

Middle School Mathematics Professional Development Impact Study

Findings After the Second Year of Implementation

Middle School Mathematics Professional Development Impact Study

Findings After the Second Year of Implementation

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: be: Garet, 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 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.

CONTENTS

Executive Summary	xv
Overview of the PD Program	xvi
Study Design	xviii
Study Findings After Two Years of Treatment	xxiii
Exploratory Analyses.....	xxvii
Overall Study Summary.....	xxviii
Chapter 1 Overview of the Study	1
Background and Importance of the Study	1
Recruitment and Random Assignment in the First and Second Years of the Study	4
Characteristics of the Schools in the Two-Year Districts	6
Second-Year Analysis Samples	9
Sources of Data	13
Analytical Approaches	15
Chapter 2 Design and Implementation of the PD Program	19
Design of the PD Program	19
Summer Institute and Seminar Series.....	22
Coaching.....	23
Implementation of the PD Program	23
Comparison of the PD Experienced by Treatment and Control Groups.....	31
Summary	33
Chapter 3 Impact of the PD Program	35
Equivalence of Treatment and Control Groups.....	35
Impact on Teacher Knowledge.....	38
Impact on Student Achievement	41
Summary	43
Chapter 4 Summary of Findings and Exploratory Analyses	45
Summary of Impact Findings	45
Results by Provider, Interactions With Baseline Characteristics, and Correlational Results	50
Summary	53
References	55
Appendix A Details of the Study Samples	A-1
Comparison of Schools, Teachers, and Students in Two-Year and One-Year Districts.....	A-2
Student Sample	A-6
Appendix B Details of Data Collection and Analytical Approaches	B-1
Details of Data Collection.....	B-1
Response Rates	B-12
Technical Notes on Analytic Approaches	B-14
Addressing Risks Associated With Multiple Hypothesis Tests	B-17

Appendix C Supplemental Information on the Design and Implementation of the PD Program	C-1
Scheduled Coverage of Mathematics Topics	C-1
Detailed Specifications of Each PD Provider’s Approach to Institutes and Seminars and to Coaching.....	C-1
Content and Structure of Institutes and Seminars	C-4
Supplemental PD Implementation Results by PD Provider	C-12
Teacher Participation in the PD Program by Provider and Date of Entry	C-22
Supplemental Data on Service Contrast	C-25
Appendix D Supporting Tables and Figures for Impact Analyses.....	D-1
Equivalence of Treatment and Control Group Characteristics	D-1
First-Year Impacts for One-Year Districts.....	D-5
Robustness Checks for Impact Estimates	D-7
Variation in the Impact of the PD Program across Districts	D-10
Unadjusted Means and Standard Deviations of Second-Year Outcome Measures for Treatment and Control Groups	D-13
Appendix E Exploratory Analyses: Approaches and Additional Results	E-1
Analysis of the One-Year Effect of the PD Program on Teacher Knowledge at the End of the Second Year	E-1
Analysis of the Per-Year Effect on Teacher Knowledge	E-2
Treatment-Control Differences in Baseline Teacher Knowledge.....	E-7
Analysis of the Average Annual Effect on Student Achievement	E-9
Effects of the PD Program on Teacher Knowledge and Student Achievement, by Provider, for the Pooled Sample	E-10
Differential PD Effect Based on Baseline Teacher Knowledge and Years of Teaching Experience	E-13
MDESs for Test of Differential PD Effects.....	E-24
Relationships Between Teacher Knowledge and Student Achievement.....	E-25
Relationships Among Teacher Knowledge and Student Achievement Using Four-Level Model. E-	

EXHIBITS

Exhibit 1-1. Theory of Action	3
Exhibit 1-2. Teacher Turnover During the Two Years of the Study: Two-Year Districts.....	10
Exhibit A-1. Student Turnover During the Second Year of the Study	A-8
Exhibit A-2. Construction of the Second-Year Student Impact Analysis Sample.....	A-10
Exhibit B-1. PD Characteristics Scales Used in Analysis of Service Contrast.....	B-6
Exhibit B-1. PD Characteristics Scales Used in Analysis of Service Contrast (continued)	B-7

FIGURES

Figure ES-1. Impact of the PD Program on Teacher Knowledge at the End of the Second Year	xxv
Figure ES-2. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year.....	xxvi
Figure B-1. Test Duration for Student Test Administration, by Test Wave in the First Year of the Study: First-Year Student Baseline and Impact Analysis Samples	B-10
Figure B-2. Distribution of Standard Errors by Total RIT Score on Fall 2007 NWEA Rational Number Test: First-Year Student Baseline Analysis Sample.....	B-11
Figure D-1. Impact of the PD Program on Teacher Knowledge at the End of the Second Year: Total Score, by District: Second-Year Teacher Impact Analysis Sample.....	D-11
Figure D-2. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year: Total Score, by District: Second-Year Student Impact Analysis Sample.....	D-12

TABLES

Table ES-1. Days and Per-Teacher Hours of PD Provided in First and Second Years of the Study.....	xviii
Table ES-2. Numbers of Schools, Teachers, and Students in Second-Year Impact Analysis Sample, Overall and by Treatment Status	xx
Table ES-3. Characteristics of Schools in Two-Year Districts and All Eligible Schools in Large Districts	xxi
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.....	xxii
Table 1-1. Characteristics of Schools in Two-Year Districts and All Eligible Schools in Large Districts	8

Table 1-2. Characteristics of Teachers in Second-Year Teacher Impact Analysis Sample and Mathematics Teachers of Seventh-Grade Students in Eligible Schools in Large Districts.....	12
Table 1-3. Numbers of Schools, Teachers, and Students in Second-Year Impact Analysis Sample, Overall and by Treatment Status	13
Table 1-4. Second-Year Minimum Detectable Effect Sizes (MDES) for Core Outcomes: Second-Year Teacher and Student Impact Analysis Samples	18
Table 2-1. Days and Per-Teacher Hours of PD Offered During the First and Second Years of the Study.....	19
Table 2-2. Teacher Institutes and Seminars—Percent of Intended Time Implemented and Actual Hours Implemented: Two-Year Districts	25
Table 2-3. Teacher Institutes and Seminars—Approximate Hours of Implemented Time Covering Specific Content Areas: Two-Year Districts.....	26
Table 2-4. Teacher Institutes and Seminars—Mean Reallocated Hours and Percent of Planned Segments Omitted and Abbreviated: Two-Year Districts	27
Table 2-5. Coaching—Percent of Intended Time Implemented and Mean Actual Hours per Teacher per Visit: Second-Year Teacher Impact Analysis Sample.....	28
Table 2-6. Percent of Implemented PD Hours Attended by the Average Teacher: Second-Year Teacher Impact Analysis Sample.....	30
Table 2-7. Treatment and Control Group Contrast in Hours of Mathematics-Related PD: Second-Year Teacher Impact Analysis Sample	32
Table 3-1. Teacher Characteristics, by Treatment Status: Second-Year Teacher Impact Analysis Sample.....	37
Table 3-2. Student Characteristics, by Treatment Status: Second-Year Student Impact Analysis Sample.....	38
Table 3-3. Impact of the PD Program on Teacher Knowledge at the End of the First and Second Years: First- and Second-Year Teacher Impact Analysis Samples—Two-Year Districts	40
Table 3-4. Impact of the PD Program on Student Mathematics Achievement at the End of the First and Second Years: First- and Second-Year Student Impact Analysis Samples—Two-Year Districts	42
Table 4-1. Impact on Teacher Knowledge (Effect Size) as Estimated for Different Time Periods and Samples (Main Impact Estimates Highlighted)	47
Table 4-2. Impact on Instructional Practice (Effect Size), as Estimated for Different Time Periods and Samples (Main Impact Estimates Highlighted)	49
Table 4-3. Impact on Student Achievement (Effect Size), as Estimated for Different Time Periods and Samples (Main Impact Estimates Highlighted)	50
Table 4-4. Standardized Regression Coefficients for the Relationships Between Teacher Knowledge and Student Achievement, Pooled Sample	53
Table A-1. Characteristics of Schools in One-Year and Two-Year Districts	A-3

Table A-2. Characteristics of Teachers in One-Year and Two-Year Districts: First-Year Teacher Impact Analysis Sample.....	A-4
Table A-3. Characteristics of Students in One-Year and Two-Year Districts: First-Year Student Impact Analysis Sample.....	A-5
Table A-4. Characteristics of Students Included and Not Included in the Second-Year Student Impact Analysis Sample: Second-Year Spring Expanded Student Sample	A-11
Table B-1. Distribution of Items on NWEA Rational Number Test.....	B-8
Table B-2. Average Test Duration (Minutes) for NWEA Rational Number Test, by Treatment Status and Test Wave in the First Year of the Study: First-Year Student Baseline and Impact Analysis Samples.....	B-9
Table B-3. Response Rates for All Teacher and Student Measures, by Treatment Status: Two-Year Districts ^a	B-13
Table B-4. Missing Data for Teacher and Student Characteristics Used as Covariates in the Impact Models, Second-Year Teacher and Student Impact Analysis Samples.....	B-17
Table C-1. Teacher Institutes and Seminars—Approximate Hours of Implemented Time Covering Specific Content Areas, by PD Provider: Two-Year Districts.....	C-13
Table C-2. Teacher Institutes and Seminars—Mean Reallocated Hours and Percent of Planned Segments Omitted and Abbreviated, by PD Provider: Two-Year Districts.....	C-14
Table C-3. Percent of Teacher Institute and Seminar Days on Which Features of the PD Matched the Plan, Overall and by PD Provider: Two-Year Districts	C-16
Table C-4. Coaching—Percent of Intended Time Implemented and Mean Actual Hours per Teacher per Visit, by PD Provider: Second-Year Teacher Impact Analysis Sample	C-17
Table C-5. Percent of Coaching Visits With Specified Features and Time Spent in Coaching With These Features: Second-Year Teacher Impact Analysis Sample.....	C-19
Table C-5. Percent of Coaching Visits With Specified Features and Time Spent in Coaching With These Features: Second-Year Teacher Impact Analysis Sample (continued)	C-20
Table C-6. Percent of Coaching Visits With Specified Features and Time Spent in Coaching With These Features, by PD Provider: Second-Year Teacher Impact Analysis Sample.....	C-21
Table C-6. Percent of Coaching Visits With Specified Features and Time Spent in Coaching With These Features, by PD Provider: Second-Year Teacher Impact Analysis Sample (continued)	C-22
Table C-7a. Percent of Implemented PD Hours Attended by the Average Teacher: Second-Year Teacher Impact Analysis Sample—America’s Choice	C-23
Table C-7b. Percent of Implemented PD Hours Attended by the Average Teacher: Second-Year Teacher Impact Analysis Sample—Pearson Achievement Solutions	C-24
Table C-8. Maximum Possible PD Dosage Based on Teacher PD Program Entry Dates: Second-Year Teacher Impact Analysis Sample	C-25
Table C-9a. Treatment and Control Group Contrast in Hours of Mathematics-Related PD, for Districts Served by America’s Choice: Second-Year Teacher Impact Analysis Sample.....	C-26

Table C-9b. Treatment and Control Group Contrast in Hours of Mathematics-Related PD, for Districts Served by Pearson Learning Solutions: Second-Year Teacher Impact Analysis Sample.....	C-27
Table C-10. Treatment and Control Group Contrasts for Features of Mathematics-Related PD: Second-Year Teacher Impact Analysis Sample (unstandardized).....	C-28
Table C-10. Treatment and Control Group Contrasts for Features of Mathematics-Related PD: Second-Year Teacher Impact Analysis Sample (unstandardized) (continued)	C-29
Table D-1. Teacher Characteristics, by Treatment Status: First-Year Teacher Impact Analysis Sample—Two-Year Districts	D-2
Table D-2. Student Characteristics, by Treatment Status: First-Year Student Impact Analysis Sample—Two-Year Districts	D-3
Table D-3. Teacher Characteristics, by Treatment Status: First-Year Teacher Impact Analysis Sample—One-Year Districts.....	D-4
Table D-4. Student Characteristics, by Treatment Status: First-Year Student Impact Analysis Sample—One-Year Districts.....	D-5
Table D-5. Impact of the PD Program on Teacher Knowledge at the End of the First Year: First-Year Teacher Impact Analysis Sample—One-Year Districts	D-6
Table D-6. Impact of the PD Program on Student Mathematics Achievement at the End of the First Year: First-Year Student Impact Analysis Sample—One-Year Districts	D-7
Table D-7. Impact of the PD Program on Teacher Knowledge at the End of the Second Year, Without Covariates: Second-Year Teacher Impact Analysis Sample	D-8
Table D-8. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year, Without Covariates: Second-Year Student Impact Analysis Sample.....	D-9
Table D-9. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year, With Teacher Covariates: Second-Year Student Impact Analysis Sample	D-9
Table D-10. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year, Using Teacher as Middle Level of Multilevel Model: Second-Year Student Impact Analysis Sample.....	D-10
Table D-11. Unadjusted Means and Standard Deviations on Teacher Knowledge and Student Mathematics Achievement: Second-Year Teacher and Student Impact Analysis Samples	D-13
Table E-1. One-Year Effect of the PD Program on Teacher Knowledge at the End of the Second Year: Second-Year Teacher Impact Analysis Sample.....	E-2
Table E-2. Per-Year Effect of the PD Program on Teacher Knowledge: Pooled Sample.....	E-4
Table E-4. Comparison of Teacher Characteristics Between Teachers in First-Year Impact Sample Only, Teachers in Second-Year Impact Sample Only, and Teachers in Both Impact Samples: Pooled Sample.....	E-6
Table E-5. Interaction Between Sample Membership (Teachers in First-Year Impact Sample Only, Teachers in Second-Year Impact Sample Only, and Teachers in Both Impact Samples) and Treatment Effect: Pooled Sample.....	E-7

Table E-6. Treatment-Control Difference in Baseline Teacher Knowledge for Teacher Samples Used in Specific Analyses.....	E-7
Table E-6. Treatment-control Difference in Baseline Teacher Knowledge for Teacher Samples Used in Specific Analyses (continued)	E-8
Table E-7. Average Annual Effect of the PD Program on Student Achievement: Pooled Sample.....	E-9
Table E-8. Student Characteristics, by Treatment Status: Pooled Sample	E-10
Table E-9. Effect of the PD Program on Teacher Knowledge: Pooled Sample, America’s Choice.....	E-11
Table E-10. Effect of the PD Program on Student Achievement: Pooled Sample, America’s Choice.....	E-12
Table E-11. Effect of the PD Program on Teacher Knowledge: Pooled Sample, Pearson Achievement Solutions.....	E-12
Table E-12. Effect of the PD Program on Student Achievement: Pooled Sample, Pearson Achievement Solutions.....	E-13
Table E-13. Effects of the Interaction Between Treatment Status and Baseline Teacher Knowledge on Teacher and Student Outcomes: Pooled Sample	E-16
Table E-14. Effects of the Interaction Between Treatment Status and Years of Teacher Experience on Teacher and Student Outcomes: Pooled Sample	E-17
Table E-15. Effects of the Linear and Quadratic Interaction Between Treatment Status and Baseline Teacher Knowledge on Teacher and Student Outcomes: Pooled Sample, Augmented Model 1	E-19
Table E-16. Effects of the Linear and Quadratic Interaction Between Treatment Status and Baseline Teacher Knowledge on Teacher and Student Outcomes: Pooled Sample, Augmented Model 2.....	E-20
Table E-17. Effects of the Quadratic Interaction Between Treatment Status and Years of Teaching Experience on Teacher and Student Outcomes: Pooled Sample, Augmented Model 1.....	E-21
Table E-18. Effects of the Linear and Quadratic Interaction Between Treatment Status and Years of Teaching Experience on Teacher and Student Outcomes: Pooled Sample, Augmented Model 2.....	E-22
Table E-19. Effects of the Interaction Between Treatment Status and Baseline Student Achievement on Student Outcomes: Pooled Sample.....	E-24
Table E-20. Minimum Detectable Effect Sizes (MDEs) for Interaction Between Treatment Status and Baseline Teacher Knowledge, Years of Teaching Experience, and Student Achievement: Pooled Sample.....	E-25
Table E-21. Variance Decomposition of Standardized Student Spring Total NWEA Test Scores by Data Structure Level: Pooled Sample.....	E-29
Table E-22. Interaction Between Teacher Knowledge and Sample Membership (Teachers in First-Year Impact Sample Only, Teachers in Second-Year Impact Sample Only, and	

Teachers in Both Impact Samples) in Regression Predicting Student Achievement: Pooled Sample..... E-30

Table E-23. Variance Decomposition of Standardized Student Spring Total NWEA Test Scores by Data Structure Level, Four-Level Model: Pooled Sample E-32

Table E-24. Standardized Regression Coefficients for the Relationships Between Teacher Knowledge and Student Achievement, Using Four-Level Model: Pooled Sample..... E-33

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

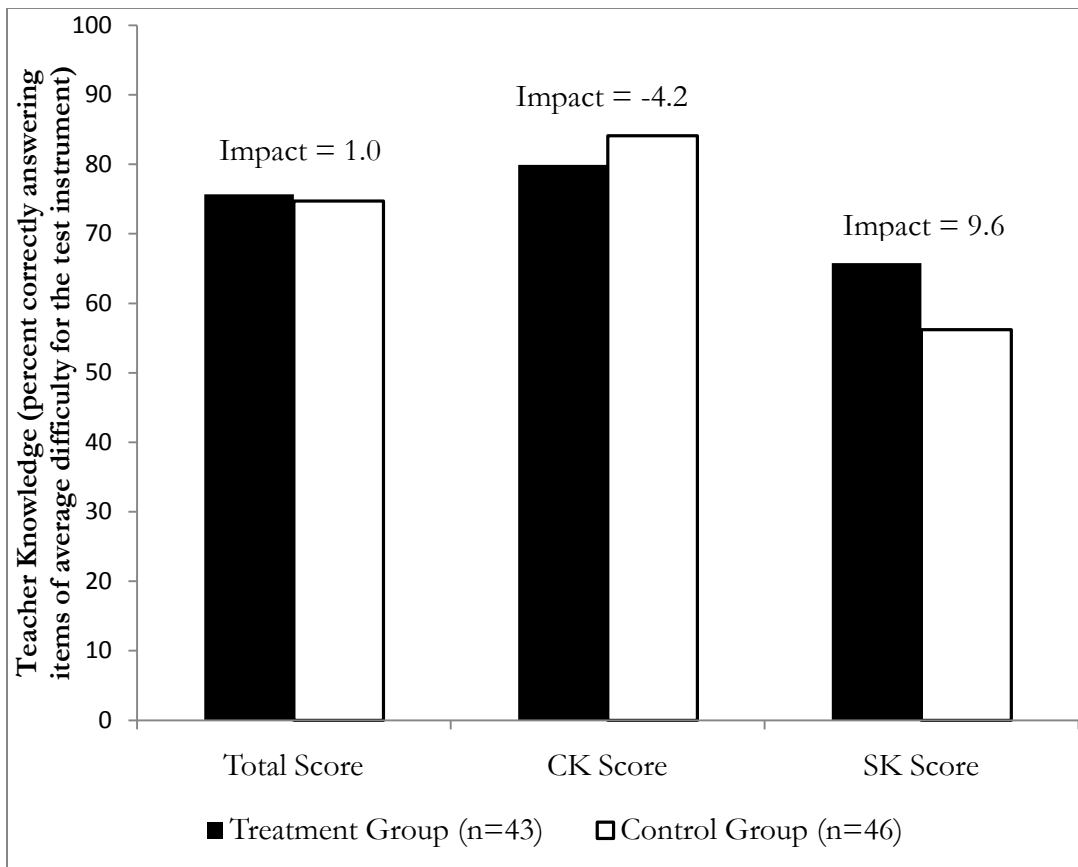
Impact Findings

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.

