearson Achievement Solutions. Both providers built on their existing materials that addressed topics in rational numbers. Facilitator guides were refined through a year-long pilot and review process. The study's external advisors reviewed both providers' facilitator guides, focusing on the accuracy, appropriateness, and coherence of the mathematics content presented to teachers.

Study Design

The Middle School Mathematics PD Impact Study was conducted in 12 districts. The study used an experimental design with random assignment of schools to treatment and control conditions within each participating district. The difference in outcomes between the treatment schools and the control schools can be interpreted as the effect of the study's PD model relative to "business as usual" in each participating district.

Study Sample

The study focused on districts using one of three specific mathematics curricula so that the PD could be designed to be relevant to the curricula that teachers were using in their classrooms. The three curricula were identified by determining the most commonly used curricula in the districts that met the study's size criteria. The most commonly used curricula fell into two categories. The sample was therefore constructed to form two parallel substudies of the same design but in different curricular contexts.[8] One substudy took place in 6 districts using either *Glencoe McGraw-Hill Mathematics: Applications and Concepts* or *Prentice Hall Mathematics* (referred to jointly as *Glencoe/PH Mathematics*); a parallel substudy took place in 6 districts using *Connected Mathematics* (*CMP*). The two categories of curricula differ in organization, lesson components, instructional approaches supported, and content emphasized, so the impact of the PD may differ by curriculum type.

Each of the two PD providers—America's Choice and Pearson Achievement Solutions—was assigned to work with 6 of the 12 districts participating in the study. Providers were assigned to districts to balance the allocation of districts using *Glencoe/PH Mathematics* and *CMP* across providers.[9] Thus, as shown in Table ES-1, the 6 districts using *Glencoe/PH Mathematics* were split between the two providers (three for America's Choice and three for Pearson Achievement Solutions), and the six districts using *CMP* were similarly split, so that the effect of the PD in either curricular context would be derived from the services of both organizations.

Twelve eligible districts in nine states agreed to participate in the study. Each district provided 4 to 8 study schools, producing a total sample size of 77 schools. Within these schools, the spring 2008 analysis sample included 195 teachers and 11,479 students, distributed across treatment and control groups as shown in Table ES-2.

---

[8] Although the study was conducted in two identifiable curricular contexts, the study is not designed to test the effectiveness of the mathematics curricula used in the participating districts. Rather, it is a study of the impact of the specific PD program used.

[9] Note that the assignment of districts to providers was not random. Among the districts using *Glencoe* and *PH Mathematics*, we assigned the three districts using *Glencoe* to America's Choice and the three districts using *PH Mathematics* to Pearson Achievement Solutions on the basis of providers' prior experiences working with those curricula. Among the districts using *CMP*, we took into account the geographic proximity of provider staff to the study districts.

Table ES-1. Allocation of the 12 Study Districts Across PD Providers and Across Mathematics Curricula

	Professional Development Provider	
	America's Choice	**Pearson Achievement Solutions**
Mathematics Curriculum		
Glencoe/PH Mathematics [a]	3 Districts	3 Districts
CMP	3 Districts	3 Districts

NOTES: [a] America's Choice served the three districts that used *Glencoe*. Pearson Achievement Solutions served the three districts that used *PH Mathematics*.

Table ES-2. Number of Schools, Teachers, and Students in Spring 2008 Impact Analysis Sample, Overall and Treatment Status

		Number of Seventh-Grade Teachers		Number of Seventh-Grade Students	
Treatment Status	**Number of Schools**	**Total Number**	**Average Per School**	**Total Number**	**Average Per School**
Treatment	40	100	2.5	5,858	146.4
Control	37	95	2.5	5,621	151.9
Total	**77**	**195**	**2.5**	**11,479**	**149.0**

SOURCE: Teacher Rosters; District Enrollment Records.

All eligible teachers teaching at least one regular seventh-grade mathematics class in each school in the 2007–2008 school year were members of the teacher sample for the study, and all seventh-grade students in their regular seventh-grade mathematics classes were members of the student sample.[10,11] This definition of the teacher and student samples implies that the study is a test of the impact of mandatory PD, as opposed to PD selected by individual teachers.

