

Middle School Mathematics Professional Development Impact Study

Findings After the Second Year of Implementation

Executive Summary

Middle School Mathematics Professional Development Impact Study

Findings After the Second Year of Implementation

Executive Summary

May 2011

**Michael S. Garet
Andrew J. Wayne
Fran Stancavage
James Taylor
Marian Eaton
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 2011-4024
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

Rebecca Maynard

Commissioner

May 2011

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.

This report is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should be: Gare, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., and Doolittle, F. (2011). *Middle School Mathematics Professional Development Impact Study: Findings After the Second Year of Implementation* (NCEE 2011-4024). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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 are also fortunate to have had 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 appreciate the advice we received from Hyman Bass, University of Michigan, and others associated with the Learning Mathematics for Teaching project as well as from 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 American Institutes for Research (AIR) 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 for her support of those efforts and her 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 INTEREST¹

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 the Institute of Education Sciences (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

This is the second and final report of the Middle School Mathematics Professional Development Impact Study, which examines the impact of providing a professional development (PD) program in rational number topics to seventh-grade mathematics teachers. An interim report (Garet et al. 2010) described the findings after one year of PD. The current report documents the impact after providing a second year of PD in a subset of the original participating districts and includes supplemental analyses that use data from both years of the study.

To improve teachers' knowledge and skill, federal policymakers have committed significant resources to teacher PD. In 2004–2005, for example, states and districts spent \$1.5 billion in federal funds on teacher PD (Birman et al. 2007). There has, however, been only limited research evidence regarding the impact of PD on teacher and student outcomes.

Over the past decade, 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 et al. 2007). However, the most recent review identified only 9 out of 1,343 studies of PD that had the types of rigorous designs—randomized control trials (RCTs) or quasi-experimental designs (QEDs)—that allow causal inferences to be made about the effectiveness of the PD strategies they examined. Four of those studies addressed the effect of teacher PD on mathematics achievement, but none focused on middle school mathematics (Yoon et al. 2007).

The U.S. Department of Education's National Center for Educational Evaluation and Regional Assistance (NCEE)—within the Institute of Education Sciences (IES)—initiated the Middle School Mathematics PD Impact Study to learn more about the role of PD in improving teacher effectiveness. Specifically, the study examines the impact of two years of a PD program for seventh-grade mathematics teachers that focuses on teachers' knowledge of rational number topics, including specialized mathematics knowledge that may be useful for teaching these topics. Rational numbers—fractions, decimals, percent, ratio, and proportion—are interrelated topics that are challenging for many seventh-grade students and are considered an essential foundation for algebra (National Mathematics Advisory Panel 2008).

The study also tests the effect of a PD program when implemented with a relatively large sample, in varied settings, and using multiple facilitators. The PD was delivered to approximately 100 treatment teachers in 12 districts in the first year of the study and approximately 50 treatment teachers in 6 districts in the second year. Ten facilitators from two separate PD organizations were involved over the course of the study. By contrast, the 9 studies with rigorous designs identified by Yoon and colleagues (2007) involved smaller samples of 5 to 44 teachers, and the PD programs were delivered by the individuals who developed them.

The second year of the study was designed to address two questions:

- What cumulative impact did providing two years of the specified PD program have on teacher knowledge of rational number topics?

- What cumulative impact did providing two years of the specified PD program have on student achievement in rational number topics?

The study produced the following core second-year results:

- **The study's PD program was implemented as intended, but teacher turnover limited the average dosage received.** On average, the treatment teachers in the second-year impact sample received 68 percent of the full intended dosage. Because some teachers left the study schools and others entered as the study progressed, not all teachers had the opportunity to experience the full dose of PD. (In particular, 22 of the 45 treatment teachers present at the end of the two-year PD program were not present at its beginning.) Relative to the hours of PD that each teacher could possibly have attended (that is, relative to the hours of PD that occurred after the teacher entered a study school), the teachers in the second-year impact sample averaged 89 percent of the possible dosage.
- **At the end of the second year of implementation, the PD program did not have a statistically significant impact on teacher knowledge.** There were no significant impacts on teachers' total score on a specially constructed teacher knowledge test (effect size = 0.05, p-value = 0.79) or on either of the test's two subscores. On average, 75.7 percent of the teachers in the treatment group correctly answered test items that were of average difficulty for the test instrument, compared with 74.7 percent of the teachers in the control group.
- **At the end of the second year of implementation, the PD program did not have a statistically significant impact on average student achievement in rational numbers.** There were no significant impacts on students' total score on a customized rational numbers test (effect size = -0.01, p-value = 0.94) or on either of the test's two subscores.