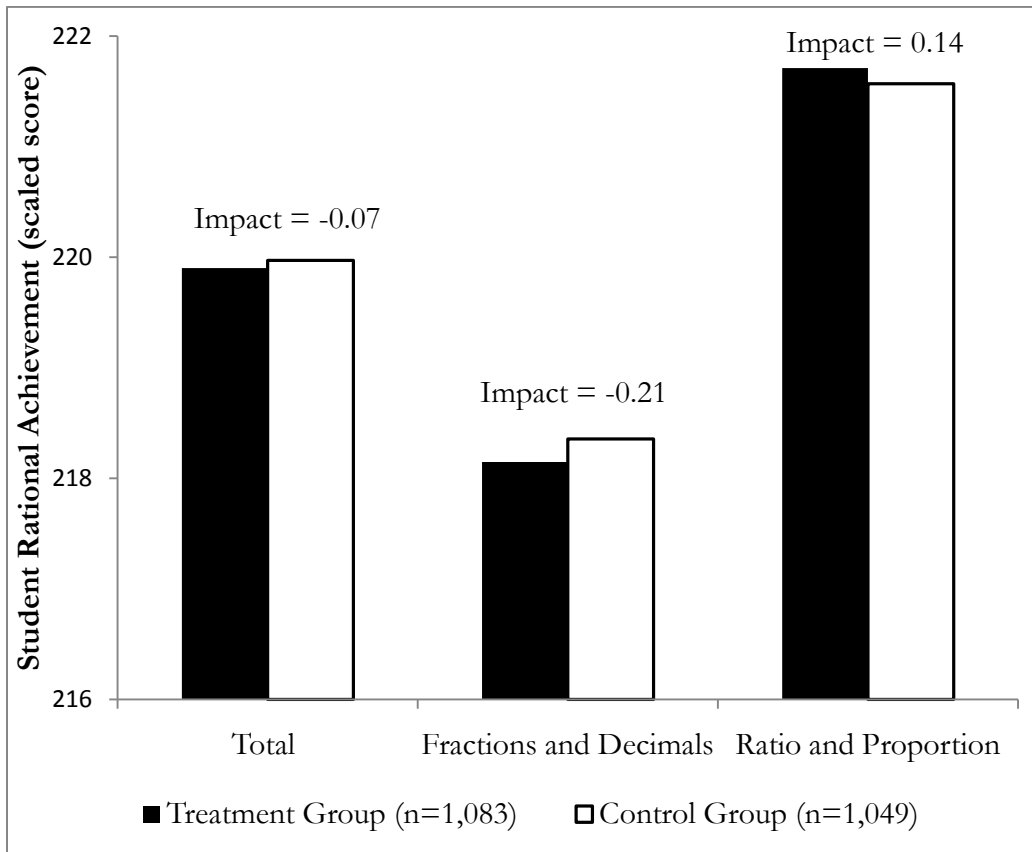
P-values are based on t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

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.

CHAPTER 1

OVERVIEW OF THE STUDY

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.

Background and Importance of the Study

Student achievement in mathematics has been a focal concern in the United States for many years. As part of an overall strategy to boost achievement in this discipline, two major reports in the past decade have called for all students to learn algebra by the end of eighth grade (National Research Council 2001; National Mathematics Advisory Panel 2008). Both reports 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 grade 7, “students should be able to solve problems involving percent, ratio, and rate, and extend this work to proportionality” (p. 20). Recommendations from these two reports are reflected in the 2010 Common Core State Standards (CCSS), which present rational number topics in grades 3 through 7 to allow sufficient time for students to master the concepts before studying algebra in depth in grade 8 (Common Core State Standards Initiative 2010).

One impediment to students’ mastery of rational numbers may be deficits in teachers’ knowledge. A recent study of elementary teachers showed that the teachers’ mathematics knowledge correlated with their student’s gains in mathematics (Hill, Rowan, and Ball 2005). Another study, which examined rational number knowledge among 136 pre-service elementary teachers, found that many had difficulty dividing fractions, although they were more successful with addition, subtraction, and multiplication of fractions (Tirosh, Fischbein, Graeber, and Wilson 1999). The prospective teachers in this study also demonstrated weak conceptual understanding and had trouble constructing representations of key concepts. For example, 43 percent claimed that there is no number between $1/5$ and $1/4$, and very few were able to go beyond area model representations of fractions to construct set models, number lines, or ratio models. Newton (2008) has reported similar findings for a different sample of pre-service elementary teachers.

No nationally representative studies have focused on middle school mathematics teachers’ knowledge of rational number topics. However, national survey data show that in 2003–2004, some 66.6 percent of public school teachers assigned to teach seventh-grade mathematics did not hold a degree in mathematics.¹⁵ Further, one recent study, which administered a mathematics test to a random sample of middle school mathematics teachers in the United States, found that those

¹⁵ Authors’ tabulations from the 2003–2004 Schools and Staffing Survey.

teaching in low-income schools had lower levels of mathematics knowledge than their peers in more affluent schools (Hill 2007).

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. This is one of two studies of content-focused teacher PD that NCEE has initiated to add to the research base on the effectiveness of PD. The first, published in 2008, examined the impact of a PD program that focused on the teaching of early reading (Garet et al. 2008).

Like the first study of content-focused PD published by NCEE (Garet et al. 2008), the Middle School Mathematics PD Impact Study tests the effect of a PD program when implemented with a relatively large sample, in varied settings, and using multiple facilitators. The study 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 et al. (2007) involved smaller samples of 5 to 44 teachers, and the PD programs were delivered by the individuals who developed them.

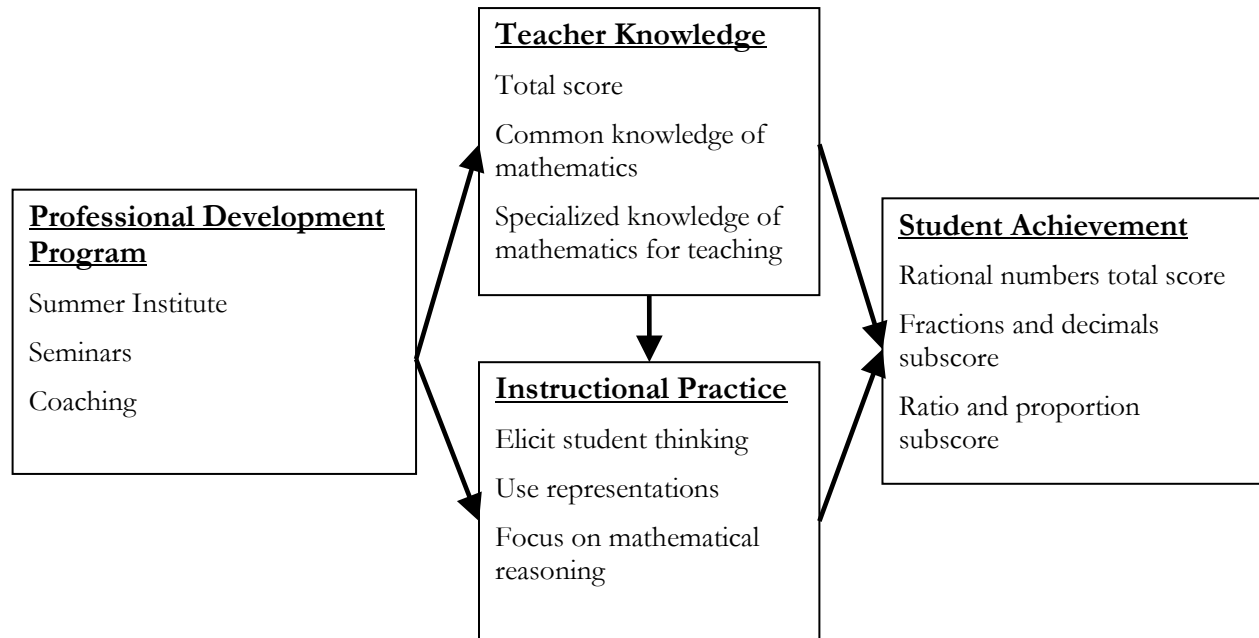
Overview of the First- and Second-Year Study Designs

The Middle School Mathematics PD Impact Study used a cluster random assignment design in which schools within each of 12 districts were randomly assigned either to receive the PD program tested by the study or to continue with whatever PD was normally provided by the district. In the first year, the study addressed three research questions:

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

The hypothesized relationship among the PD program and the three classes of outcomes that address these research questions are shown in the simplified theory of action diagram in Exhibit 1-1.

Exhibit 1-1. Theory of Action



The main findings of the first year of the study are summarized below.

Findings From the First Year of the Study (Garet et al. 2010)

- The first year’s PD program was implemented as intended in terms of the proportion of planned hours of PD delivered, the proportion of delivered PD hours attended by treatment teachers, and the degree to which the total mathematics-related PD hours for the year reported by treatment teachers exceeded those reported by control teachers (the magnitude of the service contrast).
- The one-year PD program provided by the study did not have a statistically significant impact on teacher knowledge of rational number topics.
- The one-year PD program had a statistically significant and positive impact on the frequency with which treatment teachers engaged in instructional activities that elicited student thinking, but not on the other two practice outcomes tracked by the study.
- The one-year PD program did not have a statistically significant impact on student achievement in rational number topics.

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 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 the PD was offered in only half of the original participating districts, and the observation-based measure of instructional practice was dropped.¹⁶ Thus 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?

This report focuses on findings after the second year of the study in the six districts in which two years of PD were delivered. The remainder of this chapter provides an overview of the recruitment and assignment of districts, schools, and teachers over the two years of the study, as well as information about the data collected and the analysis methods employed during the second year. Chapter 2 describes the design and implementation of the PD program and the extent of service contrast between the treatment and control groups. Chapter 3 presents the analyses of the cumulative impact of the PD program after the second year of the study. Chapter 4 summarizes the impact findings across the two years of the study and describes the results of supplemental analyses that provide context for the main results.

Finally, a word about nomenclature: In this report, the terms *first-year* and *second-year* refer to the two years in which the PD was implemented: the 2007–2008 and 2008–2009 school years, respectively. (For example, we refer to the first-year PD, the first-year impact analysis sample, and the first-year findings.) We refer to the six districts that continued in the study for both years as the *two-year districts* and the six districts that participated in the study only during the first year as the *one-year districts*. Findings from the first year of the study can be subdivided into findings obtained in the one-year districts and findings obtained in the two-year districts. Second-year findings, by necessity, are restricted to the two-year districts.

Recruitment and Random Assignment in the First and Second Years of the Study

First Year of the Study

The districts that participated in the second year of the study are a subset of the districts recruited for the first year of the study. 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 sufficient power to detect impacts

¹⁶ See Appendix A for a comparison between the school, teacher, and student samples in the six districts that participated in both years and the school, teacher, and student samples in the six districts that participated in the first year of the study only.

of the anticipated magnitude in teacher and student outcomes.¹⁷ (See Appendix A for additional detail about the selection and recruitment of districts for the first year of the study.)

The districts for the first year of the study were selected from a national sample of districts that met the following criteria:

- The district contained at least four qualifying schools (regular schools in which at least one-third of students were eligible for free or reduced-price lunch and in which at least 150 seventh-grade students were enrolled).^{18,19}
- The district used one of the three mathematics curricula that were in widest use among districts meeting the size criteria, namely, *Connected Mathematics (CMP)*, *Glencoe McGraw Hill Mathematics: Applications and Concepts (Glencoe)*, or *Prentice Hall Mathematics (PH Mathematics)*.²⁰
- The district did not provide districtwide PD in mathematics instruction of the same type and level of intensity as that being provided by the study.²¹

The 12 districts and 77 schools recruited to participate in the first year of the study were located in nine states. Six of the districts used *CMP* and the other six used *Glencoe* or *PH Mathematics*.²² The study's two PD providers—America's Choice and Pearson Achievement Solutions—were each assigned to implement the PD in three districts from each curriculum category.

Within each district, schools were randomly assigned to treatment and control groups. In six of the districts, officials asked that the assignment process ensure that schools with particular characteristics (e.g., geographic location, demographic characteristics, past academic performance) be equally represented in the treatment and control conditions. Schools within these districts were grouped into two or three blocks of schools with similar characteristics, and half the schools within

¹⁷ Relevant studies of PD interventions focused on teachers' content knowledge that were reviewed by Kennedy (1998) obtained effect sizes of 0.4 or larger for some student outcomes – substantially greater than the 0.2 minimum detectable effect size target for our design, but with interventions of greater intensity and volunteer teachers. Another review of the effects of mathematics and reading PD programs on student achievement, subsequently published as Yoon et al. (2007), also calculated average effect sizes that were larger than our minimum detectable effect size targets.

¹⁸ As defined in the *Common Core of Data (CCD)*, a regular school is a public elementary/secondary school that does not focus primarily on vocational, special, or alternative education (Sable and Plotts 2010, pg C-13).

¹⁹ This grade-level enrollment criterion was used because schools serving at least 150 seventh-grade students are likely to have more than one teacher assigned to seventh-grade mathematics. The study design called for multiple teachers per school to allow computation of key variance components needed for the impact analyses.

²⁰ The number of curricula was constrained to facilitate the task of designing the PD to be relevant to the curricula of participating teachers.

²¹ Districts remained eligible for the study if they provided PD in mathematics instruction that targeted teachers of students in grades other than seventh, involved fewer than 10 hours of training, was attended by individual teachers rather than teams of teachers from the same schools, or focused on topics such as classroom management rather than the theory and practices of mathematics instruction. Districts that assigned mathematics coaches to support the entire teaching staff of one or more schools or to support teachers of students in the seventh grade were eligible for the study, provided that the district's coaching would not create scheduling problems or excess burden for teachers participating in the study.

²² *CMP* differed from *Glencoe* and *PH Mathematics* on a number of dimensions, including chapter organization, lesson components, instructional approaches supported, and content emphasized. To investigate the possibility of an interaction between curricular context and PD impact, the sample in the first year was constructed to support two parallel substudies of the same design but in different curricular contexts.

each block were randomly assigned to the treatment group. The first-year school sample consisted of 40 treatment and 37 control schools.²³

Once schools were randomly assigned, all eligible teachers teaching at least one regular seventh-grade mathematics class in each school in the 2007–2008 school year became members of the teacher sample for the first year of the study, and all seventh-grade students in the teachers’ regular seventh-grade mathematics classes became members of the first-year student sample.²⁴ The first-year impact analyses were based on a sample of 100 treatment and 95 control teachers who were teaching eligible classes in spring 2008.

Second Year of the Study

Three criteria guided the selection of the six districts to participate in the second year of the study. First, the study team wanted to avoid districts in which large numbers of student transfers were expected to affect the study schools during the second year of the study.²⁵ The second criterion was to balance the number of districts assigned to each PD provider—three districts for each provider, and the third criterion was to maximize the school sample size in order to also maximize statistical precision. Therefore, after excluding districts that were expecting large numbers of student transfers to affect study schools, we selected the three districts for each provider with the largest number of schools in the sample.

The districts for the second year of the study were selected before the first-year results were known. Thus, findings about the impact of the first year of 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.

Characteristics of the Schools in the Two-Year Districts

The district sample for the second year of the study (two-year districts) comprised six districts, including some that used *CMP* and others that used either *Glencoe* or *PH Mathematics*.²⁶ Schools remained in the treatment groups to which they had been assigned in the first year. Table 1-1 provides descriptive information about the characteristics of the sample of 39 schools (20 treatment and 19 control) 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.

- The schools in the two-year districts were largely urban (87 percent were located in large or mid-sized cities) and served many disadvantaged students (two-thirds of the schools were Title I schools, and on average two-thirds of their students were eligible for free or reduced-price lunch).

²³ See Garett et al. (2010) for a more detailed discussion of the recruitment and random assignment procedures used in the first year of the study.

²⁴ “Eligible teachers” were defined as regular teachers and long-term substitutes, but not short-term substitutes. “Regular mathematics classes” were defined to exclude advanced classes and classes for students with special education needs.

²⁵ Such disruptions were expected to occur, for example, when a study school was scheduled to be converted to serve a different grade range or a specialized population of students.

²⁶ The proportions of *CMP* and *Glencoe/PH Mathematics* districts served by the two providers were identical.

- Relative to the national sample of schools in large districts eligible for the study, the schools in the two-year districts were less likely to be in the South and more likely to be in the Northeast and to be in cities rather than 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).

In short, the schools in districts recruited for the second year of the study served high proportions of disadvantaged students. However, the characteristics of these schools diverged in several respects from those of the national sample from which they were drawn and consequently should not be considered as representative of that sample.^{27,28}

²⁷ Table A-1 in Appendix A compares schools in the two-year and one-year districts. There were statistically significant differences between schools in the two district samples in terms of location (region of the United States), urbanicity, Title I status, number of seventh-grade students enrolled, and school type, but not in terms of race/ethnic composition or proportion of students eligible for free or reduced-price lunch.