The 77 study schools are from all four regions of the United States, and they are predominantly in large or mid-sized cities, as shown in Table ES-3. The average rate of student eligibility for free or reduced-price lunch was 66 percent, and 77 percent of the schools were designated Title I schools. In study classrooms, in fall 2007, average student performance on a computer-adaptive test of rational numbers content used in the study was at the 19th percentile, relative to all test takers in the data base maintained by the test developer.

[10] "Eligible teachers" are defined as regular teachers, not short-term substitutes. (Long-term substitutes were included.)

[11] At each school, the study focused on seventh-grade teachers who taught regular, middle-track seventh-grade mathematics classes. This focus excluded advanced classes, such as gifted and talented programs and algebra, as well as remedial classes and self-contained special education classes.

Table ES-3. School Background Characteristics for Study Sample Schools and Eligible Schools in Large Districts

Characteristics	Study Sample	Eligible Schools in Large Districts[a]
Geographic Region (percent of schools)		
Northeast	18.2	8.8*
South	53.2	55.8
Midwest	11.7	9.0
West	16.9	26.4
Urbanicity (percent of schools)		
Large or Middle-Sized City	76.6	59.1*
Urban Fringe and Large Town	18.2	30.7*
Small Town and Rural Area	5.2	10.2
Title I Status (percent of schools)	76.6	67.8
Free and Reduced-Price Lunch (school average percent of students)	66.4	65.3
Race/Ethnicity (school average percent of students)		
White	33.7	27.9*
Black	36.2	31.1
Hispanic	24.7	33.5*
Asian	2.7	5.5*
Other	1.2	0.9
Male (school average percent of students)	50.7	50.7
Total School Enrollment	754.9	919.5*
Number of Seventh-Grade Students	232.3	310.9*
Number of Full-Time-Equivalent Teachers (All Grades)	45.9	54.9*
School Type (percent of schools)[b]		
Middle School Only	81.8	95.2*
Elementary and Middle	16.9	2.9*
Middle and High	1.3	1.7
Elementary and Middle and High	0.0	0.2

Sample Size: N = 77 schools in study sample; 2,710 eligible schools.

SOURCE: 2006–2007 *Common Core of Data* (CCD).

NOTES: [a] This sample was restricted to schools in districts that satisfy the following criteria: there were at least four regular schools with at least 150 seventh-grade students each, and the percentage of students eligible for free or reduced-price lunch was at least 33 percent for the whole school.

[b] To classify school type, preK–grade 3 are considered elementary school grades, grades 4–9 are considered middle school grades, and grades 10–12 are considered high school grades.

Percentage values for characteristics with multiple categories may not sum to 100 due to rounding.

Statistical significance was determined based on t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

Table ES-4. Teacher Background Characteristics for Study Sample Teachers and Teachers in Eligible Schools in Large Districts

Description of Mathematics Teachers of Seventh-Grade Students	Study Sample	Eligible Schools in Large Districts
Standard Certification (percent)	76.6	73.4
Bachelors Degree (percent)a	100.0	100.0
Masters Degree (percent)a	34.8	40.7
Mathematics Major (percent)	12.8	29.3
Mathematics-Related Major (percent)	11.2	16.2
Years of Teaching Experience (percent)		
3 years or fewer	30.3	37.4
4–10 years	31.9	26.9
11–20 years	23.9	15.7
More than 20 years	13.8	20.1
Sample Size: N = 188 teachers in study sample; 10,700 teachers in eligible schools.		

SOURCE: Fall 2007 Teacher Survey (Teacher Baseline Analysis Sample); 2003–2004 *Schools and Staffing Survey* (SASS), Public School Teacher Data Files.

NOTES: [a]N = 187 teachers.

Percentage values for characteristics with multiple categories may not sum to 100 due to rounding.

Statistical significance was determined based on t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

On some key characteristics, the study sample schools were statistically different from the pool of eligible schools from which they were selected. The study sample schools were significantly more likely to be in the Northeast region and to be located in large- or middle-sized cities. The students in the study sample schools were more likely to be White and less likely to be Hispanic or Asian. Study schools enrolled fewer seventh-grade students and had fewer teachers than did eligible schools; study schools were also more likely than eligible schools to combine elementary and middle grades.

Despite these differences, the teachers in study schools were not statistically distinguishable from those teaching seventh-grade mathematics in the pool of eligible schools from which the study schools were selected, on any of teacher characteristics presented in Table ES-4.