Overview of the PD Program

The PD program delivered in this study focused entirely on rational number topics and was designed to develop teachers' capability to teach positive rational number topics effectively. For each rational number topic area, the PD program design emphasized using precise definitions and the properties and rationales underlying common procedures used with rational numbers. In addition, the PD emphasized developing teachers' ability to explain rational number concepts and procedures, identify and address persistent student misconceptions (often by presenting students with problems designed to reveal their thinking), and use representations of rational number concepts in teaching.

Two providers—America's Choice and Pearson Achievement Solutions—were selected through a competitive process to produce and deliver the PD.² Both providers worked with a common set of guidelines regarding the structure of the PD program, the knowledge to be

² PD provider candidates responded to a solicitation that laid out the basic parameters of the PD intervention. Selection of the winning candidates was guided by an expert panel and was based on the extent to which the candidates had existing PD materials pertaining to rational numbers and the alignment between their existing materials and the goals and specifications of the planned intervention. The decision to use two providers had two bases: first, a desire to ensure that there was sufficient capacity to deliver high quality PD to 12 districts, and second, a desire to test the impact of the PD design by allowing two different instantiations of the same basic design features.

developed, and key aspects of the delivery of the PD while also building on their existing PD materials that addressed topics in rational numbers. Facilitator guides were refined through a yearlong 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.

As shown in Table ES-1, during each year of the study, the study-provided PD included a summer institute, a series of one-day follow-up seminars held during the school year, and in-school coaching visits conducted in association with the seminar days and delivered by the seminar facilitators. The specification of the PD program was guided by the literature, which is largely based on correlational research and practitioner experience.³

The PD program provided to teachers who participated in both years of the study was designed to deliver 114 contact hours (68 hours in the first year and 46 hours in the second year). For teachers who entered the study in the second year, the PD provided 58 contact hours, including the 46 hours offered to all teachers and a 12-hour “makeup” institute that provided a condensed version of the summer institute from the first year of the study. The amount of PD in mathematics offered annually by the study was more than most mathematics teachers typically receive in a single year.⁴

³ 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.

⁴ 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).

Table ES-1. Days and Per-Teacher Hours of PD Provided in First and Second Years of the Study

Activity	First Year (2007–2008)	Second Year (2008–2009)
PD for All Participating Teachers		
Summer Institute	3 days (18 hours)	2 days (12 hours)
Seminars During the School Year	5 days (30 hours)	3 days (18 hours)
Intensive In-School Coaching ^a	10 days (20 hours)	8 days (16 hours)
Total Hours of PD	68 hours	46 hours
Makeup PD for Teachers Who Joined the Study After the First-Year Summer Institute		
Special Summer Institute		2 days (12 hours)

NOTES: ^a Each teacher was expected to receive two hours of individual or group coaching per day of in-school coaching.

For the summer institutes and seminars, 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 coaching visits, which were scheduled to occur within a few days of each of the seminar days, employed both individual and group activities and were designed to help the teachers apply material covered in the institutes and seminars to their classroom instruction.

The PD was not presented to teachers as an opportunity to improve their understanding of rational number content, and the PD did not offer an opportunity for teachers to explicitly evaluate their own knowledge of rational numbers (by assigning a test of rational numbers, for example).⁵ Further, the PD did not generally require teachers to spend time outside the institutes and coaching activities studying rational number content or practicing pedagogical techniques.

Study Design

The study used an experimental design with random assignment of schools to treatment and control conditions within each participating district. Schools remained in the same treatment condition for both years of the study. 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.

Midway through the first implementation year, results from the NCEE study of PD in early reading became available (see Garet et al. 2008). The results showed that although the single year of PD tested in the study had a statistically significant impact on some dimensions of teacher knowledge and instructional practice at the end of the year in which the PD was implemented, the PD did not produce a statistically significant impact on student achievement and did not produce a statistically significant impact on teachers' knowledge, teachers' instructional practices, or student achievement in the year following the year of the PD. That is, the study had no statistically

⁵ The results of the teacher knowledge test used in the evaluation were not shared with the teachers or the providers.

significant impact on student achievement, and the impact of the PD on teachers' knowledge and instructional practice was not sustained.