²⁸ For more context on the differences between the two-year and one-year districts, Tables A-2 and A-3 in Appendix A compare the characteristics of teachers and students in the first-year impact samples from the one-year and two-year districts. Teachers in the two-year districts scored higher at baseline on the teacher knowledge test of rational number topics and had taken more college mathematics and mathematics education courses than their counterparts in the one-year districts. Students in the two-year districts scored higher on the mathematics test administered by the study at baseline and also on their sixth-grade state accountability tests. They were younger than students in the one-year districts and included more students with English as a second language or special education status.

Table 1-1. 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 (20 treatment, 19 control) 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.

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 (*).

See Table A-1 in Appendix A for a comparison of the characteristics of schools in the one-year and two-year districts.

Second-Year Analysis Samples

Teacher sample. The *second-year teacher impact analysis sample* consists of the 92 teachers in the two-year districts who were teaching eligible classes in the spring of the second year of the study (spring 2009). Among these teachers, 51 (including 23 in the treatment group and 28 in the control group) had participated in the study since baseline (fall 2007). Another 44 teachers (including 22 in the treatment group and 22 in the control group) who were present at the end of the first year of the study left the study schools before the end of the second year and thus were not included in the second year teacher impact analysis sample. Exhibit 1-2 provides more detail about the turnover of teachers in the two-year districts across the two years of the study.

Exhibit 1-2. Teacher Turnover During the Two Years of the Study: Two-Year Districts

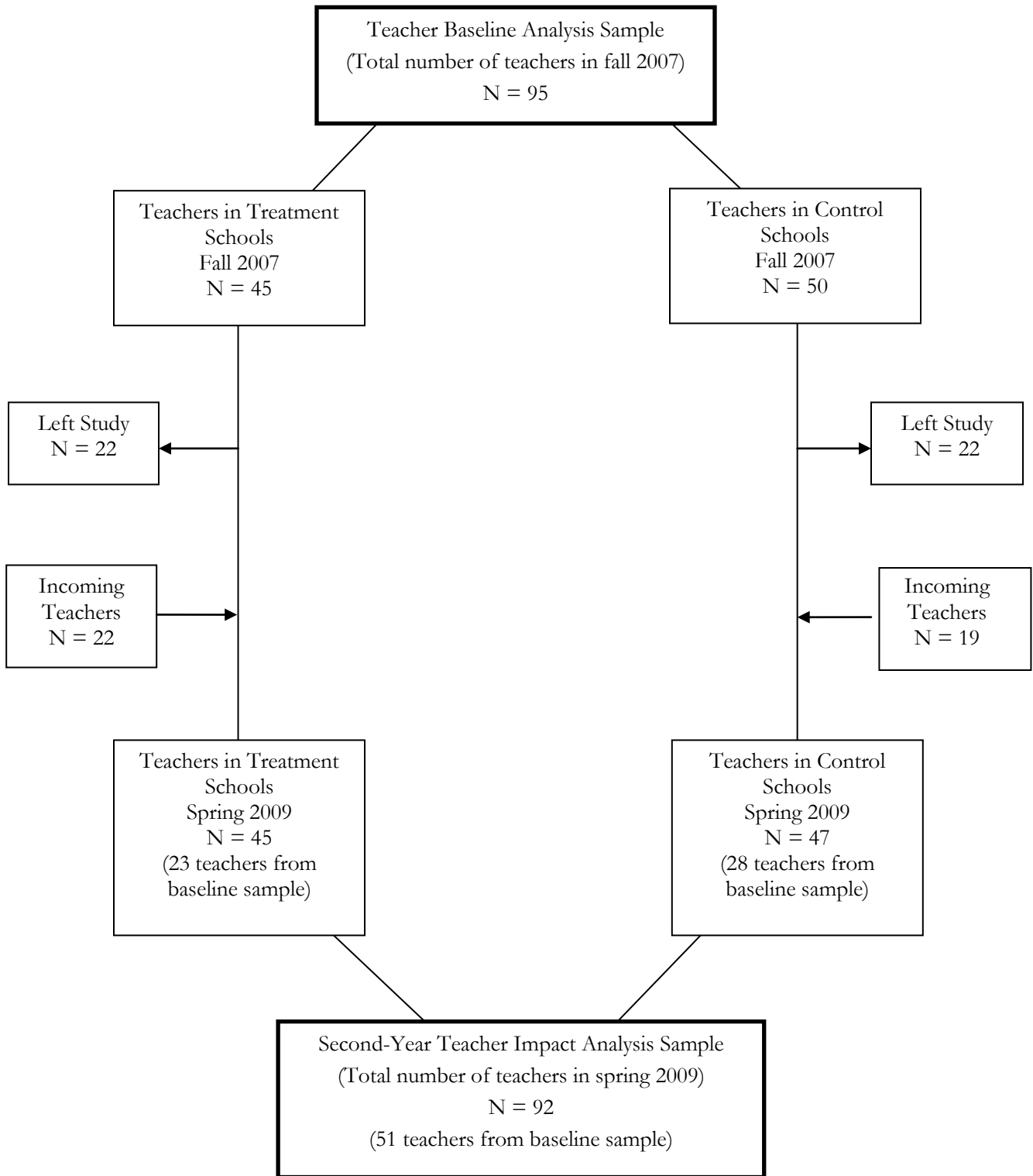


Table 1-2 provides descriptive information about the characteristics of the teachers in the second-year impact sample compared with the characteristics of seventh-grade mathematics teachers working in similar large districts.

- All teachers in the second-year impact sample had at least a bachelor's degree and 42 percent had a master's degree. Less than one-fifth (19 percent) had majored in mathematics, and over two-thirds (70 percent) had had four or more years of teaching experience when they entered the study.
- There were no statistically significant differences between the sample of seventh-grade mathematics teachers in the second-year impact sample and seventh-grade mathematics teachers in the national sample of schools in large districts eligible for the study on any of the characteristics for which the two samples could be compared, including certification, degrees, majors, and years of teaching experience.

Thus, although the *schools* in the two-year districts differed in some ways from schools in the national sample of similar districts from which the study districts were recruited, the *teachers* in the second-year impact sample resembled teachers in a national sample of seventh-grade mathematics teachers in large urban school districts.

Table 1-2. 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 Seventh-Grade Mathematics Teachers 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
Years of Teaching Experience in Current School	4.0	
Years of Teaching Experience in Middle School Mathematics	7.8	
Years of Experience With Current Mathematics Curriculum	6.4	
Number of Postsecondary Mathematics Courses Taken	7.3	

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

Student sample. The focus of the second year student impact analysis was all students who, in the spring of 2009, were enrolled in the teachers' eligible seventh-grade classes.²⁹ However, because of logistical and budgetary constraints, it was not possible for the study to administer the computer-based rational number test (described below) to all these students. Instead, the study team randomly selected a representative sample of eligible students from each regular seventh-grade mathematics class to take the rational number test that was used as a major outcome variable in the study.^{30,31} The *second-year student impact analysis sample* consists of the 2,132 students who were

²⁹ Thus, the students in the second-year impact analysis sample are distinct from the seventh-grade students who participated in the first year of the study.

³⁰ Students who were sampled for potential testing were evaluated by school personnel to determine whether testing was appropriate. Some students in regular mathematics classes might not have been able to participate meaningfully in testing under the conditions offered by the study because of their disabilities or English learner status. When school personnel determined that a student could not participate meaningfully, the student was removed from the sample. Fewer than five percent of sampled students were excluded on this basis.

³¹ Unless their parents denied permission, we also obtained district records for *all* students enrolled in the teachers' eligible seventh-grade mathematics classes in spring 2009. The district records contained demographic information and scores on district-administered mathematics tests.

randomly selected to take the rational number test in the spring of the second study year (spring 2009).^{32,33} Table 1-3 summarizes impact analysis sample sizes for the second year study.

Table 1-3. 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.

Sources of Data

Data sources used in the second year of the study were designed to serve the following three purposes:

1. To document the implementation of the PD program and the extent of service contrast between treatment and control teachers across the two years of the study
2. To describe the characteristics of participants at baseline and provide covariates for the impact analyses
3. To measure two of the intended outcomes of participation in the PD: teacher knowledge and student achievement³⁴

The following instruments were used as data sources.

PD implementation records. To gauge the implementation of the PD, the study team collected three types of data. Study team members completed an implementation form for each institute or seminar day to record the amount of time devoted to each instructional segment and the use of intended instructional materials. After each coaching event, the facilitators completed coach logs to record the amount of contact time with each teacher and the kinds of coaching activities pursued. Finally, detailed attendance records kept for each PD event were combined with data from the coaching logs to calculate the treatment dosage received by each teacher in the treatment group.

³² A separate sample of students was randomly selected to take the rational numbers test in the fall of the second study year (fall 2008). Data from the fall administration (including both school average test scores and individual student test scores, when available) were used as covariates in the impact analyses.

³³ Each student sample was representative of the classrooms as they were constituted at the time of testing, and the spring sample contained the same balance of continuing and incoming students as the classrooms as a whole. However, in drawing the continuing students for the spring sample, priority was given to students who had also been selected for the fall testing. This served to maximize the numbers of students who had fall test scores to use as covariates in the impact analyses. See Exhibits A-1 and A-2 in Appendix A for a summary of student movement into and out of the second-year study classrooms and for a more detailed summary of the sampling of students to take the fall 2008 and spring 2009 NWEA rational number tests. Table A-4 in Appendix A compares the characteristics of students selected to take the spring 2009 NWEA test (i.e., selected for the second-year impact analysis sample) with those of students who were not selected.

³⁴ As noted previously, teacher practice measures were not collected in the second year of the study, so the cumulative impact of PD on teacher practice could not be evaluated.

Teacher surveys. Teacher surveys dealing with teacher characteristics and/or the teachers' professional development experiences were administered four times during the study—at the beginning of the first year, the end of the first year, the beginning of the second year (only for teachers who were new to the study), and the end of the second year. *Teacher characteristics* were addressed through survey questions about the education and years of teaching experience that the teachers had when they joined the study, as well as questions about the professional development in which the teachers had participated during the year preceding entry into the study.³⁵ To assess whether the program provided a meaningful *contrast in service* between the treatment and control groups, the surveys included questions on the nature and extent of mathematics-related PD that the teachers experienced *during* their period of participation in the study.

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).³⁶

Tests administered to treatment and control teachers at their point of entry into the study provided baseline data on teachers' knowledge of rational numbers topics (and were included as covariates in the impact analyses). Tests administered to all treatment and control teachers in spring 2009 provided one of the two main outcomes for the second-year study.

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 the larger NWEA item bank of scaled, operational mathematics items. 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*.

Fall 2008 scores provided descriptive information about the students' level of achievement at baseline and were included as covariates in impact analyses. Spring 2009 NWEA scores constituted a second key outcome for the second year of the study.

District records. To further characterize the students in the study schools and to provide additional student-level covariates for the impact analyses, prior-year state achievement scores and basic demographic data were requested from district records for all students enrolled in the sampled classes unless parental permission was withheld.

Further details on each of the instruments used in the study are provided in Appendix B.

³⁵ Teacher characteristics were included as covariates in impact and exploratory analyses. Teachers who did not complete the fall teacher surveys (because, for example, they entered the study mid-year) were given supplemental questions appended to their spring surveys for the purpose of gathering these data.

³⁶ 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.

Analytical Approaches

This section discusses the analytic methods used in the second year of the study to estimate impact. It first describes the statistical models used to estimate the impact of the program, next explains how the impact results are presented, and finally reviews the statistical power of the study (i.e., the precision with which the analysis can measure program impact).

Statistical Models for Estimating Impact

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 6 two-year districts in a single analysis, treating the districts as fixed effects.³⁷ 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.³⁸ 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 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. Since the impact estimates are based on an “intent to treat” analysis (as explained further below), the results also do not necessarily reflect the results that might have been obtained if all treatment teachers had had a full, two-year dose of the study PD.

The second year of the study focused on two outcome domains: teacher knowledge and student achievement. The impact analyses were conducted using the full (impact) sample of teachers and students present in the study schools as of the spring 2009 data collection period.³⁹ For outcomes that were measured at the teacher level, a two-level hierarchical linear model was used, with teachers nested within schools. Every teacher within a district was weighted equally (i.e., an implicit weight of 1 was applied for each teacher). To improve the precision of the impact estimates, the model included a set of teacher-level covariates that were considered likely to be related to teacher knowledge. In addition to the baseline teacher knowledge total scores, these covariates were teacher's experience, teacher's education level, whether the teacher had a mathematics major, and number of postsecondary mathematics and mathematics education courses a teacher had taken (all as measured at the time of the teacher's entry into the study).

For the student achievement outcomes, we used a three-level hierarchical model, in which students were nested within teachers' classrooms and classrooms were nested within schools. Each student in the sample was weighted equally. Because equal numbers of students were sampled from each class, weighting each student equally was approximately equivalent to weighting each sampled class equally.⁴⁰ The covariates in the student achievement model included a single school-level covariate—the school average baseline (fall 2008) NWEA test score—as well as the following

³⁷ Schools, classes, and students were treated as random effects.

³⁸ This approach used the data for all six districts in a single analysis, assuming a common set of school-, teacher-, and student-level error terms across districts. This method allowed us to examine how the impact of the PD program varied across districts and whether these differences were statistically significant.

³⁹ As noted previously, because of resource constraints related to the computer-administered student outcome measure, the impact sample of students was a random sample of all the students present in the eligible classrooms in spring 2009.

⁴⁰ For a detailed discussion of student sampling, see Appendix A.

student-level covariates: baseline (fall 2008) achievement score, gender, age, race/ethnicity, student's English as a second language (ESL)/limited English proficiency (LEP) status, special education status, and free or reduced-price lunch status.⁴¹

Teachers or students with missing outcome measures were dropped from the impact analysis for which they lacked data.⁴² In cases with missing covariate measures, the missing data were replaced with zeros and a dichotomous variable indicating the missing status of a given covariate for each observation was added to the impact analysis model.⁴³

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 regular seventh-grade mathematics 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.⁴⁴

Additional analyses described in Chapter 4 report first-year impact estimates separately for the one-year and two-year districts, examine the one-year effect of the PD at the end of the second year, and pool all cases across both years of the study in order to increase the statistical power of the analyses. The pooled sample allowed us to address questions for which the size of the second-year impact sample was insufficient, such as the impact of the PD separately by provider, the interaction between teacher or student characteristics and the effect of the PD, and the relationship between teacher knowledge and student achievement. Taken together, these analyses provide context for the main study findings and suggest hypotheses that might be worth exploring in future efforts.

Understanding the Impact Tables in This Report

To provide context for interpreting estimated impacts, tables that report the estimated impacts of the PD present the regression-adjusted mean outcome levels for the treatment and the control groups. The mean program impacts were estimated using the impact models described above, which used all available observations from both the treatment group and the control group. (See Black et al. 2008 and Garet et al. 2008 for a discussion of this calculation method.)

To calculate the *regression-adjusted mean outcome levels* for the two groups, the mean outcome levels were adjusted by using the observed mean covariate values for the treatment group in the

⁴¹ This school average baseline NWEA test score variable was calculated using all valid and usable fall 2008 student NWEA test scores.

⁴² Of the 92 teachers who were teaching eligible classes in spring 2009 and therefore were included in the second-year teacher impact analysis sample, 3 were missing data on the teacher knowledge outcome measures. Of the 2,449 students on the spring 2009 class rosters who were sampled and eligible for testing, 2,132 had usable test data and were included in the second-year student impact analysis sample. These 2,132 students comprised 86 percent of the sampled and eligible students in the treatment group and 88 percent of the sampled and eligible students in the control group. See Exhibit A-2 in Appendix A for details of the construction of the second-year student impact analysis sample.

⁴³ For more detailed information about the statistical models, see Appendix B. Table B-4 shows the number of cases in the second-year teacher and student impact analysis samples for which data were missing on characteristics used as covariates in the impact models.

⁴⁴ For teachers in the second-year impact analyses, the maximum duration of study participation was two years (school year 2007–2008 and school year 2008–2009); for students the maximum duration of study participation was one year (school year 2008–2009).

estimated impact model. In other words, means for both groups were “regression adjusted” using the treatment group’s observed means as a common set of covariate values.

- For schools randomly assigned to the treatment group, the regression adjusted mean outcome levels equal the mean outcome levels for treatment schools. (That is, adjusting mean values of the treatment group outcomes by observed treatment group covariate values leaves the mean outcome values unchanged.)
- For schools randomly assigned to the control group, the regression adjusted mean outcome levels represent how the treatment group schools would have performed had they been randomly assigned to the control group. In other words, they represent the counterfactual.

In the impact tables and the relevant text in the report, the regression-adjusted mean outcomes for the treatment group are labeled as “treatment group” outcomes, and the regression-adjusted mean outcomes for the control group are labeled as “control group” outcomes.

Sample Sizes and Statistical Power

A common way to represent statistical precision is as a minimum detectable effect (MDE), which is the smallest true effect that an estimator has a “good chance” of detecting (Bloom 1995). We use the standard convention of defining a minimum detectable effect as the smallest true impact that has an 80 percent chance of being found to be statistically significant at the 5 percent level of statistical significance for a two-tailed test. When a minimum detectable effect is measured in standard deviation units, it is referred to as a minimum detectable effect size (MDES).⁴⁵

Table 1-4 reports MDES estimates for the program impact on the teacher and student outcomes for the second-year impact sample, which was about half the size of the first-year impact sample. These estimates are based on available data from the second year of program implementation; that is, they represent the realized precision of the study in the second year. The MDES for teacher knowledge scores ranged from 0.52 to 0.59 standard deviations. The MDES for student mathematics achievement were all 0.20 standard deviations.

⁴⁵ Throughout this report, the standard deviations of the outcome measure for the control group members in the full first-year impact sample (12-district sample) are used in calculating effect sizes. This was done to allow comparison of effect sizes between the first and second years of the study.

Table 1-4. Second-Year Minimum Detectable Effect Sizes (MDES) for Core Outcomes: Second-Year Teacher and Student Impact Analysis Samples

Outcome Measures	MDES
Teacher Knowledge	
Total Score (logits)	0.59
Common Knowledge of Mathematics (CK) Score (logits)	0.52
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.59
Student Mathematics Achievement	
NWEA Rational Number Test Total Score (scale score)	0.20
Fractions and Decimals Score (scale score)	0.20
Ratio and Proportion Score (scale score)	0.20

SOURCE: Spring 2009 Teacher Knowledge Test; Spring 2009 NWEA Rational Number Test.