Data Collection and Outcome Measures

Data were collected from teachers and students in the study schools in the fall, winter, and spring of the 2007–2008 school year. The three main outcome measures were constructed as follows:

- **Teacher knowledge of rational numbers content and pedagogy.** Teacher knowledge was measured for all treatment and control teachers using a specially constructed *teacher knowledge test*. The test was first administered in summer 2007 to treatment teachers and fall 2007 to control teachers to provide descriptive information on the sample and to serve as a covariate in the impact analysis. It was also administered in spring 2008 to provide an outcome measure. The test was designed to measure two constructs aligned with the purpose of the professional development program: knowledge of rational numbers content typically taught in seventh grade (common knowledge of mathematics, or CK) and additional knowledge that may be useful for teaching rational number topics (specialized knowledge of mathematics for teaching, or SK).[12]

- **Teachers' instructional practices.** To measure instructional practice for treatment and control teachers, one *classroom observation* was conducted for each teacher after the treatment teachers in that district had had at least 5 of the 8 scheduled days of institutes and seminars. The observations produced three primary measures of instructional practice, which documented the frequency with which the teacher employed several key behaviors encouraged by the PD program.[13] The first measure, *Teacher elicits student thinking*, encompassed such behaviors as asking other students whether they agree or disagree with a particular student's response and also included behaviors elicited from the students such as offering additional justifications or strategies. The second measure, *Teacher uses representations*, counted the number of times the teacher displayed and explained a visual representation of mathematics, such as number lines or ratio tables, as well as the number of different types of representations the teacher used. The third measure, *Teacher focuses on mathematical reasoning*, counted the number of times that the teacher asked questions such as Why does this procedure work? Why does my answer make sense? or Why isn't 3/4 a reasonable answer to this problem?
- **Student achievement in rational numbers.** A customized, computer-adaptive *student achievement test* was constructed for the study by the Northwest Evaluation Association (NWEA). The test developed for this study was restricted to positive rational numbers content and drew on a customized item base that contained nearly 1,200 rational numbers items abstracted from the larger NWEA item bank of scaled, operational items.[14]

We also surveyed teachers to gather data on their backgrounds and on the amount and type of PD in mathematics they participated in during the study period. Study staff obtained information on the implementation of the PD by observing the institute and seminars and by reviewing logs maintained by coaches that recorded the nature of each coach interaction with each teacher.

Analytic Approaches

The basic analytic strategy for assessing the impact of the PD program was to compare outcomes for schools that were randomly assigned within each district to each of the two study conditions. Because we used nested data, three-level models (with students nested within teachers' classrooms nested within schools) were used to estimate the impact of professional development on student achievement and two-level models (with teachers nested within schools) were used to

---

12 Each form included 24 multiple-choice or short-response items, equally divided between CK and SK and equally divided between the two major domains of rational numbers on which the PD focused: (1) fractions and decimals and (2) ratio, rate, proportion, and percent.

13 These measures, although related to the goals of the PD program, do not provide comprehensive coverage of the behaviors the PD hoped to affect. Some desired behaviors did not lend themselves to observation in the course of a single class session (e.g., continuity, follow-up), and others could not be rated reliably by our observers, who did not have specific expertise in mathematics and mathematics teaching. We did not attempt to measure the accuracy of the mathematics presented or the quality of the teacher's actions.

14 Each individual student was presented with 30 items from the customized item base, chosen adaptively from four topic areas: fractions (11 items), decimals (4 items), percents (4 items), and ratios/proportions (11 items). Within each topic area, items were selected for presentation in a manner that ensured distribution across the cognitive categories of concepts, operations, and applications. To aid interpretation of the total score results, NWEA also constructed customized, seventh-grade norms by reanalyzing data from its Growth Research Database—a large data base compiled from NWEA testing.

estimate the impact on the teacher measures. The impact model used the sample of teachers and students present in the study schools as of the spring 2008 data collection period. The estimates provide an intent-to-treat analysis of the impact of the PD program because they reflect impact on the targeted (or "intended") sample, whether or not all eligible teachers in the treatment schools participated fully in the PD provided.