Based on this information, NCEE elected to explore the effect of extending the implementation of the seventh-grade mathematics PD to two years. Because of resource constraints, the second year of PD was offered in only half of the originally participating districts.

Study Sample

The process used to recruit 12 districts for the first year of the study was designed to produce a sample that was relevant to federal education programs—which tend to target low-income students—and large enough to provide power to detect impacts of the anticipated magnitude in teacher and student outcomes.

For the second year sample of 6 districts, we wanted to maintain the balance between PD providers. After excluding districts in which we expected the composition of the study schools to change as a result of restructuring initiatives, we selected the 3 districts for each provider with the largest number of schools in the sample, thus maximizing the statistical precision. Districts were selected before the first-year results were known, so findings about the impact of the first year of the PD on teachers and students—overall or in specific districts—did not inform the choice of districts to participate in the second year of the study.

Thirty-nine schools participated in the second year of the study. The second-year impact analysis sample included 92 teachers and 2,132 students, distributed across treatment and control groups as shown in Table ES-2. Among the 92 teachers, 51 (23 in the treatment group and 28 in the control group) had participated in the study since baseline (fall 2007).

Table ES-2. Numbers of Schools, Teachers, and Students in Second-Year Impact Analysis Sample, Overall and by Treatment Status

Treatment Status	Number of Schools	Number of Seventh-Grade Teachers		Number of Seventh-Grade Students	
		Total Number	Average Per School	Total Number	Average Per School
Treatment	20	45	2.4	1,083	54.2
Control	19	47	2.4	1,049	55.2
Total	39	92	2.4	2,132	54.7

SOURCE: Teacher Rosters; Study District Records.

All eligible teachers teaching at least one regular seventh-grade mathematics class in spring 2009 were members of the second-year teacher impact sample, and a random sample of all seventh-grade students who were in the teachers' regular seventh-grade mathematics classes in spring 2009 were members of the second-year student impact sample.^{6,7} 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.

Table ES-3 provides descriptive information about the characteristics of the sample of 39 schools in the two-year districts compared with the characteristics of schools serving seventh-grade students in the national sample of similar districts from which the original 12-district sample was recruited for the study. On some key characteristics, the study sample schools were statistically different from the larger pool of eligible schools. The study sample schools were less likely to be in the South and more likely to be in the Northeast region and to be in cities rather than in urban fringe communities, towns, or rural areas. On average, they had smaller enrollments than schools in the national sample (753 students vs. 920 students) and smaller teaching staffs (48.5 FTEs vs. 54.9 FTEs). The schools in the two-year districts also were less likely than schools in the national sample to be middle schools (67 percent vs. 95 percent) and more likely to serve a combination of elementary and middle school grades (33 percent vs. 3 percent).

⁶ "Eligible teachers" are defined as regular teachers, not short-term substitutes. (Long-term substitutes were included.)

⁷ 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. Characteristics of Schools in Two-Year Districts and All Eligible Schools in Large Districts

School Characteristics	Schools in Two-Year Districts	All Eligible Schools in Large Districts^a
Geographic Region (percent of schools)		
Northeast	35.9	8.8*
South	35.9	55.8*
Midwest	12.8	9.0
West	15.4	26.4
Urbanicity (percent of schools)		
Large or Middle-Sized City	87.2	59.1*
Urban Fringe, Large or Small Town, or Rural Area	12.8	40.9*
Title I Eligible (percent of schools)	66.7	67.8
Free or Reduced-Price Lunch (school average percent of students)	66.1	65.3
Race/Ethnicity (school average percent of students)		
White	34.7	27.9
Black	34.7	31.1
Hispanic	25.4	33.5
Asian	2.6	5.5
Other	1.4	0.9
Male (school average percent of students)	51.6	50.7
Total School Enrollment	752.6	919.5*
Number of Seventh-Grade Students	207.9	310.9*
Number of Full Time Equivalent Teachers (all grades)	48.5	54.9*
School Type (percent of schools) ^b		
Middle School Only	66.7	95.2*
Elementary and Middle	33.3	2.9*
Middle and High	0.0	1.7
Elementary and Middle and High	0.0	0.2

Sample Size: N = 39 schools in second-year 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 In classifying 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 owing to rounding.