NOTES: MDESs are based on the standard errors and standard deviations of the second-year impact estimates.

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

Effect sizes were calculated using the control group standard deviation for the first-year impact analysis sample (12-district sample). On the Teacher Knowledge Test, the control group standard deviation was 0.97 for the *Total Score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*. On the NWEA Rational Number Test, the control group standard deviation was 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for the *Ratio and Proportion Score*.

CHAPTER 2

DESIGN AND IMPLEMENTATION OF THE PD PROGRAM

This chapter describes the content and structure of each component of the professional development (PD) program, examines the implementation of the PD program, and compares the mathematics PD experienced by treatment and control teachers for both years of the study. How and how well the PD program was implemented are important factors in understanding the impact that the program had on teachers and students.

Design of the PD Program

The PD program delivered across the two years of the study was designed to increase teachers' capability to teach positive rational number topics effectively. In each year, the PD included summer institute days, seminar days scheduled during the school year, and in-school coaching for all eligible seventh-grade teachers in treatment schools. (See Table 2-1.) The intended dosage of content-focused PD in each year (68 hours in the first year and 46 hours in the second year) was higher than the dosage of content-focused PD that most mathematics teachers typically receive in a single year.^{46, 47}

Table 2-1. Days and Per-Teacher Hours of PD Offered During the 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.

As noted in Chapter 1, a number of teachers left or joined the study during the course of the intervention. Most of the turnover occurred during the school break between the first and second years of the study. The design of the PD program attempted to address teacher turnover by offering teachers who joined the study after the first-year summer institute a supplemental makeup

⁴⁶ A national survey of teachers completed in 2005–2006 found that 11 percent of elementary teachers and 22 percent of secondary mathematics teachers participated in professional development in mathematics lasting more than 24 hours (U.S. Department of Education 2009, p. 95).

⁴⁷ The planned duration of the PD in the second year differed from the planned duration in the first year, both to accommodate district preferences and to ensure that the PD addressed the needs of both continuing and new teachers. Consistent with these goals, the second year of PD was somewhat shorter than the first year, but still more intensive than what teachers typically receive. Continuing teachers were offered 46 hours of PD in the second year, and teachers new to the study were offered 46 hours plus an additional 12 hours of summer institute to help them catch up with the content delivered in the first year.

version of that institute. The makeup institute, which was offered just prior to the second-year summer institute for all teachers, lasted two days and covered a subset of the content and activities that had been included in the three days of the first-year summer institute. Like the rest of the PD, the makeup institute focused on activities designed to strengthen teachers' understanding of rational number topics for teaching (specialized knowledge of mathematics for teaching [SK]). The first day of the makeup institute was devoted to fractions and decimals, and the second day was devoted to ratio, proportion, and percent.^{48,49}

In each district, facilitators delivered the summer institutes and seminars to training groups that included the seventh-grade mathematics teachers from all the district's treatment schools as well as the mathematics teacher leaders, department chairs, and resource teachers who worked with them. The seminars and coaching visits for each district were scheduled, to the extent possible, to align with periods in which rational number topics were being covered in that district's seventh-grade mathematics curriculum.⁵⁰

Two providers—America's Choice and Pearson Achievement Solutions—were selected through a competitive process to deliver the PD program in both years according to a common set of guidelines regarding the structure of the PD program, the knowledge to be developed, and key aspects of the delivery of the PD. The study design required both PD providers to deliver the same intended dosage and to adhere to a common set of objectives, rational number topics, and PD features, described in more detail below. But because the providers, by design, each built on their own expertise and pre-existing materials to address topics in rational numbers, they differed in how they planned to structure teacher learning activities and present the content to teachers.⁵¹

Within the domain of rational numbers, the PD program was designed to focus on fractions, decimals, ratio, rate, proportion, and percent. Across the eight institute and seminar days in the first year of the study, the program was designed to provide equal coverage to fractions and decimals (four days) and ratio, rate, proportion, and percent (four days) and to emphasize 12 key understandings in rational numbers.⁵² Based on the PD providers' reflections on the first year of the program, the providers selected material for the second year that they believed would reinforce and deepen the teachers' understanding, particularly in areas where the teachers had seemed weakest during the first year. PD provider America's Choice designed its five days of institutes and seminars

⁴⁸ Appendix C includes a complete list of the topics covered by each provider on each day of the makeup institute.

⁴⁹ No explicit makeup was given for the content missed in the first-year seminars. The makeup institute was not intended to fully replace the PD that new teachers did not experience, but rather to equip them with the foundational content needed to make use of the second-year PD institute, seminars, and coaching that were to follow.

⁵⁰ Districts differed in the amount of time devoted to rational number content, in the order in which topics were covered in the curriculum, and in the timing of this content during the school year. According to district curriculum pacing guides in effect during the period of the study, in the six districts that participated in both years of the PD (the two-year districts), positive rational numbers were the planned topic of instruction for an average of 36 percent of the school year. Time on rational numbers ranged from 15 percent in the district that planned to spend the least time on rational numbers to 54 percent in the district that planned the most. Appendix C provides further detail on the percentage of the school year allocated to rational number topics in each district.

⁵¹ 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.

⁵² See the discussion of the teacher knowledge test in Appendix B for a description of the 12 key understandings that underlay both the knowledge test and the PD.

in the second year to allocate more coverage to ratio, rate, and proportion (four days) than to fractions and decimals (one day), while Pearson Achievement Solutions designed its second-year institutes and seminars to provide equal coverage to fractions and decimals (two and a half days) and ratio, rate, and proportion (two and a half days).⁵³

The content of the PD was consistent with the two recommendations in the IES Practice Guide on Fractions (which appeared after our study was under way) that had the strongest levels of supporting evidence—“Moderate Evidence” (Siegler et al. 2010). For example, for each rational number topic area, the PD program design emphasized the use of precise definitions and the properties and rationales underlying common procedures used with rational numbers. This emphasis on properties and rationales for procedures is consistent with recommendation 3 of the Practice Guide—“Help students understand why procedures for computations with fractions make sense.” The Guide cites several studies in support of this recommendation, including Rittle-Johnson, Siegler, and Alibali (2001) and Rittle-Johnson and Koedinger (2002; 2009), which found that students’ computational skills in decimals improved when they gained conceptual understanding. The National Research Council (2001) also suggested that conceptual understanding and procedural fluency should be presented in an integrated fashion, rather than separately.

The PD also emphasized the importance of using number lines to represent rational numbers and help students see that numbers exist beyond the whole numbers. The use of the number line is recommendation 2 in the IES Practice Guide—“...Use number lines as a central representational tool in teaching [that fractions are numbers]...” Supporting studies for recommendation 2 that are cited by the Guide include Siegler and Booth (2004), which showed that students’ ability to locate decimals accurately on a number line was positively associated with grades.

In addition, the PD emphasized developing teachers’ ability to identify and address persistent student misconceptions, often by presenting students with problems designed to reveal their thinking. The PD also focused on teachers’ ability to explain rational numbers concepts and procedures, helping teachers know the correct language to use when responding to students or delivering direct instruction. The pedagogical techniques that received the most attention were eliciting and responding to student thinking, using charts to keep track of particular student misconceptions, and using strategies for summarizing (for students) the core mathematical ideas of a lesson.

Although the PD was thus designed to address both rational number topic areas and pedagogical techniques, the focus of the presentation in both years of the study was on SK, and instruction in common knowledge of mathematics (CK) content was mainly implicit.⁵⁴ That is, 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).⁵⁵ Further, the PD did not require teachers to spend time outside the institutes and coaching activities studying rational number content or practicing pedagogical techniques. The only exception to this

⁵³ Appendix C includes a complete list of the topics covered by each provider in each institute and seminar day in each year of the study.

⁵⁴ As explained in Chapter 1, 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.

⁵⁵ The results of the teacher knowledge test used in the evaluation were not shared with the teachers or the providers.

rule was that America's Choice assigned teachers brief readings and/or problem sets to be completed between some of the institute and seminar days. These homework assignments were not collected or graded, but the PD sometimes included 15 minutes for teachers to discuss their homework with one another.

The next two sections provide more detailed descriptions of the design and development of each of the PD components and highlight the specific approaches planned by each PD provider for each year of the study.

Summer Institute and Seminar Series

Each institute and seminar day was divided into segments, and the facilitator guides prepared by each PD provider specified the segment structure as well as the mathematical content, activities, and suggested timing for each segment. Facilitator guides for the first year were refined through a yearlong pilot and review process; the second-year facilitator guides used the same structure as the first-year guides. The study's external advisors reviewed both providers' facilitator guides in both years, focusing on the accuracy, appropriateness, and coherence of the mathematics content presented to teachers.⁵⁶

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.

During each PD segment, teachers were to be provided specific preplanned participant materials from the facilitator's guide, such as mathematics problem sets and worksheets, templates for planning and reflecting on lessons and monitoring student thinking, journals and handouts for teacher reflection, and supplemental readings in rational number content and pedagogy. At the end of each PD segment, facilitators were to summarize the main mathematical ideas covered in that segment. Each day of the PD program was also designed to provide opportunities to link the content of the PD to the seventh-grade mathematics textbooks being used by the teachers participating in the PD.

Although both PD providers incorporated the overall design features described above, the providers' planned PD differed in some specific elements. For example, in America's Choice PD segments, teachers were asked to solve sets of mathematics problems, working individually or in small groups. The problem sets were designed to lay the groundwork for or reinforce definitions of key rational number concepts and to illustrate common student misconceptions. Facilitators were directed to emphasize these same ideas in subsequent structured discussions. In contrast, Pearson Achievement Solutions used a single, more open-ended problem or task to structure each PD segment. Each task was designed to elicit multiple approaches and to fuel extended discussions about the core ideas, common student approaches, and potential misconceptions associated with the task. Following these discussions, teachers collaboratively planned how they would teach a lesson related to the task. Facilitators were told to use their expertise to determine how to structure the

⁵⁶ The review process drew on materials on rational numbers developed by mathematicians Sybilla Beckman, James Milgram, and Hung-Hsi Wu, specifically Beckman (2005), Milgram (2005), and Wu (2002a, 2002b, 2005).

discussions and whether to extend the length of any given PD segment. Greater detail on each PD provider's planned approach to the PD is provided in Appendix C.

Coaching

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 provided through 5 two-day visits to each school in the first year and 4 two-day visits to each school in the second year. In the first year, each two-day coaching visit was scheduled to begin no later than the third school day after one of the five seminar days and was designed to link to the preceding seminar. In the second year, because the PD consisted of three seminars and 4 two-day coaching visits, one coaching visit was unattached to a seminar day.

Each provider prepared a coaching manual or coaching plan that described the intended structure and focus of each day's coaching activities. According to the manual or plan, facilitators were expected to use both individual and group delivery formats and a range of coaching activities, including planning, observing, instructing, and debriefing.

As with the summer institutes and seminars, however, the two providers structured their coaching activities differently. In both years of the PD, America's Choice used a coaching model that was designed to work with whatever topics teachers were teaching according to the district pacing guide. Pearson Achievement Solutions focused each coaching visit on a specific rational number lesson that was planned collaboratively during the seminar days. America's Choice planned to engage in a range of individual and group coaching activities on each of the two-day coaching visits, whereas the Pearson Achievement Solutions coaching plan emphasized coach observation and after-school group debriefing discussions. The specific approaches to coaching planned by each PD provider are described in Appendix C.

Implementation of the PD Program

This section describes the facilitators who delivered the PD and documents the degree to which the PD was delivered as planned.

Professional Development Facilitators

In the first year, 10 facilitators (6 for America's Choice and 4 for Pearson Achievement Solutions) delivered the institutes, seminars, and coaching in the 12 study districts. In the second year, some of these facilitators left their organizations and the remaining facilitators were redeployed as necessary to deliver the PD in the 6 remaining study districts.

Each year, a designated pair of facilitators led the institutes and seminars for a given district. The pair of facilitators split up to conduct the coaching, with one facilitator assigned to half the treatment schools in the district and the other assigned to the other half.⁵⁷

Eight of the 10 facilitators had an undergraduate degree or concentration in mathematics or mathematics education, and all 10 were certified to teach secondary mathematics. Eight facilitators held master's degrees, and five of the eight master's degrees were in mathematics education. The

⁵⁷ Whenever facilitators remained with the same district in the second year, they also remained with the same schools for coaching as they had worked with in the first year.

remaining three master's degrees were in technology, administration, and curriculum. The facilitators had 7 to 39 years of experience teaching mathematics and 2 to 16 years of experience providing professional development.

At the conclusion of the first implementation year, the facilitators took the same teacher knowledge test that was administered to the teachers participating in the study. On average, 92.7 percent of the facilitators correctly answered teacher knowledge test items of average difficulty for the test instrument, compared with 45.7 percent of teachers in the first-year treatment group and 50.6 percent of teachers in the first-year control group, as measured at baseline.^{58,59}

Before the first year of implementation, the facilitators working with each provider participated in weeklong summer training programs taught by the providers' lead developers (who had created and compiled the PD program materials). The training sessions offered time for the facilitators to become familiar with the key goals and structure of the professional development, read the facilitator and participant materials, work through the activities and problem sets, and practice delivering segments. The facilitators also participated in summer training prior to the second year of implementation, but the training was less formal and less intensive compared with the summer training that occurred prior to the first year. However, in that second summer the facilitators from both providers contributed to the development and refinement of the second-year PD materials.

Implementation

In the sections that follow, the focus is on the overall implementation of the PD program, and only major differences in implementation by provider are noted. A full presentation of implementation results by PD provider is given in Appendix C.

Institutes and seminars. To measure the degree to which the institutes and seminars were implemented as planned, study staff collected detailed attendance records and observed all 90 days of professional development (i.e., 15 days in each of the 6 two-year districts, including the makeup institutes for new teachers offered at the beginning of the second year). For each observation, they completed a detailed, closed-ended protocol, the results of which could then be compared with the day's prespecified training agenda. The observation protocol (Institute and Seminar Implementation Form) focused on seven dimensions of the PD: the duration of each planned segment; whether each planned segment was covered or skipped; the delivery formats (e.g., individual, small group, whole group, or teacher presentation); the use of participant materials; the extent to which the main ideas of each segment were summarized; the extent to which links were made to the mathematics curriculum used in the district schools; and the level of teacher engagement. To support the implementation form, observers received a coding guide and were trained in its use. The observation protocol measured the degree to which each provider's plan was implemented, but it did not measure the quality of the delivery or the accuracy of the mathematics presented.

Across the six district-training groups in the two-year districts, the number of seventh-grade teacher participants in the institutes and seminars ranged from 3 to 11 and averaged 7 in the first

⁵⁸ The differences are statistically significant. Two-tailed t-tests indicate that both p-values are < 0.01.

⁵⁹ 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 indicates that the relatively small knowledge gain produced by the intervention was not due to a ceiling on the test instrument.

year of the study; the total number of participants ranged from 6 to 16 and averaged 10.⁶⁰ In the second year of the study, the number of seventh-grade teacher participants in the regular institutes and seminars ranged from 3 to 8 and averaged 6; the total number of participants ranged from 4 to 13 and averaged 8.

Table 2-2 summarizes data on the total duration of the institutes and seminars as delivered in the 6 two-year districts, and Table 2-3 provides information on the duration of coverage for specific content areas. Content area coverage was approximated based on the time devoted to each agenda section and the primary focus of each section.

On average, the PD providers delivered 93 percent of the intended institute and seminar hours in each year of the study. This percentage represents 44.8 hours of the intended 48 hours for the first-year summer institute and seminars and 39.0 hours of the intended 42 hours for the second-year summer institute, makeup institute, and seminars.

During the first year, an average of 23.2 hours of the institutes and seminars focused on fractions and decimals, and 21.6 hours focused on percent, ratio, rate, and proportion. In the second year, the averages were 14.4 for fractions and decimals and 24.6 for percent ratio, rate, and proportion, but the between-district variation in content area coverage was greater because of the different designs adopted by America’s Choice and Pearson Achievement Solutions.⁶¹

Table 2-2. Teacher Institutes and Seminars—Percent of Intended Time Implemented and Actual Hours Implemented: Two-Year Districts

	Percent of Intended Hours Implemented	Intended Hours	Mean Actual Hours	S.D.	Minimum	Maximum
First Year						
Total Hours Across Topics	93.2	48.0	44.8	1.90	42.2	47.3
Second Year						
Total Hours Across Topics	92.8	42.0	39.0	2.47	35.4	42.2

Sample Size: N = 6 two-year districts.

SOURCE: 2007–2008 Institute and Seminar Implementation Form; 2008–2009 Institute and Seminar Implementation Form.

NOTE: Second-year estimates include hours for the makeup institutes that were offered to teachers who joined the study after the first-year summer institutes.

⁶⁰ Mathematics teacher leaders, department chairs, and resource teachers were also invited to participate in the institutes and seminars. However, these participants were not part of the impact study, and no data were collected on their baseline characteristics or outcomes.

⁶¹ See Appendix C for the approximate implemented hours of teacher institutes and seminars covering specific content areas by PD provider. As discussed above, America’s Choice designed its five days of institutes and seminars in the second year to allocate more coverage to ratio, rate, and proportion (four days) than to fractions and decimals (one day), whereas Pearson Achievement Solutions designed its second year institutes and seminars to provide equal coverage to fractions and decimals (two and a half days) and ratio, rate, and proportion (two and a half days).

Table 2-3. Teacher Institutes and Seminars—Approximate Hours of Implemented Time Covering Specific Content Areas: Two-Year Districts

Content Area	Mean Hours	S.D.	Minimum	Maximum
First Year				
Fractions, Decimals	23.2	0.87	22.2	24.5
Percent, Ratio, Rate, Proportion	21.6	1.36	20.1	23.5
Second Year				
Fractions, Decimals	14.4	4.46	10.3	19.4
Percent, Ratio, Rate, Proportion	24.6	6.57	15.9	31.6

Sample Size: N = 6 two-year districts.

SOURCE: 2007–2008 Institute and Seminar Implementation Form, 2008–2009 Institute and Seminar Implementation Form.

NOTES: Hours per topic are an approximation based on the primary focus of each agenda section.