Study Findings After One Year of Treatment

Implementation Findings for First Year of Treatment

- **Across the study's 12 districts, the average number of hours of institutes, seminars, and coaching delivered was 67.6 hours—approximately the number intended.** During the institutes and seminars, the PD providers delivered an average of 45.2 hours of professional development, 94 percent of the intended 48 hours. During the coaching, the treatment group teachers received an average of 4.5 hours of coaching per 2-day coaching visit, 112 percent of the intended 4 hours per visit. Almost 84 percent of the coaching hours were spent on topics that were a focus of the study's PD program.
- **The treatment group teachers attended an average of 83 percent of the implemented hours of the study-provided PD program and reported participating in 55.4 hours more mathematics-related PD than the control group teachers.** Institute and seminar attendance records and coach logs recorded the extent of participation in the study-provided PD program. When asked to report on all mathematics-related PD received between summer and spring—including both study-provided PD and PD not related to the study—treatment group teachers reported receiving significantly more hours of mathematics-related institutes, seminars, and coaching than control group teachers (76.5 hours compared with 21.2 hours).

Impact Findings After One Year of Treatment

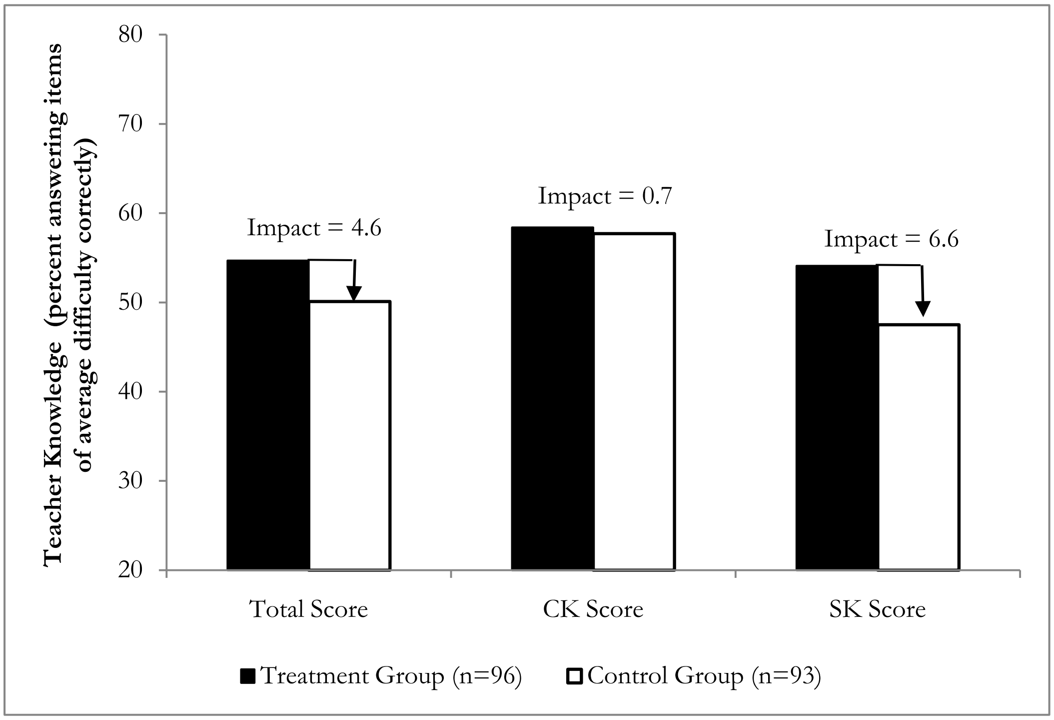
Impact on Teachers' Knowledge of Rational Number Topics and How to Teach Rational Number Topics

- **During the first year of implementation, the PD program did not have a statistically significant impact on overall teacher knowledge (effect size = 0.19, p-value = 0.15).** On average, 54.7 percent of teachers in the treatment group answered test items of average difficulty correctly, compared with 50.1 percent for teachers in the control group. (See Figure ES-1.) To put these results into context, the study also administered the teacher knowledge test to the PD provider staff (i.e., the staff who delivered the institutes, seminars, and coaching). On average, 92.7 percent of the PD provider staff answered test items of average difficulty correctly.[15]

---

[15] As described in Chapter 2, the difficulty level of the teacher knowledge test was intentionally aligned with the average knowledge level of the study population. The much higher performance of the PD facilitators on this same instrument provides perspective on the estimated size of the knowledge gain that was effected by the PD program.