Despite these differences between study schools and all eligible schools, the *teachers* in study schools were not statistically distinguishable from teachers in a national sample of seventh-grade mathematics teachers in large urban school districts on any of the teacher characteristics presented in Table ES-4.

Table ES-4. Characteristics of Teachers in Second-Year Teacher Impact Analysis Sample and Mathematics Teachers of Seventh-Grade Students in Eligible Schools in Large Districts

Teacher Characteristics	Teachers in Second-Year Impact Analysis Sample ^a	All Teachers of Seventh-Grade Students in Eligible Schools in Large Districts
Standard Certification (percent)	72.2	73.4
Bachelor's Degree (percent)	100.0	100.0
Master's Degree (percent)	45.6	40.7
Mathematics Major (percent)	18.9	29.3
Mathematics-Related Major (percent)	4.4	16.2
Years of Teaching Experience (percent)		
3 years or fewer	30.0	37.4
4–10 years	40.0	26.9
11–20 years	21.1	15.7
More than 20 years	8.9	20.1

Sample Size: N = 92 teachers in second-year impact analysis sample; 10,700 teachers in eligible schools.

SOURCE: Teacher Survey; 2003–2004 Schools and Staffing Survey (SASS), Public School Teacher Data Files.

NOTES: ^a Characteristics of study teachers were measured at time of entry into the study.

Percentage values for characteristics with multiple categories may not sum to 100 owing 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 (*).

Data Collection and Outcome Measures

Data were collected from teachers and students in the study schools in fall and spring of the 2007–2008 and 2008–2009 school years. The two main outcome measures used in the second year of the study were constructed as follows:

- **Teacher knowledge test.** Teacher knowledge was measured for all treatment and control teachers using a test constructed specifically for the study. The test consisted of multiple-choice and short-response items that were designed to measure knowledge of rational number topics. Three alternate forms of the test were administered so that individual teachers would receive different forms (i.e., different items) at each administration. In addition to a total score, the teacher knowledge test yielded two subscores for each participant, aligned with the two types of knowledge that were targeted by the PD: common knowledge of mathematics (CK) and specialized knowledge of mathematics for teaching (SK).⁸
- **Student achievement test.** A customized, computer-adaptive rational number test was constructed for the study by the Northwest Evaluation Association (NWEA). The NWEA Rational Number Test was restricted to positive rational number content and drew on a customized item bank of nearly 1,200 rational number items abstracted from

⁸ 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. SK is the additional knowledge of rational numbers that may be useful for teaching rational number topics.

the larger NWEA item bank of scaled, operational mathematics items.⁹ Three Item Response Theory (IRT)-based scores were computed for each participant: a total score, a fractions and decimals score, and a ratio and proportion score.

We also surveyed treatment and control teachers to gather data on their professional backgrounds and on the amount and type of PD in mathematics they participated in during the two-year study period. Study staff obtained information on the implementation of the PD by collecting attendance records, observing the institutes and seminars, and reviewing logs maintained by coaches that recorded the nature and extent of each coach interaction with each teacher.

Analytic Approaches

The basic strategy for the impact analysis was to estimate the difference in outcomes between the treatment and control groups, adjusting for the blocking used in random assignment and for teacher- and student-level covariates. Because random assignment was conducted separately within each of the six school districts participating in the second year of the study, the study comprised six separate random assignment experiments. To obtain the impact estimates, we pooled the data for all six study districts in a single analysis, treating the districts as fixed effects.¹⁰ Separate program impact estimates were obtained for each district and then averaged across the six districts, weighting each district's estimate in proportion to the number of treatment schools from the district in the study sample. Findings in this report therefore represent the impact on the performance of teachers and students in the average treatment school in the 6 two-year study districts. The results do not necessarily reflect what the treatment effect would be in the wider population of districts from which those in the study were selected.

The impact estimates provide an “intent to treat” analysis of the impact of the program; that is, the estimates reflect the program impact on all teachers and students in the targeted classrooms in the study schools, even though some of those teachers and students were not present for the full duration of the study and some of the teachers did not take full advantage of the opportunity to participate in the study-provided PD even though they were present.