Second-year estimates include hours for the makeup institutes that were offered to teachers who joined the study after the first-year summer institutes.

As described earlier in the chapter, each day’s PD was divided into segments. The segment structure varied by provider and by day; overall, 6 to 12 segments were planned per day, with each segment scheduled to last from 5 to 145 minutes. To assess the content coverage, we examined whether the planned segments took place and, if so, whether the time devoted to each segment met or exceeded the planned time.

The results presented in Table 2-4 indicate that across the two PD providers, on average, 1 hour of planned PD segments was shifted to other content or skipped each day, because of either omitted or abbreviated segments. As shown in Appendix C, America’s Choice, which used a prescriptive plan that stressed coverage of all segments, reallocated 0.6 and 0.5 hours of planned segments per day, respectively, in the first and second years of the program. Pearson Achievement Solutions, which planned flexibility that allowed some segments to run long and others to be omitted, reallocated an average of 1.5 hours per day in each year of the program.

Table 2-4. Teacher Institutes and Seminars—Mean Reallocated Hours and Percent of Planned Segments Omitted and Abbreviated: Two-Year Districts

	Mean Hours Reallocated	Percent of Segments Omitted	Percent of Segments Highly Abbreviated (lasted 50 percent or less of intended time)	Percent of Segments Abbreviated (lasted 51–75 percent or less of intended time)
First Year	1.1	2.5	16.5	13.7
Second Year	1.0	5.1	6.4	15.4

Sample Size: N = 393 planned PD segments in the first year (160 for America’s Choice; 233 for Pearson Achievement Solutions); 312 planned PD segments in the second year (159 for America’s Choice; 153 for Pearson Achievement Solutions); 48 institute and seminar days in the first year (24 for America’s Choice; 24 for Pearson Achievement Solutions); 42 institute and seminar days in the second year (21 for America’s Choice; 21 for Pearson Achievement Solutions).

SOURCE: 2007–2008 Institute and Seminar Implementation Form; 2008–2009 Institute and Seminar Implementation Form.

NOTES: The results were calculated across PD days, with each PD day weighted by the number of planned PD segments. Reallocated hours include the intended duration for omitted segments and the difference between the intended and actual duration for abbreviated segments (i.e., segments that did not last for the intended duration). Minutes reallocated from one segment may have been shifted to another segment or skipped and never delivered. Results presented elsewhere indicate that the majority of the reallocated hours were shifted to other segments rather than skipped entirely.

Second-year estimates include hours for the makeup institutes that were offered to teachers who joined the study after the first-year summer institutes.

Each institute and seminar day was designed to include a combination of individual, small-group, and whole-group activities, as well as teacher presentations; use a planned set of materials for each segment; summarize main ideas at the end of each segment; and explicitly link the PD content and the specific seventh-grade mathematics curriculum used in the study schools.

Based on observation records, each of these planned features was implemented during more than 70 percent of the institute and seminar days provided by America’s Choice. (See Appendix C.) More than 70 percent of the institute and seminar days provided by Pearson Achievement Solutions exhibited the planned formats in the first year and the planned explicit links in the second year; however, the remaining features were implemented less often. For example, 80 percent or more of the planned materials were used on 8 percent of the first-year PD days and on 67 percent of the second-year PD days provided by Pearson Achievement Solutions. The variation from plan exhibited by Pearson Achievement Solutions reflects its facilitators’ decisions about the best use of time within particular segments, but also Pearson’s development of more focused sets of planned materials for the second year.

Finally, we examined the overall level of teacher engagement for each day of the institutes and seminars.⁶² On 98 percent of the first-year days and 100 percent of the second-year days, at least 80 percent of the participating teachers were engaged in the PD, as measured by the study observers.

Coaching. To describe how the coaching was implemented, the study collected coach logs on which the coaches reported the duration of each interaction with individual study teachers, as

⁶²The Institute and Seminar Implementation Form included an item on teacher engagement that had five possible responses: 20 percent or less, 40 percent, 60 percent, 80 percent, or 100 percent of participating teachers were actively engaged for the majority of the day. Observers were to record teacher engagement at least four times across the day. Teachers were to be counted as actively engaged if they were watching the facilitator, working problems, or listening to or contributing to the discussion. To be actively engaged, teachers did not need to be enthusiastic, just attentive.

well as emphasis on topics in rational numbers, emphasis on pedagogical topics highlighted by the study’s PD, delivery format, and use of intended activities.

Table 2-5 summarizes the amount of coaching delivered per teacher per visit, based on all visits in which teachers from the second-year impact sample participated or could have participated.⁶³ According to the coach log data, these coaching events delivered an average of 3.9 hours of coaching per two-day coaching visit in the first year (97 percent of the intended 4 hours per visit) and 5.3 hours of coaching per two-day coaching visit in the second year (132 percent of the intended 4 hours per visit). Included in these averages are hours spent coaching one-on-one as well as hours spent coaching in a group format or with pairs of teachers.

Table 2-5. Coaching—Percent of Intended Time Implemented and Mean Actual Hours per Teacher per Visit: Second-Year Teacher Impact Analysis Sample

Coaching Time per Visit	Percent of Intended Hours Implemented	Mean Actual Hours	S.D.
First Year			
Hours Coached per Teacher per Visit	97.1	3.9	2.30
Second Year			
Hours Coached per Teacher per Visit	132.1	5.3	2.27

Sample Size: N = 125 two-day coaching visits available to teachers in the second-year impact sample during the first year; 180 two-day coaching visits available to teachers in the second-year impact sample during the second year.

SOURCE: 2007–2008 Coach Log, 2008–2009 Coach Log.

On average across PD providers, coaches covered topics in rational numbers in 81 percent of the first-year and 93 percent of the second-year coaching visits. (See Appendix C.) Overall, “common student misunderstandings” and “using representations” were the most common pedagogical foci, featured in over 80 percent of coaching visits. The coaching was delivered using a mix of individual and group formats. One-on-one coaching was included in 87 percent of first-year visits and 84 percent of second-year visits. Coaching as part of a group occurred in 69 percent and 86 percent of first-year and second-year visits, respectively. The most common coaching activities used in both first-year and second-year coaching were debriefing after a lesson, planning lessons, and observing teachers’ instruction; each featured in greater than 75 percent of the visits overall. In contrast, modeling activities that involved the coach instructing students while the teacher observed or co-taught were used in 58 percent of the first-year coaching visits and 31 percent of the second-year coaching visits. Modeling and co-teaching were particularly rare activities during Pearson Achievement Solutions’ coaching visits, occurring during 14 percent and 7 percent of its first-year and second-year visits, respectively.

⁶³ The total number of two-day visits included in this analysis was 125 for the first year and 180 for the second year. These numbers represent the number of visits in which teachers in the second-year impact sample could have participated, based on program entry dates. Visits on which a teacher could have participated, but did not, contribute zero hours to the calculation of average hours of coaching per teacher per visit.

Teacher Participation in the PD Program

In the previous sections, we reported on the duration and other features of the PD as delivered. In this section, we focus on the average *dosage of PD received* by the 45 teachers who were teaching regular seventh-grade mathematics classes in the treatment schools in spring 2009.⁶⁴ For the institutes and seminars, the dosage was determined from teacher sign-in sheets; for the coaching, the dosage was determined from the coach logs.

The dosage received by the average treatment teacher in this sample is reported in Table 2-6. (Separate results for each provider are presented in Appendix C.) The average treatment teacher attended 77 hours of study PD, or 66 percent of the total 118 PD hours implemented across the two years. This included 62 percent of the institute hours implemented, 62 percent of the seminar hours implemented, and 73 percent of the coaching hours implemented.

Teacher turnover limited the average dosage of PD participation. Evidence of this can be seen in the percentages of teachers in the second-year impact sample who attended less than 75 percent of the PD in each year of the study. As shown in the far right column of Table 2-6, more than 60 percent of the teachers in the second-year impact sample attended less than 75 percent of the first-year PD. In fact, of the 45 treatment teachers teaching regular seventh-grade mathematics classes in spring 2009, 22 were not present and teaching that course at the beginning of the study and therefore were not eligible to receive the full 118 hours implemented. Among treatment teachers teaching eligible classes in spring 2009, the maximum possible PD dosage based on program entry dates was 87 hours on average. Therefore, the average treatment teacher, having attended 77 hours of study PD, received *89 percent of the maximum possible* PD dosage. Details are shown in Appendix C.

⁶⁴ These 45 teachers, combined with 47 teachers in control schools, are the members of the second-year impact sample, which serves as the basis for the impact findings reported in Chapter 3.

Table 2-6. Percent of Implemented PD Hours Attended by the Average Teacher: Second-Year Teacher Impact Analysis Sample

PD Type (Implemented Hours)	Percent of Implemented PD Hours Attended by the Average Treatment Teacher	Percent of Treatment Teachers Attending		
		100% or More of PD ^a	75–99% of PD	Less Than 75% of PD
First Year				
All PD (67 hours)	52.7	19.6 ^c	20.0 ^c	60.3 ^c
Institute (17 hours) ^b	62.1	39.0 ^c	6.4 ^c	54.5 ^c
Seminars (27 hours)	49.5	39.2	6.1	54.7
Coaching (23 hours)	51.3	25.8	19.9	54.3
Second Year				
All PD (51 hours)	83.5	20.6	52.0	27.4
Institute (12 hours)	63.1	55.1 ^c	6.9 ^c	37.9 ^c
Seminars (17 hours)	84.3	60.7 ^c	12.2 ^c	27.0 ^c
Coaching (22 hours)	94.2	41.6 ^c	47.8 ^c	10.5 ^c
Total (First and Second Year)				
All PD (118 hours)	65.9	15.3 ^c	26.1 ^c	58.7 ^c
Institute (29 hours)	62.1	24.6 ^c	28.6 ^c	46.9 ^c
Seminars (44 hours)	62.4	27.8	15.5	56.7
Coaching (45 hours)	73.1	27.6	22.3	50.1

Sample Size: N = 20 schools; 45 teachers.

SOURCE: 2007–2008 Participation Form; 2008–2009 Participation Form; 2007–2008 Institute and Seminar Implementation Form; 2008–2009 Institute and Seminar Implementation Form; 2007–2008 Coach Log; 2008–2009 Coach Log.

NOTES: For each district, the mean total number of hours that program teachers were coached was used in the denominator when calculating the percent of implemented hours of PD attended by treatment teachers.

The row headings contain, in parentheses, the weighted average actual number of hours implemented for each type of PD across the districts. Districts are weighted by the numbers of treatment schools.

^a Because the calculations for coaching and all PD use the average total coaching hours implemented in the denominator, the percentage of PD attended may exceed 100 percent.

^b Teachers who did not participate in the first-year summer institute because of absence, refusal, or entry into the program subsequent to delivery of the institute were provided a two-day makeup institute in the second year of the program. Because this makeup institute addressed content covered during the first-year institute, teachers' participation in the makeup institute is treated as participation in first-year PD.

^c Numbers do not sum to 100 percent owing to rounding.

Comparison of the PD Experienced by Treatment and Control Groups

In addition to the PD program provided by the study to teachers in treatment schools, teachers in both the treatment and control groups could have participated in other PD provided in their district or elsewhere. Several factors could reduce the treatment-control contrast, including the possibility that teachers in the treatment group could have attended fewer nonstudy PD opportunities than control group teachers. Teacher turnover could also have limited the treatment-control contrast. As noted in the previous section, 22 of the 45 treatment teachers teaching eligible classes the end of the PD program were not present and teaching that course at the beginning of the study, meaning that they missed varying amounts of the institutes, seminars, and coaching provided.

To assess whether the PD program as implemented for the study did in fact result in the intended service contrast between treatment and control groups, we relied on data from the teacher surveys administered in fall 2007, spring 2008, fall 2008, and spring 2009. These surveys asked teachers in both groups to report the number of hours they spent in any type of PD related to mathematics or mathematics education. The survey questions were intended to capture the hours that the treatment group teachers spent in the study PD as well as the hours that teachers in either group spent in any other mathematics-related PD.^{65,66}

As intended, and in spite of the influence of teacher turnover on treatment dosage, teachers in treatment schools experienced more hours of mathematics PD than teachers in control schools. Specifically, teachers in treatment schools reported experiencing 63.6 more hours of mathematics PD during the two-year study period than teachers in control schools, including more of the kinds of PD emphasized by the study—institutes and seminars (defined as PD sessions lasting one-half day or longer but excluding courses such as college courses that last for several weeks) and coaching. (See Table 2-7.) Relative to the control group, treatment teachers who were present in spring 2009 reported receiving 37.7 more hours of total mathematics PD during the first year, when the full intended first-year treatment dosage was 68 hours. Also relative to the control teachers, treatment teachers reported receiving 25.8 more hours of total mathematics PD during the second year, when the full intended second-year treatment dosage was 46 hours. (All reported differences were statistically significant.)

⁶⁵ For purposes of the service contrast, we wanted to compare the total amount of PD experienced by all treatment and control teachers in the second-year impact sample over the course of the two years when the study was ongoing (summer 2007 through spring 2009). For teachers who entered the study late, we were able to reconstruct their PD during the period before they entered the study on the basis of responses they gave on their baseline teacher surveys.

⁶⁶ To estimate the program effect on hours of mathematics PD received, we formulated a two-level model paralleling the models used for the impact analyses of teacher knowledge and instructional practice, as described in Chapter 1. However, in estimating the effect on hours of mathematics PD, we included only covariates for treatment group by district and an indicator for block.

Table 2-7. Treatment and Control Group Contrast in Hours of Mathematics-Related PD: Second-Year Teacher Impact Analysis Sample

Type of Mathematics-Related PD	Treatment Group Weighted	Control Group Weighted	Estimated Difference	Standard Error of the Estimated Difference	Estimated Difference Effect Size	P-value
First Year						
Summer 2007						
Institutes or Seminars ^a (hours)	12.9	2.5	10.4*	3.57	1.25	0.01
2007–2008 School Year						
Institutes or Seminars ^a (hours)	23.9	4.8	19.2*	5.96	1.79	< 0.01
Coaching (hours)	10.6	4.7	5.9	3.51	0.62	0.11
Other PD (hours)	4.9	3.1	1.8	2.87	0.15	0.53
Summer 2007, 2007–2008 School Year						
TOTAL PD (hours)	52.3	14.5	37.7*	9.42	1.45	< 0.01
Second Year						
Summer 2008						
Institutes or Seminars ^a (hours)	12.8	9.4	3.4	5.83	0.15	0.57
2008–2009 School Year						
Institutes or Seminars ^a (hours)	24.8	7.0	17.8*	4.17	1.36	< 0.01
Coaching (hours)	14.3	4.7	9.6*	2.28	0.88	< 0.01
Other PD (hours)	4.9	9.4	–4.5	3.51	–0.28	0.21
Summer 2008, 2008–2009 School Year						
TOTAL PD (hours)	56.9	31.0	25.8*	8.95	0.68	0.01
Summer 2007, 2007–2008 School Year and Summer 2008, 2008–2009 School Year						
TOTAL PD (hours)	109.2	45.6	63.6*	15.40	1.11	< 0.01

Sample Size: N = 38 schools (20 treatment, 18 control); 83 teachers (40 treatment, 43 control).

SOURCE: Fall 2007 Teacher Survey; Spring 2008 Teacher Survey; Fall 2008 New Teacher Survey; Spring 2009 Teacher Survey.

NOTES: ^a Institutes or seminars are defined as PD sessions lasting one-half day or longer but excluding courses such as college courses that last for several weeks.

The analyses are based on a two-level model controlling for random assignment block.

Effect sizes were calculated using the control group standard deviation. The control group standard deviation was 8.3 for summer 2007 institutes or seminars; 10.7 for 2007–2008 institutes or seminars; 9.5 for coaching; 12.0 for other PD; 26.1 for the total PD during summer 2007 and school year 2007–2008; 13.1 for 2008–2009 institutes or seminars; 10.9 for coaching; 16.2 for other PD; 38.1 for the total PD during summer 2008 and school year 2008–2009; and 57.4 for the total PD during summer 2007, 2007–2008 school year, summer 2008, and 2008–2009 school year.

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

We also examined the service contrast between treatment and control teachers in terms of the *features* of the mathematics related PD they attended. Eleven different features were analyzed, including three related to the content emphasis of the PD, two related to the pedagogical emphasis, two related to structural features of the PD (active participation and collective participation), two

related to the perceived relevance and clarity of purpose of the PD, and two having to do with features of the coaching.

For the first year, when many treatment teachers in the final impact analysis sample were not yet participating in program PD (because they were late entrants to the study), we observed few treatment and control group contrasts in terms of the features of the mathematics-related PD that they attended, and not all contrasts favored the treatment group.

In the second year, when teacher turnover was more limited, the mathematics-related institutes and seminars received by the treatment teachers more often emphasized topics in rational numbers (summer 2008 and 2008–2009 school year); more often emphasized pedagogical topics (2008–2009 school year only); and more often involved active participation (summer 2008 and 2008–2009 school year).⁶⁷ Compared with control group teachers, treatment group teachers received mathematics-related coaching in the second year that more often used elements of the PD treatment’s coaching cycle (i.e., plan, observe, and debrief), but they did not receive coaching that more often involved the teacher observing coaches and other teachers. (See Appendix C, Table C-10 for details of service contrasts based on features of the mathematics-related PD. All reported differences are statistically significant.)

Summary

In summary, the PD program focused on rational number content and consisted of summer institute days, seminar days scheduled during the school year, and in-school coaching delivered over two years. Results indicate that the PD program was implemented as intended. Teacher turnover in the studied grade over the course of two years limited the maximum possible PD dosage and the magnitude of the treatment-control group service contrast. (Twenty-two of the 45 treatment teachers teaching eligible classes at the end of the two-year PD program were not present at the beginning of the PD.) Nevertheless, the average treatment teacher included in the second-year impact sample received 66 percent of the full intended dosage and 89 percent of the maximum possible dosage given their date of entry into the study. Also over the two years of the study, there was a cumulative service contrast of 63.6 hours of PD between the treatment and control conditions.

⁶⁷ For example, on a four-point Likert scale in which “1” equaled “never” and “4” equaled “often,” treatment group teachers had a mean score of 3.10 for active participation during the 2008-2009 school year, and control group teachers had a mean score of 2.38.