Figure ES-1. First-Year Impact of the PD Program on Teacher Knowledge



SOURCE: Spring 2008 Teacher Knowledge Test (Teacher Impact Analysis Sample).

NOTES: The impact analysis for teacher knowledge was conducted using measures scaled in logits. The estimated impacts are based on a two-level model controlling for random assignment block and teacher-level covariates. The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for teachers in the treatment group as the basis for the adjustment.

The treatment group and the control group values presented in the figure are transformed means, and each impact value presented is the difference in these transformed means. The values for the percent answering items of average difficulty correctly correspond to the estimated treatment and control group means, scaled in logits.

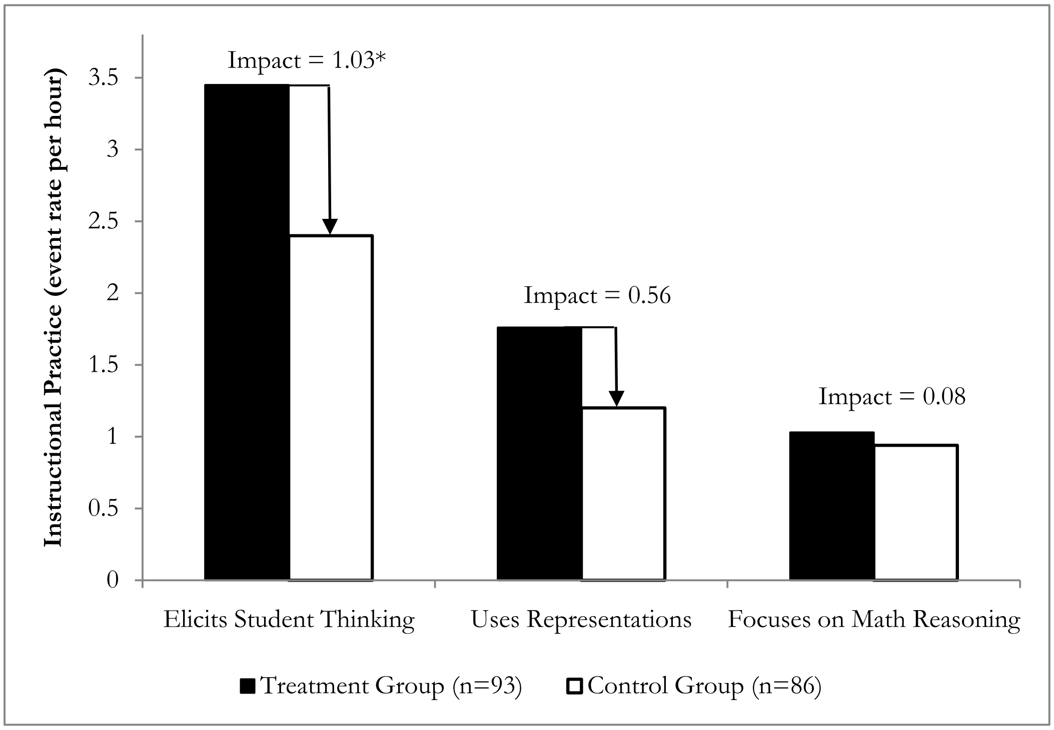
Statistical significance was determined on the basis of t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

- **The PD program did not have a statistically significant impact on either of the teacher knowledge subscale scores.** On average, 58.4 percent of treatment group teachers answered CK test items of average difficulty correctly, compared with 57.7 percent of control group teachers (effect size = 0.02, p-value = 0.88). On average, 54.7 percent of treatment group teachers answered SK test items of average difficulty correctly, compared with 47.5 percent of control group teachers (effect size = 0.23, p-value = 0.14). (See Figure ES-1.)

Impact on Teachers' Instructional Practices

- **During the first year of implementation, there was a statistically significant and positive impact of the PD program on the frequency with which teachers engaged in activities that elicited student thinking (effect size = 0.48).** Treatment teachers on average engaged in 1.03 more activities per hour that elicited student thinking. On average, teachers in the treatment group engaged in such activities 3.45 times per hour, compared with 2.42 times per hour for teachers in the control group. (See Figure ES-2.)

Figure ES-2. First-Year Impact of the PD Program on Instructional Practice



SOURCE: 2007–2008 Classroom Observation Protocol (Teacher Impact Analysis Sample).