A common way to represent statistical precision is as a minimum detectable effect size (MDES), which is the smallest true effect that an estimator has a good chance of detecting (Bloom 1995). The second year of the study was powered to detect an effect size of 0.59 for teacher knowledge and 0.20 for student achievement

Study Findings After Two Years of Treatment

Implementation Findings

- **Across the six districts that participated in the study for two years, the average number of hours of institutes, seminars, and coaching delivered was 118 hours, which was 4 hours more than the intended dosage of 114 hours. During the**

⁹ 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 database compiled from NWEA testing (NWEA 2003).

¹⁰ Schools, classes, and students were treated as random effects.

institutes and seminars, the PD providers delivered an average of 93 percent of the intended hours of professional development in each year of the study. With regard to coaching, the treatment group teachers received an average of 97 percent of the intended hours in the first year and 132 percent of the intended hours in the second year.

- **The treatment group teachers attended an average of 77 hours of study PD and reported participating in 63.6 hours more mathematics-related PD than the control group teachers.** The average hours of study PD attended represented 68 percent of the intended dose of 114 hours and 66 percent of the total 118 PD hours implemented across the two years. However, relative to the hours of PD that each teacher could possibly have attended (that is, relative to the hours of PD that occurred after the teacher entered a study school), the teachers in the second-year impact sample averaged 89 percent of the possible dosage.
- **Teacher turnover limited the maximum possible PD dosage and the magnitude of the treatment-control group service contrast.** Twenty-two of the 45 treatment teachers teaching regular seventh-grade mathematics classes at the end of the two-year PD program were not present at its beginning. Most turnover occurred over the summer between the two years of implementation.¹¹