CHAPTER 3

IMPACT OF THE PD PROGRAM

The primary focus of this chapter is an examination of the impact of the overall (two-year) professional development (PD) program on the two types of outcomes that were the focus of the second year of the study: teacher knowledge and student mathematics achievement. Because our report of first-year findings was based on all 12 districts in the first-year sample (Garet et al. 2008), whereas the second-year impact analyses are necessarily restricted to the 6 districts that were included in the second year of the study (the two-year districts), we also present the separate impact findings for the 6 two-year districts at the end of the first year to provide a better context for understanding the second-year results.⁶⁸ Before turning to the impact results, we first review the equivalence of the treatment and control groups in the second-year impact sample.

Equivalence of Treatment and Control Groups

As explained in Chapter 1, the study randomly assigned schools to treatment and control groups, and all impact estimates are based on an intent-to-treat analysis that includes all eligible seventh-grade teachers (and their eligible seventh-grade students), who were present in the sample schools at the time of outcome data collection. Any systematic change in the study sample over the course of the two years of the study that is correlated with treatment status could lead to differences between the treatment and control groups in the impact analyses.

To address this concern, we conducted an analysis of the equivalence of the treatment and control group participants included in the second-year impact analysis samples. For teachers, equivalence is based on selected characteristics of the teachers at the point of entry into the study as well as on the percentage of teachers in each group who entered the study in fall 2007. For students, equivalence is based on selected characteristics as of fall 2008.

The results of these tests, which are reported in Tables 3-1 and 3-2, indicate that there were no statistically significant differences on any of the measured characteristics between the treatment and control groups for either the teachers or the students in the second-year impact analysis sample. In addition to testing for differences in each separate variable, we conducted a chi-square test for all teacher-level variables and another chi-square test for all student-level variables. These tests indicate that there was no overall difference in the characteristics of the treatment and control group teachers or students included in the impact analysis sample.⁶⁹

We also conducted equivalence tests to compare the students in the second-year impact sample with students on the spring 2009 class rosters who were not in the impact sample, but for whom we had district data on student characteristics. For these comparisons, which are shown in

⁶⁸The districts included in the second-year analysis were not chosen randomly. See Chapter 1 for more information about the selection process.

⁶⁹In testing for differences in individual teacher or student characteristics, we conducted many hypothesis tests. Conducting multiple tests increases the probability of concluding that a particular difference is statistically significant when, in fact, the true difference is zero. In particular, we would expect to see a “false positive” for every 20 hypotheses tested. For this reason, we conducted an overall likelihood ratio chi-square test to test for a systematic, or overall, difference between the characteristics of the treatment and control groups. (The test was based on a logit regression, predicting treatment status based on the measured variables.) The p-values for the chi-square are 0.98 for the teachers and 0.72 for students.

Appendix A, Table A-4, there were three significant differences out of eleven characteristics compared. Students in the impact sample were younger than students not in the impact sample, they were less likely to be classified as receiving special education services, and they had higher mathematics scores on the sixth-grade state accountability test. (All p-values < 0.01.)

Finally, we also conducted equivalence tests of the *first-year impact analysis samples* for the 6 two-year districts. There were no statistically significant differences between the treatment and control groups for teachers; for students, there were fewer Hispanic students in the treatment group (p-value = 0.03). The detailed results can be found in Appendix D, Tables D-1 and D-2.

Table 3-1. Teacher Characteristics, by Treatment Status: Second-Year Teacher Impact Analysis Sample

Teacher Characteristics	Treatment Group	Control Group	Estimated Difference	P-value for Estimated Difference
Baseline Teacher Knowledge ^a				
Total Score (logits)	0.32	0.58	-0.26	0.27
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>57.8</i>	<i>64.1</i>	<i>-6.3</i>	
Common Knowledge of Mathematics (CK) Score (logits)	0.60	0.90	-0.30	0.38
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>67.4</i>	<i>73.5</i>	<i>-6.2</i>	
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	-0.01	0.35	-0.37	0.08
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>46.5</i>	<i>55.7</i>	<i>-9.2</i>	
Years of Teaching Experience (percent)				
3 years or fewer	24.7	28.8	-4.1	0.72
4–10 years	40.6	49.1	-8.5	0.48
More than 10 years	34.7	22.3	12.4	0.29
Years of Teaching Experience in Middle School Mathematics	6.6	6.3	0.3	0.85
Educational Level: M.A. and Above (percent)	53.0	49.0	4.0	0.75
Mathematics Major (percent)	24.4	14.8	9.6	0.25
Number of Postsecondary Mathematics Courses Taken	7.4	7.8	-0.4	0.51
Number of Postsecondary Mathematics Education Courses Taken	2.0	2.3	-0.3	0.25
Teachers Who Entered the Study in Fall 2007 (percent) ^b	49.8	64.2	-14.4	0.27

Sample Size: N = 90 teachers (44 treatment, 46 control).

SOURCE: Fall 2007 Teacher Knowledge Test; Fall 2008 Teacher Knowledge Test; Teacher Survey.

NOTES: ^a Sample Size: N = 83 teachers (39 treatment, 44 control).

^b Sample Size: N = 92 teachers (45 treatment, 47 control).

Baseline teacher knowledge was assessed at the beginning of the teacher's first year in the study. For treatment group teachers, teacher knowledge was assessed prior to the summer PD (or prior to the teacher's first seminar day if the teacher missed the summer PD but entered the study within the first 10 weeks of the school year). For control group teachers, teacher knowledge was assessed within the first 10 weeks of the school year. Teachers who entered the study later in the school year were not tested for baseline knowledge.

The values for the percentage correctly answering items of average difficulty for the test instrument correspond to the estimated treatment and control group means, scaled in logits. Educational experience items reflect the teacher's circumstances as of the teacher's entry into the study; teaching experience was calculated as the number of years of experience at the start of the 2008–2009 school year.

Values in the columns represent unadjusted means for the groups. Percentage values for characteristics with multiple categories may not sum to 100 owing to rounding.

The analyses are based on a two-level model controlling for random assignment block.

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

Table 3-2. Student Characteristics, by Treatment Status: Second-Year Student Impact Analysis Sample

Student Characteristics	Treatment Group	Control Group	Estimated Difference	P-value for Estimated Difference
Age (year) ^a	12.57	12.61	-0.04	0.24
Students Eligible for Free or Reduced-Price Lunch (percent)	70.97	76.44	-5.47	0.21
Race/Ethnicity (percent)				
White, Non-Hispanic	35.69	30.77	4.92	0.29
Black, Non-Hispanic	34.02	37.61	-3.59	0.46
Hispanic	24.95	28.53	-3.58	0.37
Asian/Pacific Islander	3.00	1.83	1.16	0.21
Other	2.35	1.19	1.15	0.31
Male (percent)	51.28	48.68	2.60	0.31
English as a Second Language (percent)	14.04	16.65	-2.61	0.41
Special Education Status (percent)	9.51	10.33	-0.81	0.67
Sixth-Grade Mathematics Scores on State Accountability Assessment (standardized)	0.25	0.15	0.10	0.28
Fall 2007 Student Mathematics Achievement				
NWEA Total Score (scale score)	215.88	214.94	0.94	0.40
<i>Corresponding Percentile Rank</i>	22	20		
Fractions and Decimals Score (scale score)	215.45	214.30	1.14	0.33
Ratio and Proportion Score (scale score)	216.38	215.55	0.83	0.47

Sample Size: N = 2,132 students (1,083 treatment, 1,049 control).

SOURCE: Fall 2008 NWEA Rational Number Test; Study District Records.

NOTES: ^a Age was calculated as the age (in years) of a student as of September 1, 2008.

Values in the columns represent unadjusted means for the groups. Percentage values for characteristics with multiple categories may not sum to 100 owing to rounding.

The analyses are based on a three-level model controlling for random assignment block.

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

Impact on Teacher Knowledge

At the end of both the first and second years of the PD program, teacher knowledge was measured and compared for all treatment and control group teachers using a test constructed specifically to measure teachers' knowledge of rational numbers.⁷⁰ The Interim Report presented the first-year impacts on teacher knowledge across all 12 districts that participated in the first year of the PD. The top panel of Table 3-3 displays the impacts on teacher knowledge within the 6 two-year

⁷⁰ For a detailed description of the teacher knowledge test used to measure impact, see Appendix B. Alternate forms of this test were also used to measure teachers' knowledge at baseline.

districts at the end of the first year.⁷¹ As with the full sample of 12 districts, there were no statistically significant impacts found at the end of the first year for these 6 districts.⁷²

The bottom panel of Table 3-3 presents the impacts of the PD program on teacher knowledge after two years of implementation.⁷³ The second-year analysis is based on the 89 seventh-grade mathematics teachers who were teaching at study schools at the end of the second year and who completed the spring 2009 teacher knowledge test. Of those, 48 teachers—including 21 teachers in the treatment group and 27 teachers in the control group—were enrolled in the study since the beginning of the first year. The rest of the teachers in this sample entered sometime during the first or second year.⁷⁴

⁷¹ The regression model used in this table is described by Equation B-1 in Appendix B.

⁷² See Appendix D, Table D-5, for the first-year impacts of the program on the six districts not included in the second-year analysis (one-year districts). The characteristics by treatment status for this sample of teachers are reported in Table D-3.

⁷³ The covariates used in the model are listed in Appendix B. As a robustness check, Table D-7 in Appendix D presents the impacts of the PD program on teacher knowledge using a similar model that does not include the covariates. There were no statistically significant findings using this approach.

⁷⁴ See Exhibit 1-2 in Chapter 1 for more information regarding teacher entry and exit during the course of the study.

Table 3-3. Impact of the PD Program on Teacher Knowledge at the End of the First and Second Years: First- and Second-Year Teacher Impact Analysis Samples—Two-Year Districts

Outcome Measure	Treatment Group	Control Group	Estimated Impact	Standard Error of the Estimated Impact	Estimated Impact Effect Size	P-value for the Estimated Impact
End of the First Year						
Total Score (logits)	0.41	0.43	-0.02	0.19	-0.02	0.92
<i>Percent correctly answering items of average difficulty for the test instrument</i>	60.0	60.5	-0.4			
Common Knowledge of Mathematics (CK) Score (logits)	0.55	0.62	-0.08	0.31	-0.06	0.80
<i>Percent correctly answering items of average difficulty for the test instrument</i>	66.2	67.9	-1.7			
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.42	0.49	-0.08	0.26	-0.07	0.77
<i>Percent correctly answering items of average difficulty for the test instrument</i>	57.3	59.1	-1.9			
Sample Size: N = 38 schools (20 treatment, 18 control); 90 teachers (41 treatment, 49 control).						
End of the Second Year						
Total Score (logits)	1.13	1.08	0.05	0.20	0.05	0.79
<i>Percent correctly answering items of average difficulty for the test instrument</i>	75.7	74.7	1.0			
CK Score (logits)	1.26	1.54	-0.29	0.24	-0.21	0.25
<i>Percent correctly answering items of average difficulty for the test instrument</i>	79.9	84.1	-4.2			
SK Score (logits)	0.78	0.37	0.41	0.23	0.36	0.09
<i>Percent correctly answering items of average difficulty for the test instrument</i>	65.8	56.2	9.6			
Sample Size: N = 38 schools (20 treatment, 18 control); 89 teachers (43 treatment, 46 control).						

SOURCE: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test.

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 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 values for the percentage correctly answering items of average difficulty for the test instrument correspond to the estimated treatment and control group means, scaled in logits.

Effect sizes were calculated using the control group standard deviation for the first-year teacher impact analysis sample. The control group standard deviation was 0.97 for the *Total score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*.

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

As shown in the table, the PD program did not have a statistically significant impact on teachers' teacher knowledge total score at the end of the second year. 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 of teachers in the control group. The impacts of the PD program on teachers' common knowledge of mathematics (CK) and specialized knowledge of mathematics for teaching (SK) subscores were also not statistically significant.

These impact results reflect the average impact of the PD program across the 6 two-year districts, with each district weighted by the number of treatment schools in the study sample.⁷⁵ These overall impacts might mask differences in impact across the six districts. Therefore, we examined the site-by-site variation in impact at the end of the second year and found no statistically significant variation across the six districts for any of the teacher knowledge measures.⁷⁶

Impact on Student Achievement

The primary student achievement outcome is the total score on a customized test of rational number mathematics developed by the Northwest Evaluation Association (NWEA). Because the PD focused on topics in rational numbers, this measure is a key indicator of the impact of the PD program on student achievement. In addition to the NWEA *Total score*, we measured two subscales for specific topics: (1) *Fractions and Decimals Score* and (2) *Ratio and Proportion Score*.⁷⁷

Table 3-4 displays the impacts of the PD program on student mathematics achievement at the end of the first and second years of implementation for the 6 two-year districts.⁷⁸ The student achievement impact analysis for the first year of the study was based on a sample of the seventh-grade students that the study teachers taught during the 2007–2008 school year, and the impact analysis for the second year was based on a sample of the students these teachers taught during the 2008–2009 school year. More specifically, a randomly chosen sample of the seventh-grade students enrolled in the participating schools at the end of each study year was recruited to complete the customized NWEA test and was included in the student impact analysis sample for that study year.⁷⁹

The top panel of Table 3-4 presents the student achievement impacts at the end of the first year for the 6 districts that participated in both years of the PD. As with the full sample of 12 districts, there were no statistically significant impacts on any of the student achievement measures.⁸⁰

The bottom panel of Table 3-4 displays the second-year impact of the PD program on the students' NWEA total score and the two subscale scores.⁸¹ In the second year of the study, 87

⁷⁵ Table D-11 in Appendix D includes the unadjusted means and standard deviation for all teacher and student outcomes.

⁷⁶ Figure D-1 in Appendix D graphically illustrates the second-year impact estimates and 95 percent confidence intervals for teacher knowledge *Total Score* by site. This figure provides a visual representation of the variability in impacts as well as the uncertainty in the estimate for each district. Statistical tests suggested no statistically significant variations across sites. The p-values for these tests are 0.28 for the *Total Score*, 0.31 for the *CK Score*, and 0.29 for the *SK Score*.

⁷⁷ See Appendix B for a more detailed description of the NWEA rational number test. This computer adaptive test was also used to measure student achievement at baseline. Software prevented a student from seeing the same item more than once on successive administrations.

⁷⁸ The regression model used in this table is described by Equation B-2 in Appendix B.

⁷⁹ For more information on the samples of students included in the impact analyses, see Chapter 1 and Appendix A.

⁸⁰ See Appendix D, Table D-6 for the first-year impacts on student achievement for those districts that did not participate in the second year of implementation (one-year districts). The student characteristics by treatment status for this sample of students are reported in Table D-4.

percent of the 2,132 students in the student impact analysis sample were also at the study schools when the fall class rosters were collected, indicating that they were in the school for most or all of the school year.

Table 3-4. Impact of the PD Program on Student Mathematics Achievement at the End of the First and Second Years: First- and Second-Year Student Impact Analysis Samples—Two-Year Districts

Outcome Measure	Treatment Group	Control Group	Estimated Impact	Standard Error of the Estimated Impact	Estimated Impact Effect Size	P-value for the Estimated Impact
End of the First Year						
NWEA Total Score (Scale Score)	218.79	218.93	-0.13	0.95	-0.01	0.89
<i>Corresponding Percentile Rank</i>	21	21				
Fractions and Decimals Score (scale score)	217.50	217.38	0.12	0.96	0.01	0.90
Ratio and Proportion Score (scale score)	220.12	220.51	-0.39	1.01	-0.03	0.70
Sample Size: N= 39 schools (20 treatment, 19 control); 2,203 students (1,094 treatment, 1,109 control).						
End of the Second Year						
NWEA Total Score (Scale Score)	219.90	219.97	-0.07	0.99	-0.01	0.94
<i>Corresponding Percentile Rank</i>	22	22				
Fractions and Decimals Score (scale score)	218.15	218.36	-0.21	1.04	-0.01	0.84
Ratio and Proportion Score (scale score)	221.71	221.57	0.14	1.03	0.01	0.89
Sample Size: N= 39 schools (20 treatment, 19 control); 2,132 students (1,083 treatment, 1,049 control).						

SOURCE: Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test.

NOTES: The impact analyses for student mathematics achievement were conducted using scale scores. The estimated impacts are based on a three-level model controlling for random assignment block and student-level covariates.

The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the treatment group as the basis for the adjustment.

The values for the corresponding percentile ranks are derived from the treatment and control group means in scale scores.

Effect sizes were calculated using the control group standard deviation from the first-year student impact analysis sample. The control group standard deviation was 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for *Ratio and Proportion Score*.

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

As shown in Table 3-4, there were no statistically significant impacts on any of the measures of student achievement during the second year of the study. Based on the spring norming sample for the NWEA rational number test, the average spring *Total Score* for both treatment and control group students in the second year of the study corresponded to the 22nd percentile. This indicates that the students in the study sample on average performed at the low end of the student achievement distribution.⁸²

⁸¹ Tables D-8 through D-10 in Appendix D present several robustness checks of this impact analysis. Table D-8 uses the same model but does not include any of the covariates (the covariates used in the original analysis are listed in Appendix B). Table D-9 includes the teacher-level covariates used in the teacher knowledge model (also listed in Appendix B) along with the student- and school-level covariates. Table D-10 displays the impacts using teacher instead of class as the middle level of a multilevel model. No statistically significant differences between treatment and control groups were found in any of these checks.

⁸²See Appendix B for more information on the determination of percentile ranks on the NWEA rational number test.

The impact analyses for students reflect the average impact of the PD program across the 6 two-year districts, with each district weighted by the number of treatment schools in the study sample. These overall impacts might mask differences in impact across the six districts. Therefore, we also examined the site-by-site variation in impact and found no statistically significant variation across the six districts for any of the student achievement measures.⁸³

Summary

In general, no statistically significant impacts were found on any of the teacher or student outcome measures at the end of the first or second year for the subset of districts that participated in two years of PD.

⁸³ Figure D-2 in Appendix D graphically illustrates the second-year impact estimates and 95 percent confidence intervals for student achievement *Total Score* in the study by site. This figure provides a visual representation of the variability in impacts as well as the uncertainty in the estimate for each district. Statistical tests suggested no statistically significant variations across sites. The p-values for these tests are 0.68 for the *Total Score*, 0.59 for the *Fractions and Decimals Score*, and 0.70 for the *Ratio and Proportion Score*.