NOTES: As noted in Chapter 4, the impact analysis for instructional practice was conducted using measures scaled in log rate per hour. Those estimated impacts are based on a two-level model controlling for random assignment block and teacher-level covariates. The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for teachers in the treatment group as the basis for the adjustment.

The treatment group and the control group values presented in the figure are event rates per hour, and each impact value presented is the difference in these event rates per hour. The values for the event rate per hour correspond to the treatment and control group means, scaled in log rates per hour (event rate = EXP(log rate)). For the Teacher Elicits Student Thinking scale, the event rate represents the average number of times per hour that teachers engaged in activities that elicited student thinking. The event rate for the Teacher Focuses on Mathematical Reasoning scale can be interpreted similarly. For the Teacher Uses Representations scale, the event rate can be interpreted as the average number of times per hour that teachers used representations or the average number of different types of representations that teachers used per hour.

Statistical significance was determined on the basis of t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

- **The PD program did not have a statistically significant impact on teachers' use of representations (effect size = 0.30; p-value = 0.0539).**[16] Treatment teachers on average used representations 1.76 times per hour, compared with 1.21 times per hour for the control group. (See Figure ES-2.)
- **The PD program did not have a statistically significant impact on the frequency with which teachers engaged in activities that focused on mathematical reasoning (effect size = 0.19, p-value = 0.32).** Treatment teachers on average engaged in activities that focused on mathematical reasoning 1.03 times per hour, compared with 0.94 for the control group. (See Figure ES-2.)

[16] See Chapter 2 and Appendix A for more detail on the construction of the *Teacher uses representations* scale.

Impact on Student Achievement in Rational Numbers

- **During the first year of implementation, the PD program did not have a statistically significant impact on average student achievement as measured by the *Total scale score* (effect size = 0.04, p-value = 0.37).** Students in treatment schools on average scored 217.11 scale score points, compared with 216.59 for the control group.
- **The PD program did not have a statistically significant impact on either of the student achievement subscale scores.** On the *Fractions and decimals score,* students in treatment schools on average scored 215.53 scale score points, compared with 215.01 scale score points for students in control schools (effect size = 0.03, p-value = 0.38). On the *Ratio and proportion score*, students in treatment schools on average scored 218.65 scale score points, compared with 218.18 scale score points for students in control schools (effect size = 0.03, p-value = 0.46).

Examining Additional Questions Related to the Impact Findings

We examined several additional questions related to the impact findings using nonexperimental analyses. Specifically, we examined whether teacher turnover during the school year might alter the interpretation of the impact findings, because teachers who began after the beginning of the school year did not have access to all of the PD. We also examined whether outcomes may have differed if the PD had targeted teachers with low or high levels of prior knowledge, or on students with low or high levels of prior achievement. Finally, we examined whether the knowledge or practices emphasized in the PD appear to be related to student achievement, irrespective of teachers' treatment status. The study was not designed to provide a rigorous test of these questions, so the results should be viewed as suggestive.

- **Teacher Turnover.** Some teachers in the treatment group participated in nearly all the PD, whereas others participated in only some of the PD. Teachers who remained in their schools from the fall baseline data collection to the end of the school year had access to more of the PD than those teachers who came later in the school year. We compared outcomes for treatment teachers who remained in their schools from the fall baseline data collection to the spring impact data collection with outcomes for control teachers who remained in their schools over this same period. Overall, 91 percent of the teachers in the impact analyses were present in the fall; the remaining 9 percent arrived sometime later in the year. Analyses focused on the subsample of "stable" teachers and their students yielded results similar to those for the full study sample. These nonexperimental results suggest that, despite its consequences for access to the PD, *teacher turnover does not appear to alter the observed impact findings.*
- **Baseline Teacher Knowledge.** A second question is whether the PD program may have been more or less effective for teachers who began the study with different levels of baseline knowledge. Teachers with high levels of baseline knowledge may have found the PD too easy; teachers with low levels of baseline knowledge may have found the PD too hard. Nonexperimental analyses did not show a statistically significant association between teachers' initial knowledge levels and treatment-control differences in teacher knowledge, their instructional practice, or student achievement, *which suggests that targeting the PD to teachers with a particular level of mathematics knowledge would be unlikely to alter the findings.*