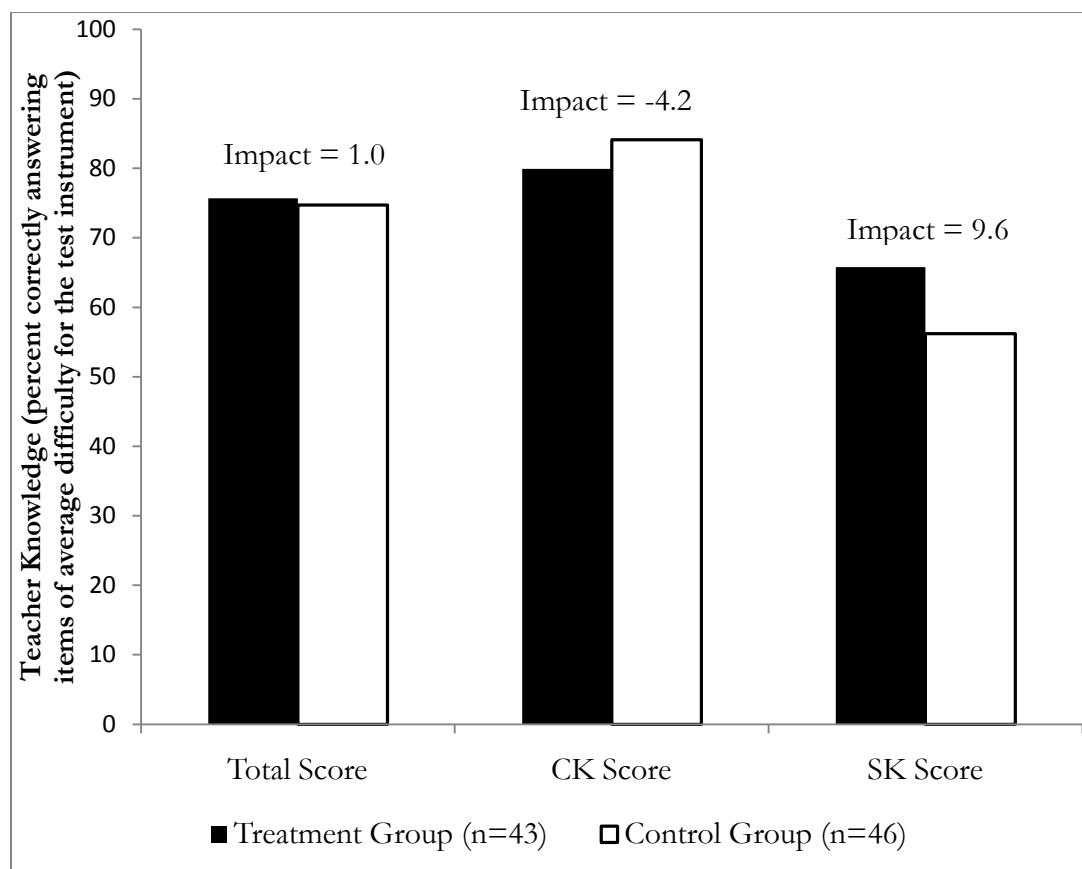
Impact Findings

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

- **At the end of the second year of implementation, the PD program did not have a statistically significant impact on overall teacher knowledge.** On average, 75.7 percent of teachers in the treatment group correctly answered test items of average difficulty for the test instrument, compared with 74.7 percent for teachers in the control group (effect size = 0.05, p-value = 0.79). (See Figure ES-1.)
- **The PD program did not have a statistically significant impact on either of the teacher knowledge subscale scores.** On average, 79.9 percent of treatment group teachers correctly answered CK test items of average difficulty for the test instrument, compared with 84.1 percent of control group teachers (effect size = -0.21, p-value = 0.25). On average, 65.8 percent of treatment group teachers correctly answered SK test items of average difficulty for the test instrument, compared with 56.2 percent of control group teachers (effect size = 0.36, p-value = 0.09). (See Figure ES-1.)

¹¹ Within the 6 two-year districts, there were 45 teachers in the treatment group at the beginning of the first year and 45 teachers in the treatment group at the end of the second year. However, between those two time points, 22 treatment teachers left the study (because they no longer taught eligible classes at the participating schools), and 22 teachers joined the study. Five of these staff transitions occurred during the first year of the program, 13 occurred over the summer between the first and second years (but before the summer institutes), and 6 occurred during the second year of the program.

Figure ES-1. Impact of the PD Program on Teacher Knowledge at the End of the Second Year



SOURCE: Spring 2009 Teacher Knowledge Test (Second Year Teacher Impact Analysis Sample).

NOTES: The impact analyses for teacher knowledge were 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 figure displays 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 values for the percent correctly answering items of average difficulty for the test instrument correspond to the estimated treatment and control group means, scaled in logits.

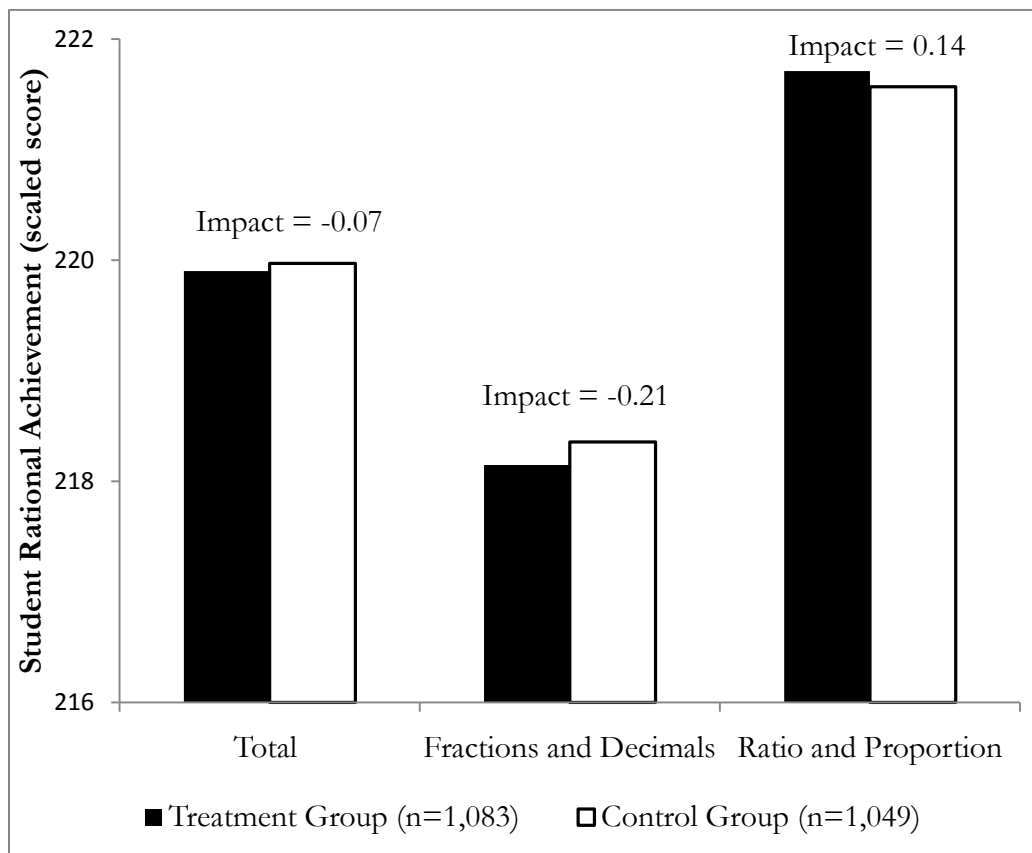
P-values are based on t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