CHAPTER 4

SUMMARY OF FINDINGS AND EXPLORATORY ANALYSES

In this chapter, we summarize the impact of the professional development (PD) program, based on results presented in the Interim Report and Chapter 3 of the current report. We then consider some exploratory analyses, using non-experimental methods, to suggest areas for potential investigation in future work.

Summary of Impact Findings

In the sections below we summarize the results for each of the three sets of outcomes examined by the study: teacher knowledge, instructional practice, and student achievement.

As indicated in Chapter 1, the first year of the PD program was implemented in 12 districts, and the Interim Report provided estimates of the impact of the initial year of PD in all 12 districts. The second year of PD was provided in 6 of the 12 districts. To make it possible to compare the findings for the first and second years, we summarize overall results for the first year, as well as separate results for the 6 districts that provided PD in the first year only (the one-year districts), and the 6 districts that provided PD in both study years (the two-year districts).

Teacher knowledge. Table 4-1 provides estimates of the impact of the PD program on the total teacher knowledge score and each of the two teacher knowledge subscores—common knowledge of mathematics (CK) and specialized knowledge of mathematics for teaching (SK). The study’s main impact estimates for the outcomes at the end of the first year (presented in the Interim Report) and the cumulative outcomes at the end of the second year (presented in Chapter 3) are highlighted in the table. Other table rows show first-year results separately for the one-year districts and the two-year districts.

The table also presents the results of two exploratory analyses of one-year effects on teacher knowledge. The first of these examines the effect on teacher knowledge during the second year of PD only; that is, it analyzes second-year outcomes, controlling for teacher knowledge at the end of the first year/beginning of the second year. The second exploratory analysis is a “per year” analysis of the effect of the PD on teacher knowledge. For the latter analysis, we pooled data from the first- and second-year impact samples in order to increase the precision of the estimates. 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 (also from the 6 two-year districts). Teachers who were in both impact analysis samples 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. (See Table E-1 and E-2 in Appendix E for full results).⁸⁴

⁸⁴ The pooled analyses for teacher outcomes, and also the pooled analysis for student outcomes, are discussed in more detail in Appendix E. Table E-4 shows data on the baseline equivalence of the three samples that are included in the pooled sample: teachers in the first-year impact sample only, teachers in the second-year impact sample only, and teachers in both impact samples. Table E-5 examines the interaction between sample membership and treatment effect; no significant interactions were found.

A few points are apparent in Table 4-1. As previously reported in the Interim Report, we did not find significant impacts of the PD program on any of the measures of teacher knowledge at the end of the first year for the full set of 12 districts. We did find a significant impact of the PD program on the teacher knowledge total score (effect size=0.38) and the SK score (effect size = 0.50), but not the CK score (effect size = 0.07), for the 6 one-year districts. We found no significant impacts at the end of the first year for the 6 two-year districts (effect sizes = -0.02 for total score, -0.06 for CK, and -0.07 for SK).

Also, as reported in Chapter 3, we found no significant cumulative impact of the PD program on teacher knowledge at the end of the second year (effect sizes = 0.05 for total score, -0.21 for CK, and 0.36 for SK). No significant impact was found either, when the analysis was restricted to the one-year impacts achieved during the second year of the study (effect sizes = 0.10 for total score, -0.17 for CK, and 0.34 for SK).

Table 4-1. Impact on Teacher Knowledge (Effect Size) as Estimated for Different Time Periods and Samples (Main Impact Estimates Highlighted)

Time Period and Sample	Estimated Impact Effect Size	P-value	Sample Size
Total Score			
End-of-First-Year Impact, All 12 Districts	0.19	0.15	189
End-of-First-Year Impact, 6 One-Year Districts	0.38*	0.03	99
End-of-First-Year Impact, 6 Two-Year Districts	-0.02	0.92	90
Cumulative End-of-Second-Year Impact, 6 Two-Year Districts	0.05	0.79	89
End-of-Second-Year One-Year Effect, 6 Two-Year Districts	0.10	0.63	89
Per-Year Effect, Pooled Sample	0.18	0.13	278
Common Knowledge of Mathematics (CK) Score			
End-of-First-Year Impact, All 12 Districts	0.02	0.88	189
End-of-First-Year Impact, 6 One-Year Districts	0.07	0.69	99
End-of-First-Year Impact, 6 Two-Year Districts	-0.06	0.80	90
Cumulative End-of-Second-Year Impact, 6 Two-Year Districts	-0.21	0.25	89
End-of-Second-Year One-Year Effect, 6 Two-Year Districts	-0.17	0.34	89
Per-Year Effect, Pooled Sample	-0.03	0.81	278
Specialized Knowledge of Mathematics for Teaching (SK) Score			
End-of-First-Year Impact, All 12 Districts	0.23	0.14	189
End-of-First-Year Impact, 6 One-Year Districts	0.50*	0.02	99
End-of-First-Year Impact, 6 Two-Year Districts	-0.07	0.77	90
Cumulative End-of-Second-Year Impact, 6 Two-Year Districts	0.36	0.09	89
End-of-Second-Year One-Year Effect, 6 Two-Year Districts	0.34	0.11	89
Per-Year Effect, Pooled Sample	0.28*	0.02	278

SOURCE: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test.

NOTES: The impact analyses for teacher knowledge were conducted using measures scaled in logits. The estimated impacts are based on a three-level model controlling for random assignment block and teacher-level covariates.

All effect sizes were calculated using the control group standard deviation for the first-year teacher impact analysis sample. The control group standard deviation was 0.97 for the *Total Score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*. P-values are based on t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

See Tables E-1 and E-2 in Appendix E for full results of the end-of-second-year, one-year impact and the per-year effect analyses. The pooled sample 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.

Before proceeding to discuss the results of the per-year analyses using the pooled sample, (which is the final set of findings shown in Table 4-1), it is worth noting several features of the study design and of the sample as it was realized. First, although schools were randomly assigned to treatment and control groups at the start of the first year, the observed treatment-control difference in baseline teacher knowledge total score, although not statistically significant, was negative in all samples analyzed—with differences ranging from -0.12 to -0.28 logits. (See Table E-6 in Appendix E, which shows the baseline differences between treatment and control teachers in all the samples

discussed.⁸⁵) To adjust for these treatment-control differences, and to improve the precision of the estimates, we used baseline knowledge as a covariate in all the analyses of impact.

A second consideration in interpreting our results is teacher mobility between the first and second years of the study. Although the second year of the study was intended as a test of the impact of two years of the PD program, only about 60 percent of the teachers in the second-year impact sample were present in the study from the start and thus had the opportunity to receive the full intended dosage of PD.

Finally, the realized teacher-level effects, although not statistically significant, were positive for teacher knowledge total score and SK in all but one of the samples examined, but also small relative to the planned minimum detectable effect size (MDE), particularly the MDE for the second year.⁸⁶

As shown in Table 4-1, the estimated effects on the teacher knowledge total score and CK for the pooled sample were not statistically significant. However, the estimated average effect of each year of the PD program on SK using the pooled sample was statistically significant (effect size = 0.28, $p = 0.02$), and the effect size falls between the estimate for the end-of-first-year impact (effect size = 0.23, not statistically significant) and the estimate for the cumulative end-of-second-year impact (effect size = 0.36, also not statistically significant). Because the sample for this analysis is larger than the samples used for the first-year or second-year impact estimates presented earlier, the precision is somewhat improved.

Instructional practice. Table 4-2 summarizes the first-year results for the impact of the PD program on instructional practice. The results are limited to the first year because instructional practice was not measured in the second year of the study.

⁸⁵ Most of the treatment-control differences in baseline CK and SK were also non-significant and negative. However, in the pooled sample, the difference for CK was negative and significant (difference = -0.43 , $p = 0.03$) and in the first-year impact sample for the one-year districts the difference for SK, though not significant, was positive (difference = 0.02 , $p = 0.94$).

⁸⁶ To put these results in context, the average gain in knowledge for the 87 treatment group teachers in the 12-district sample present in both the fall and spring of the first year was 0.39 logits for the total score, 0.35 for CK, and 0.45 for SK. For the 85 control teachers, the average gain was 0.07 logits for the total score, 0.04 for CK, and 0.10 for SK. The average gain for the 41 treatment teachers present in both the fall and spring of the second year was 0.69 for the total score, 0.54 for CK, and 0.62 for SK. For the 45 control teachers, it was 0.41, 0.45, and 0.02.

Table 4-2. Impact on Instructional Practice (Effect Size), as Estimated for Different Time Periods and Samples (Main Impact Estimates Highlighted)

Time Period and Sample	Estimated Impact Effect Size	P-value	Sample Size
Teacher Elicits Student Thinking			
End-of-First-Year Impact, All 12 Districts	0.48*	<0.01	179
End-of-First-Year Impact, 6 One-Year Districts	0.63*	<0.01	91
End-of-First-Year Impact, 6 Two-Year Districts	0.36	0.12	88
Teacher Uses Representations			
End-of-First-Year Impact, All 12 Districts	0.30	0.05†	179
End-of-First-Year Impact, 6 One-Year Districts	0.32	0.15	91
End-of-First-Year Impact, 6 Two-Year Districts	0.21	0.43	88
Teacher Focuses on Mathematical Reasoning			
End-of-First-Year Impact, All 12 Districts	0.19	0.32	179
End-of-First-Year Impact, 6 One-Year Districts	0.05	0.86	91
End-of-First-Year Impact, 6 Two-Year Districts	0.31	0.28	88

SOURCE: 2007–2008 Classroom Observation Protocol.

NOTES: The impact analyses for instructional practice were conducted using measures scaled in log rate per hour. The estimated impacts are based on a two-level model controlling for random assignment block and teacher-level covariates.

Effect sizes were calculated using the control group standard deviation. The control group standard deviations were 0.74 for *Teacher elicits student thinking*, 1.28 for *Teacher uses representations*, and 0.45 for *Teacher focuses on mathematical reasoning*.

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

† P-value = 0.054, which rounds to 0.05 but is not statistically significant at the 0.05 level.

The results in Table 4-2 show that as previously reported in the Interim Report, the PD program had a positive, statistically significant impact (effect size = 0.48, $p < 0.01$) on a measure of the extent to which the teacher elicits student thinking. This impact was significant in the six districts that participated in the first year of the study only (effect size = 0.63, $p < 0.01$), but not in the six districts that continued on to the second year of the study (effect size = 0.36, $p = 0.12$). No significant impacts were found for the other two measures of instructional practice—the *Teacher uses representations* scale and the *Teacher focuses on mathematical reasoning* scale.

Student achievement. Table 4-3 presents the estimated impact of the first year of the PD on student achievement for the total sample, the one-year districts and the two-year districts, the estimated cumulative impact on student achievement after the second year of the PD, and the estimated average effect on student achievement across the two years of the study, using a pooled sample.⁸⁷ There were no statistically significant impacts of the PD program on student achievement in either year of the study or in any of the subsamples. The estimated average effect across the two

⁸⁷ As with the pooled sample of teachers used in the analysis of the effect of the PD on teacher knowledge, a pooled sample of students was also used to increase power. However, the composition of the pooled samples for teachers and students differ in some respects. For the teacher sample, teachers who participated in both years of the study appear in the sample twice—once for the first year and once for the second year. Students, on the other hand, only appear once. This is because the student samples for each year of the study comprised the students in the teachers' current seventh-grade classes and there was no overlap between these samples except to the extent that individual students from the first year of the study might have been required to repeat seventh-grade mathematics and were assigned to a study classroom for the second year.

years using the pooled sample was also not found to be significant. (See Table E-7 in Appendix E for the full results of this analysis.)

Table 4-3. Impact on Student Achievement (Effect Size), as Estimated for Different Time Periods and Samples (Main Impact Estimates Highlighted)

Time Period and Sample	Estimated Impact Effect Size	P-value	Sample Size
Total Score			
End-of-First-Year Impact, All 12 Districts	0.04	0.37	4,528
End-of-First-Year Impact, 6 One-Year Districts	0.08	0.09	2,325
End-of-First-Year Impact, 6 Two-Year Districts	-0.01	0.89	2,203
Cumulative End-of-Second-Year Impact, 6 Two-Year Districts	-0.01	0.94	2,132
Average Effect, Pooled Sample	0.05	0.24	6,660
Fractions and Decimals Score			
End-of-First-Year Impact, All 12 Districts	0.03	0.38	4,528
End-of-First-Year Impact, 6 One-Year Districts	0.07	0.15	2,325
End-of-First-Year Impact, 6 Two-Year Districts	0.01	0.90	2,203
Cumulative End-of-Second-Year Impact, 6 Two-Year Districts	-0.01	0.84	2,132
Average Effect, Pooled Sample	0.04	0.33	6,660
Ratio and Proportion Score			
End-of-First-Year Impact, All 12 Districts	0.03	0.46	4,528
End-of-First-Year Impact, 6 One-Year Districts	0.08	0.11	2,235
End-of-First-Year Impact, 6 Two-Year Districts	-0.03	0.70	2,203
Cumulative End-of-Second-Year Impact, 6 Two-Year Districts	0.01	0.89	2,132
Average Effect, Pooled Sample	0.05	0.23	6,660

SOURCE: Spring 2008 Rational Number Test; Spring 2009 NWEA Rational Number Test.

NOTES: The impact analyses for student mathematics achievement were conducted using scale scores. The estimated impacts are based on a three-level model controlling for random assignment block and student-level covariates.

Effect sizes were calculated using the control group standard deviation from the first-year student impact analysis sample. The control group standard deviations were 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for the *Ratio and Proportion Score*.

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

See Table E-7 in Appendix E for full results of the analyses of the pooled sample.

Results by Provider, Interactions With Baseline Characteristics, and Correlational Results

The Interim Report included a number of exploratory analyses using the first-year impact analysis sample. Because the second-year impact analysis sample was only about half the size of the first-year impact analysis sample, these analyses were not repeated for the second year. Instead, we used the pooled sample described above to further investigate the questions addressed by these first-year analyses. The results are presented below. (The detailed results of the pooled sample analyses are shown in Appendix E.)

Effect of the PD program by provider. As explained in Chapter 2, we had two PD providers: America’s Choice and Pearson Achievement Solutions. Each delivered PD aligned to the same objectives, using the same format and duration. However, in other respects the providers were free to follow their own practices in designing and delivering the PD. Each provider worked with six districts in the first year and three districts in the second year. However, the districts were not randomly assigned to providers.

After the first year of implementation, for America’s Choice, there was no statistically significant impact of the PD on any of the measures of teacher knowledge or student achievement. Statistically significant impacts were observed for two of the three measures of instructional practice—the *Teacher elicits student thinking Scale* (effect size = 0.63, $p = 0.01$) and the *Teacher uses representations Scale* (effect size = 0.60, $p = 0.02$). Using the pooled sample, we found no statistically significant effects of the PD for America’s Choice. (See Tables E-9 and E-10 in Appendix E for detailed results of the pooled sample analysis of America’s Choice.)

After the first year of implementation, for Pearson Achievement Solutions, there was no statistically significant impact of the PD on any of the measures of teacher knowledge, instructional practice, or student achievement. When we used the pooled sample, we also found no statistically significant effects. (See Tables E-11 and E-12 in Appendix E.)

Interaction with baseline characteristics. After the first year, we examined whether the PD program had differential effects based on teachers’ baseline knowledge scores or students’ baseline achievement scores. No statistically significant interactions were observed for any of the first-year outcomes.

The interaction analyses of effects on teacher knowledge and student achievement were repeated using the pooled sample.⁸⁸ To do this we reestimated the models used in the analysis of the effect of the PD, adding the interaction of baseline knowledge/baseline achievement and the treatment indicators. We also looked for quadratic as well as linear relationships between baseline teacher knowledge/student achievement and the effect of the treatment, and we estimated the models using two slightly different methods as a check on robustness. None of the interaction effects were statistically significant. All the interaction analyses using the pooled samples are described in Appendix E, and the detailed results appear in Tables E-13 through E-19, with the MDEs for these analyses reported in Table E-20.

Relationships between teacher knowledge and student achievement. According to the study’s theory of action, participation in the PD is hypothesized to affect student achievement indirectly by improving teacher knowledge and classroom instruction. For the Interim Report, the relationships among these variables were examined by adding the measures of teacher knowledge and instructional practice to the impact model for student achievement in place of the treatment status indicator. These analyses produced no statistically significant associations, although most of the estimated coefficients were positive and consistent in magnitude with the associations reported in the literature.

⁸⁸ In addition to potential interactions with baseline teacher knowledge, we examined potential differential effects on teacher knowledge or student achievement outcomes associated with teachers’ prior experience. See Tables E-14, E-17, and E-18 in Appendix E.

After the second year, we repeated the analyses related to teacher knowledge and student achievement, using the pooled sample. We first added the teacher knowledge total score to the impact model in place of the treatment status indicator. Separate analyses were also conducted including the CK and SK subscale scores for teacher knowledge.⁸⁹

Table 4-4 reports the estimated association between teacher knowledge and student achievement for each of the three teacher knowledge scores and each of the three student outcome measures.⁹⁰ The table provides the estimated coefficients for each of the teacher knowledge scores, as well as the relevant statistical test information, including the standard error for each estimated coefficient and the corresponding p-values for a two-tailed t-test. All results reported in the table are standardized: each coefficient represents the magnitude of the change in achievement (in effect size using student-level standard deviation units) associated with a one standard deviation change in each of the independent variables, controlling for the other independent variables and covariates included in the model.⁹¹

The estimated association between teacher knowledge total score and two of the three student achievement measures is positive and statistically significant. More specifically, the estimated association between the teacher knowledge total score and the student Northwest Evaluation Association (NWEA) total score is 0.05 ($p = 0.02$), and the association between teacher knowledge total score and students' *Fractions and Decimals Score* is also 0.05 ($p < 0.01$), suggesting that students in a classroom taught by a teacher scoring one standard deviation above average on teacher knowledge scored 0.05 standard deviations above average on the NWEA test total score and the *Fractions and Decimals Score*. The association between teacher knowledge total score and the *Ratio and Proportion Score* is not significant (coefficient = 0.04, $p = 0.07$).⁹²

As shown in the second and third columns, neither teachers' CK score nor SK score was significantly associated with any of the three student outcomes. However, as demonstrated by an F-test in the final column, jointly, these two teacher knowledge measures were significantly associated with students' *Fractions and Decimals Score* ($p = 0.02$).⁹³

⁸⁹ In order to base the analysis on the level of teacher knowledge that the students experienced over the course of the year, the teacher knowledge measure used was the average of the fall and spring test scores for each teacher. As noted earlier, each student is in the pooled sample for only one year, and each student is paired with the teacher knowledge value for his/her teacher in that year.