- **Baseline Student Achievement.** A third question is whether the PD program may have been more or less effective for students who began the year with different levels of baseline achievement. Students of different initial achievement levels might have had different needs. Nonexperimental analyses indicated that the PD program did not appear to be more or less effective for students with low or high initial achievement, *which suggests that targeting the PD to teachers with students of a particular mathematics skill level would be unlikely to alter the findings.*

- **Teacher Knowledge, Instructional Practice, and Student Achievement.** A final question is whether the study's outcome measures captured aspects of teacher knowledge and instructional practice that are associated with student achievement. Correlational analyses show no statistically significant relationships linking the teacher knowledge measures and instructional practice measures to student achievement, although most of the coefficients were positive and consistent in magnitude with associations reported in the literature.

Summary

In summary, the study's results indicate that, during the first year of implementation, the PD program did not have a statistically significant impact on teacher knowledge. It had a significant positive impact on the frequency with which teachers engaged in activities intended to elicit student thinking, one of the study's three measures of instructional practice, but it did not have a statistically significant impact on the other two measures of instruction. The PD program did not have a statistically significant impact on student achievement in rational numbers.

Nonexperimental analyses conducted to supplement the main impact analyses suggest that the main results were not affected by teacher turnover during the implementation year. The nonexperimental analyses did not provide evidence of differential effectiveness for teachers with different levels of baseline knowledge or students with different levels of baseline achievement.

These results should be interpreted in the context of the study's design, the settings in which the PD was delivered, and the study's measures. The study was designed to examine the impact of the PD program as implemented by two providers in 12 districts. On average, students in the study schools entered seventh grade substantially below grade level, scoring at the 19th percentile on the study's measure of achievement in rational numbers. While one strength of the study is that it assessed the impact of the PD program on teacher knowledge and instruction, the instructional practice measures focused only on the frequency with which teachers engaged in specific practices, not the quality with which the practices were implemented. Further, although the study met the targets set for statistical power, the sample size and the reliability of the teacher measures limited the precision of the estimated effects on teacher knowledge and instruction.

The results reported here are based on a single year of implementation of the PD program, in the 2007-2008 school year. During the 2008–2009 school year, in 6 of the 12 study districts, teachers in schools randomly assigned to the treatment condition were provided with the opportunity to participate in a second year of PD focused on rational numbers. The next report from the Middle School Mathematics PD Impact Study will provide evidence on the impact of the full, two-year PD program.

REFERENCES

Birman, B., Le Floch, K.C., Klekotka, A., Ludwig, M., Taylor, J., Walters, K., et al. *State and Local Implementation of the No Child Left Behind Act: Vol. 2. Teacher quality under NCLB: Interim report.* Washington, DC: U.S. Department of Education; Office of Planning, Evaluation and Policy Development; Policy and Program Studies Service, 2007.

Borko, H. "Professional Development and Teacher Learning: Mapping the Terrain." *Educational Researcher*, 2004, 33(8): 3–15.

Clewell, B.C., Campbell, P.B., and Perlman, L. *Review of Evaluation Studies of Mathematics and Science Curricula and Professional Development Models*. Submitted to the GE Foundation. Washington, DC: The Urban Institute, 2004.

Kennedy, M. *Form and Substance of Inservice Teacher Education* (Research monograph no. 13). Madison: University of Wisconsin–Madison, National Institute for Science Education, 1998.

National Mathematics Advisory Panel. *Foundations for Success: The Final Report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education, 2008.

National Research Council. *Adding It Up: Helping Children Learn Mathematics*. J. Kilpatrick, J. Swafford, and B. Findell (Eds). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press, 2001.

Richardson, V., and Placier, P. "Teacher Change." In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th Ed.). New York: Macmillan, 2001: 905–947.

Supovitz, J.A. "Translating Teaching Practice into Improved Student Performance." In S. H. Fuhrman (Ed.), *From the Capitol to the Classroom: Standards-Based Reform in the States* (100th Yearbook of the National Society for the Study of Education, Part II). Chicago: University of Chicago Press, 2001: 81–98.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. *State and Local Implementation of the No Child Left Behind Act, Volume VIII—Teacher Quality Under NCLB: Final Report*. Washington, DC: 2009.

Yoon, K.S., Duncan, T., Lee, S. W.-Y., Scarloss, B., and Shapley, K. *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement* (Issues & Answers Report, No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest, 2007.