Impact on Student Achievement in Rational Numbers

- At the end of the second year of implementation, the PD program did not have a statistically significant impact on average student achievement as measured by the NWEA Rational Number Test *Total Score*. Students in treatment schools on average scored 219.90 scale score points, compared with 219.97 scale score points for the control group (effect size = -0.01 , p -value = 0.94). (See Figure ES-2.)

- **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 218.15 scale score points, compared with 218.36 scale score points for students in control schools (effect size = -0.01 ; p -value = 0.84). On the *Ratio and Proportion Score*, students in treatment schools on average scored 221.71 scale score points, compared with 221.57 scale score points for students in control schools (effect size = 0.01 ; p -value = 0.89). (See Figure ES-2.)

Figure ES-2. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year



SOURCE: Spring 2009 NWEA Rational Number Test.

NOTES: The impact analyses for student mathematics achievement were conducted using scale scores. Although the theoretical scale scores for the student achievement test range in value from negative infinity to positive infinity, typical scores fall between 150 and 300 (NWEA 2003).

The estimated impacts are based on a three-level model controlling for random assignment block and student-level covariates.

The figure displays regression-adjusted mean outcomes for each group, using the mean covariate values for students in the treatment group as the basis for the adjustment.

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

Exploratory Analyses

We conducted several additional analyses, extending the exploratory analyses conducted for the Interim Report, and using the added power of a “pooled” sample of teachers. This pooled sample comprises three mutually exclusive and collectively exhaustive groups of teachers: teachers who were in the first-year impact analysis sample only (from all 12 districts); teachers who were in the second-year impact analysis sample only (from the 6 two-year districts); and teachers who were in both impact analysis samples. Teachers who were in both impact analysis samples (also from the 6 two-year districts) are included in the pooled sample twice, once using their first-year outcomes, and once using their second-year outcomes, controlling for their knowledge scores at the end of the first year/beginning of the second year. We also constructed a pooled sample of students that includes students who were in the first-year impact analysis sample (from all 12 districts) and students who were in the second-year impact sample (from 6 two-year districts).¹²

- **One-year effects of PD on teacher knowledge.** The estimated effects of one year of PD on teacher knowledge total score and CK for the pooled sample were not statistically significant. However, the estimated average effect of one year of the PD program on SK using the pooled sample was statistically significant (effect size = 0.28, $p = 0.02$).¹³
- **Average effect of PD on student achievement.** Different groups of students experienced the effect of the PD in each year of the study. The estimated average effect of the PD on student achievement using the pooled sample was not found to be statistically significant.
- **Results by provider.** We also used the pooled sample to examine the impact of the PD program separately for the two PD providers, America’s Choice and Pearson Achievement Solutions. These analyses did not indicate significant effects of the PD program on teacher knowledge or student achievement for either provider.
- **Baseline teacher knowledge.** Similarly, we drew on the pooled analysis sample to examine whether the PD program may have been more or less effective for teachers who began the study with different levels of baseline knowledge. We hypothesized that 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. The analyses did not show a statistically significant association between teachers’ initial knowledge levels and treatment-control differences in teacher knowledge or student achievement outcomes.
- **Baseline student achievement.** We also drew on the pooled sample to examine whether the PD may have been more or less effective for students who began the year with different levels of baseline achievement. Students with different initial achievement levels may have had different needs. The analyses indicated that the PD program did not appear to be more or less effective for students with low or high initial achievement.

¹² The “pooled” sample of teachers used in the per-year effect analyses includes 138 teachers who were in the first-year impact sample only, 38 teachers who were in the second-year impact sample only, and 51 teachers who were in both the first- and second-year impact samples. Since the students in each year of the study represented the teachers’ *current* seventh-grade students, there was no overlap between the first- and second-year student samples in the pooled sample.

¹³ The effect of one year of PD was calculated as the average of the one-year effect of the first year of PD and the additional one-year effect of the second year of PD.

- **Teacher knowledge and student achievement.** Finally, we drew on the pooled analysis sample to examine whether the study's measure of teacher knowledge was associated with student achievement as was hypothesized in the study's theory of action. Correlational analyses show a statistically significant positive association between the teacher knowledge total score and the student achievement total score of 0.05 (p-value = 0.02) and between the teacher knowledge total score and the student *Fractions and Decimals Score* of 0.05 (p-value < 0.01).¹⁴

Overall Study Summary

In summary, the study results indicate that after two years of implementation, the PD program did not have a statistically significant impact on teacher knowledge or on student achievement in rational numbers. These second-year results are consistent with the results at the end of the first year. At the end of the first year, the PD program did not have a significant impact on teacher knowledge or student achievement. Observations of teachers were conducted only in the first year. In the first year, the PD program had a statistically significant impact on one measure of instructional practice (the *Teacher elicits student thinking Scale*), a nearly significant impact on a second (the *Teacher uses representations Scale*, $p = .054$), but no significant impact on the third measure of instructional practice used in the study (the *Teacher focuses on mathematical reasoning Scale*).

Exploratory analyses based on a pooled sample, which combined data from the first and second years of the study to maximize the precision of the estimated effects, suggest that on average, each year of the PD had a statistically significant positive effect on SK, one of the two dimensions of teacher knowledge measured by the study. There was no effect on CK, the other dimension of teacher knowledge. Other exploratory analyses suggest that there was no significant differential effect of the PD for teachers who differed in baseline knowledge or prior experience, or for students who differed in baseline achievement. Exploratory analyses also suggest that students taught by teachers with higher knowledge scores exhibited significantly higher achievement, after controlling for prior achievement and other student background characteristics.

Although teachers' mathematical knowledge may be associated with student achievement gains, and thus may be a useful focus for PD, the PD tested did not have an effect on teacher knowledge of a magnitude that translated into an impact on student achievement. The results suggest that teachers' SK may have improved with each year of study PD. However, it is unclear whether multiple years of PD would produce larger gains in SK, especially without configuring the PD to take into account teacher mobility. Within a given year, our impact results suggest that, in order to affect achievement outcomes, the PD would have to be more efficient than the PD tested here in improving SK on an annual basis. Finally, while our evidence and evidence from other studies indicates that there is an association between teacher knowledge and student achievement, we do not know the relative importance of SK and CK. The study PD was primarily focused on SK and was not as directly focused on CK. Providing PD that places more direct emphasis on CK is another potential avenue for future study.

¹⁴ To examine the relationship between teacher knowledge and student achievement, we incorporated the teacher knowledge total score in the impact model in place of the treatment status indicator. Separate analyses were also conducted using the CK and SK subscores rather than the total score. We then examined the estimated coefficients for each of the teacher knowledge scores and calculated the statistical significance of the coefficients using a two-tailed t-test.

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.
- Bloom, H.S. “Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs.” *Evaluation Review*, 1995, 19(5): 547–556.
- 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.
- Common Core State Standards Initiative. 2010. *The Standards: Mathematics*. Washington, DC: National Governors Association and the Council of Chief State School Officers. Available online at: <http://www.corestandards.org/the-standards/mathematics>.
- Garet, M., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement* (NCEE 2008-4030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2008.
- Garet, M., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., et al. *Middle School Mathematics Professional Development Impact Study: Findings After the First Year of Implementation* (NCEE 2010-4009). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2010.
- 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.
- Northwest Evaluation Association (NWEA). *Technical Manual for Use With Measures of Academic Progress and Achievement Level Tests*. Lake Oswego, OR: NWEA, 2003.
- 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.
- 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.