⁹⁰ The regression models used in this table are described by Equations E-7 through E-9 in Appendix E.

⁹¹ The minimum detectable effect size (MDES) for the association between teacher knowledge *Total Score* and student achievement *Total Score* or student *Fractions and Decimals Score* is 0.05, the MDES for the association between teacher knowledge *Total Score* and student *Ratio and Proportion Score* is 0.06, and the MDES for the association between teacher *CK Score* and student achievement *Total Score* or student *Ratio and Proportion Score* is 0.07, while the MDES for the association between teacher *CK Score* and student *Fractions and Decimals Score* is 0.06. The MDES for the association between teacher *SK Score* and student achievement *Total Score* or student *Ratio and Proportion Score* is 0.08, and the MDES for the association between teacher *SK Score* and student *Fractions and Decimals Score* is 0.07.

⁹² The magnitude of the association between teacher knowledge and student achievement is similar in magnitude to correlations that have been reported in the literature. For example, a study by Clotfelter, Ladd, and Vigdor (2006) found that the associations between teacher licensure test scores and fifth-grade students' mathematics achievement were 0.01 to 0.02 standard deviations. Hill, Rowan and Ball (2005) reported that first- and third-grade students gained roughly 0.05 standard deviations on the Terra Nova mathematics tests for every standard deviation difference in teachers' specialized content knowledge. Rockoff, Jacob, Kane, and Staiger (2008) found that a 1 standard deviation increase in teachers' scores on a test of mathematics knowledge for teaching is associated with a statistically significant (p -value = 0.02) increase of about 0.03 standard deviations in students' mathematics achievement.

⁹³ Table E-22 in Appendix E explores the interaction between sample membership and teacher knowledge in the regressions predicting student achievement. None of the joint p-values for the interactions was statistically significant. See Appendix E for more detail on these tests.

Because we do see a correlation between the teacher knowledge total score and student achievement, these findings suggest that programs positively affecting teacher knowledge have the potential to increase student performance. Still, the magnitude of the correlation indicates that an effective program targeting student achievement through teacher knowledge would need to have a substantial impact on teachers.

Table 4-4. Standardized Regression Coefficients for the Relationships Between Teacher Knowledge and Student Achievement, Pooled Sample

Standardized Outcomes	Mediating Variables in Model				
		TK Total Score	CK Score	SK Score	F-Test
NWEA Total Score	Coefficient	0.05*			
	standard error	0.02			
	p-value	0.02			
	Coefficient		0.03	0.03	
	standard error		0.02	0.03	
	p-value		0.26	0.31	0.07
Fractions and Decimals Score	Coefficient	0.05*			
	standard error	0.02			
	p-value	<0.01			
	Coefficient		0.04	0.02	
	standard error		0.02	0.03	
	p-value		0.08	0.42	0.02*
Ratio and Proportion Score	Coefficient	0.04			
	standard error	0.02			
	p-value	0.07			
	Coefficient		0.01	0.03	
	standard error		0.02	0.03	
	p-value		0.58	0.27	0.21

Sample Size = 76 schools (40 treatment, 36 control); 276 teachers (138 treatment, 138 control); 6,352 students (3,240 treatment, 3,112 control).

SOURCES: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test.

NOTES: Coefficients in the table are standardized regression coefficients. The coefficients were estimated on the basis of a three-level model controlling for random assignment blocks and school-, teacher-, and student-level covariates.

P-values for individual coefficients are based on t-tests, and joint tests of sets of coefficients are based on F-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

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. The 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 (*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. While these findings are suggestive, clear causal interpretations are not appropriate given the exploratory nature of the analyses and the fact that the samples used for these analyses differ from the samples specified in the original sample plan.

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.

REFERENCES

- Beckman, S. *Mathematics for Elementary Teachers*. New York: Pearson Addison Wesley, 2005.
- 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.
- Black, A.R., Doolittle, F., Zhu, P., Unterman, R., and Grossman, J.B. *The Evaluation of Enhanced Academic Instruction in After-School Programs: Findings After the First Year of Implementation* (NCEE 2008-4021). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2008.
- 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.
- Clotfelter, C.T., Ladd, H.F., and Vigdor, J.L. *Teacher-Student Matching and the Assessment of Teacher Effectiveness* (Working Paper No. 11936). Cambridge, MA: National Bureau of Economic Research (NBER), 2006.
- 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.
- Hanushek, E.A., Kain, J.F., O'Brien, D.M., and Rivkin, S.G. *The Market for Teacher Quality* (Working Paper No. 11154). Cambridge, MA: National Bureau of Economic Research (NBER), 2005.
- Hill, H.C. "Mathematical Knowledge of Middle School Teachers: Implications for the No Child Left Behind Policy Initiative." *Educational Evaluation and Policy Analysis*, 2007, 29(2): 95–114.
- Hill, H., Rowan, B., and Ball, D. "Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement." *American Educational Research Journal*, 2005, 42(2): 371–406.
- Kane, T., and Staiger, D.O. *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation* (Working Paper 14607). Cambridge, MA: National Bureau of Economic Research (NBER), 2008.

- Kennedy, M. *Form and Substance of Inservice Teacher Education* (Research monograph no. 13). Madison: University of Wisconsin–Madison, National Institute for Science Education, 1998.
- Milgram, R.J. *The Mathematics Pre-Service Teachers Need to Know*. Stanford, CA: Stanford University, Department of Mathematics, 2005.
- 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.
- Newton, K.J. “An Extensive Analysis of Elementary Preservice Teachers’ Knowledge of Fractions.” *American Educational Research Journal*, 2008, 45(4): 1080–1110.
- Northwest Evaluation Association (NWEA). *Technical Manual for Use With Measures of Academic Progress and Achievement Level Tests*. Lake Oswego, OR: NWEA, 2003.
- Northwest Evaluation Association (NWEA). *Reliability and Validity Estimates: NWEA Achievement Level Tests and Measures of Academic Progress*. Lake Oswego, OR: NWEA, 2004.
- Richardson, V., and Placier, P. “Teacher Change.” In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th Ed.). New York: Macmillan, 2001: 905–947.
- Rittle-Johnson, B., Siegler, R. S., and Alibali, M. W. “Developing Conceptual Understanding and Procedural Skill in Mathematics: An Iterative Process.” *Journal of Educational Psychology*, 2001, 93(2): 346–362.
- Rittle-Johnson, B., and Koedinger, K.R. (2002, October). *Comparing Instructional ‘Strategies for Integrating Conceptual and Procedural Knowledge*. Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Athens, GA.
- Rittle-Johnson, B., and Koedinger, K.R. “Iterating Between Lessons on Concepts and Procedures Can Improve Mathematics Knowledge.” *British Journal of Educational Psychology*, 2009, 79: 483–500.
- Rockoff, J.E. “The Impact of Individual Teachers on Student Achievement: Evidence From Panel Data.” *The American Economic Review*, 2004, 94(2): 247–252.
- Rockoff, J.E., Jacob, B., Kane, T., and Staiger, D. *Can You Recognize an Effective Teacher when you Recruit One?* (Working Paper No. 14485). Cambridge, MA: National Bureau of Economic Research (NBER), 2008.
- Sable, J., and Plotts, C. *Documentation to the NCES Common Core of Data Public Elementary/ Secondary School Universe Survey: School Year 2008-2009*. (NCES 2010-350 rev). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2010.
- Siegler, R. S., and Booth, J. L. “Development of numerical estimation in young children.” *Child Development*, 2004, 75(2): 428–444.

- Siegler, R., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., Thompson, L., and Wray, J. *Developing effective fractions instruction for kindergarten through 8th grade: A practice guide* (NCEE #2010-4039). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2010.
- Sireci, S.G., Thissen, D., and Wainer, H. "On the Reliability of Testlet-Based Tests." *Journal of Educational Measurement*, 1991, 28(3): 234–247.
- 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.
- Tirosh, D., Fischbein, E., Graeber, A.O., and Wilson, J.W. *Prospective Elementary Teachers' Conceptions of Rational Numbers*, 1999. Available online at: <http://jwilson.coe.uga.edu/Texts.Folder/tirosh/Pros.El.Tchrs.html>.
- 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.
- Wu, H. *Chapter 1: Whole Numbers*, 2002a. Available online at: <http://math.berkeley.edu/~wu/>.
- Wu, H. *Chapter 2: Fractions*, 2002b. Available online at: <http://math.berkeley.edu/~wu/>.
- Wu, H. *Key Mathematical Ideas in Grades 5–8*, 2005. Available online at: <http://math.berkeley.edu/~wu/>.
- 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.

APPENDIX A
DETAILS OF THE STUDY SAMPLES

APPENDIX A

DETAILS OF THE STUDY SAMPLES

This appendix describes the process used to recruit schools for the first year of the study, compares the characteristics of the school, teacher, and student samples in the one-year and two-year districts, and provides additional detail on the construction of study samples.

Recruitment for the First Year of the Study

The 12 districts were identified and recruited through a multistage process. In the first stage, we used information from the 2003–2004 *Common Core of Data* (CCD) to identify districts throughout the nation that operated four or more schools meeting the study criteria.⁹⁴ To be included, a school had to have at least 150 students in the seventh grade, so that there would likely be more than one teacher assigned to teach seventh-grade mathematics, and a school had to have 33 percent or more of all students eligible for free or reduced-price lunch, so that the sample would be relevant to federal education programs, which tend to target low-income students.

In the second stage, the resulting list of 311 districts was narrowed to 40 districts. Among the 311 districts, initial contact efforts focused on the 167 districts containing 6 or more eligible schools. We identified the three curricula that were most commonly used (*Glencoe McGraw-Hill Mathematics [Glencoe]: Applications and Concepts*, *Prentice Hall Mathematics [PH Mathematics]*, and *Connected Mathematics [CMP]*) and then focused on the districts that had been using one of those three curricula as the core seventh-grade mathematics program in most of their schools during the 2006–2007 year. We further focused on districts that did not provide districtwide professional development (PD) in mathematics instruction of the same type and level of intensity as that being provided by the study.⁹⁵ Contact efforts were subsequently expanded to include those districts with four or five eligible schools, until a sample of 40 eligible districts was identified.

In the third stage, study staff held informational conference calls with officials in the 40 districts identified in stage two and subsequently visited the 21 districts that expressed interest in participating in the study. A second visit was conducted in each district to present information to principals at eligible middle schools. After a final informational meeting in Washington, DC, the study team secured final commitments from district officials and principals in 12 study districts, located in nine states.

⁹⁴ To identify eligible districts within states that did not appear in the 2003–2004 CCD (i.e., New York, Tennessee, and Kentucky), the study team examined district websites and held conversations with consultants.

⁹⁵ Districts that provided professional development in mathematics instruction that targeted teachers of students in grades other than seventh, involved fewer than 10 hours of training, was attended by individual teachers rather than teams of teachers from the same schools, or focused on topics such as classroom management rather than the theory and practices of mathematics instruction were eligible for the study. Districts that assigned mathematics coaches to support the entire teaching staff of one or more schools or to support teachers of students in the seventh grade were eligible for the study, provided that the district's coaching would not create scheduling problems or excess burden for teachers participating in the study.

Comparison of Schools, Teachers, and Students in Two-Year and One-Year Districts

This section compares the schools, teachers, and students in the two-year districts with those in the one-year districts. The comparisons are based on the first-year impact analysis samples.

Table A-1 shows that schools in districts that participated in both years of PD implementation differed in several respects from schools in districts that participated only in the first year. Schools in the two-year districts were more likely than schools in the one-year districts to be located in the Northeast and less likely to be located in the South. They were more likely to be located in large or middle-sized cities and less likely to be located in urban fringe areas and towns. The two-year study schools were also more likely to be Title I eligible, and they enrolled more seventh graders and were more likely to combine middle and elementary grades.

Table A-1. Characteristics of Schools in One-Year and Two-Year Districts

School Characteristics	Two-Year Districts	One-Year Districts	Estimated Difference	P-value for Estimated Difference
Geographic Region (percent of schools)				
Northeast	35.9	0.0	35.9*	<0.01
South	35.9	71.1	-35.2*	<0.01
Midwest	12.8	10.5	2.3	0.76
West	15.4	18.4	-3.0	0.73
Urbanicity (percent of schools)				
Large or Middle-Sized City	87.2	65.8	21.4*	0.03
Urban Fringe and Large Town	7.7	28.9	-21.3*	0.02
Small Town and Rural Area	5.1	5.3	-0.1	0.98
Title I Eligible (percent of schools)	66.7	86.8	-20.2*	0.04
Free or Reduced-Price Lunch (school average percent of students)	66.1	66.7	-0.6	0.90
Race/Ethnicity (school average percent of students)				
White	34.7	32.8	1.9	0.70
Black	34.7	37.8	-3.2	0.56
Hispanic	25.4	24.0	1.4	0.80
Asian	2.6	2.7	-0.1	0.93
Other	1.4	1.1	0.2	0.75
Male (school average percent of students)	51.6	49.8	1.7	0.14
Total School Enrollment	752.6	757.4	-4.8	0.92
Number of Seventh-Grade Students	207.9	257.3	-49.4*	0.02
Number of Full Time Equivalent Teachers (all grades)	48.5	43.2	5.3	0.06
School Type (percent of schools) ^a				
Middle School Only	66.7	97.4	-30.7*	<0.01
Elementary and Middle	33.3	0.0	33.3*	<0.01
Middle and High	0.0	2.6	-2.6	0.31
Elementary and Middle and High	0.0	0.0	0.0	1.00

Sample Size: N = 39 schools in two-year districts; 38 schools in one-year districts.

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

NOTES: ^a 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.

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

Table A-2 compares the teachers in the one-year and two-year districts, based on the teachers who were present at the end of the first year of the study (the first-year impact sample). On average, teachers in the two-year districts scored higher on the baseline teacher knowledge test, were more likely to have attained a master's or higher postsecondary degree prior to entry into the study, and had taken more postsecondary mathematics and mathematics education courses than their counterparts in the one-year districts. However, there were no significant differences in years of teaching experience between teachers in the one-year and two-year districts. The two groups also did not differ in the percentage of teachers who were “stable,” that is, teacher who entered the study in fall 2007 and remained through the end of the first year.

Table A-2. Characteristics of Teachers in One-Year and Two-Year Districts: First-Year Teacher Impact Analysis Sample

Teacher Characteristics	Two-Year Districts	One-Year Districts	Estimated Difference	P-value for Estimated Difference
Baseline Teacher Knowledge ^a				
Total Score (logits)	0.10	-0.25	0.35*	0.02
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>52.4</i>	<i>43.8</i>	<i>8.7</i>	
Common Knowledge of Mathematics (CK) Score (logits)	0.28	-0.14	0.42	0.07
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>60.0</i>	<i>49.6</i>	<i>10.4</i>	
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.04	-0.25	0.29	0.05†
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>47.8</i>	<i>40.7</i>	<i>7.2</i>	
Years of Teaching Experience (percent)				
3 years or fewer	29.9	30.9	-1.0	0.88
4–10 years	35.2	29.2	6.0	0.40
11–20 years	22.1	25.8	-3.7	0.56
More than 20 years	13.2	13.9	-0.7	0.91
Years of Teaching Experience in Middle School Mathematics	6.9	8.0	-1.0	0.43
Educational Level: M.A. and Above (percent)	46.6	28.1	18.5*	0.02
Mathematics Major (percent)	17.1	10.4	6.7	0.16
Number of Postsecondary Mathematics Courses Taken	7.4	5.3	2.1*	<0.01
Number of Postsecondary Mathematics Education Courses Taken	2.1	1.7	0.4*	0.02
Teachers Who Entered the Study in Fall 2007 (percent) ^b	87.2	94.1	-6.9	0.11

Sample Size: N = 188 teachers (91 in two-year districts, 97 in one-year districts).

SOURCE: Fall 2007 Teacher Knowledge Test; Teacher Survey.

NOTES: ^a Sample Size: N = 190 teachers (93 in two-year districts, 97 in one-year districts).

^b Sample Size: N = 195 teachers (96 in two-year districts, 99 in one-year districts).

Baseline teacher knowledge was assessed at the beginning of the teacher's first year in the study. For treatment group teachers, teacher knowledge was assessed prior to the summer PD (or prior to the teacher's first seminar day if the teacher missed the summer PD but entered the study within the first 10 weeks of the school year). For control group teachers, teacher knowledge was assessed within the first 10 weeks of the school year. Teachers who entered the study later in the school year were not tested for baseline knowledge.

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.

Educational experience items reflect the teachers' circumstances as of the teacher's entry into the study; teaching experience was calculated as the number of years of experience at the start of the 2007–2008 school year.

The analyses are based on a two-level model.

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

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

† P-value = 0.0521, which rounds to 0.05 but is not statistically significant at the 0.05 level.

Finally, Table A-3 compares the characteristics of students in the districts that participated in both years of PD implementation and students in the districts that participated only in the first year of the study (based on characteristics of students who were in the first-year impact analysis sample for each district). In the first year of the study, students in the two-year districts were on average about 2.5 months younger and included higher proportions of English as a second language and special education students than their counterparts in the one-year districts. However, students in the two-year districts scored higher than their counterparts in the one-year districts both on district-administered sixth-grade mathematics achievement tests and on the Northwest Evaluation Association (NWEA) rational number test administered in fall 2007 (baseline).

Table A-3. Characteristics of Students in One-Year and Two-Year Districts: First-Year Student Impact Analysis Sample

Student Characteristics	Two-Year Districts	One-Year Districts	Estimated Difference	P-value for Estimated Difference
Age (years) ^a	12.6	12.8	-0.2*	<0.01
Students Eligible for Free or Reduced-Price Lunch (percent)	64.8	70.8	-6.0	0.07
Race/Ethnicity (percent)				
White, non-Hispanic	33.5	28.3	5.2	0.19
Black, non-Hispanic	34.1	41.7	-7.6	0.08
Hispanic	27.1	25.4	1.7	0.60
Asian/Pacific Islander	2.7	2.0	0.8	0.14
Other	2.7	2.5	0.2	0.80
Male (percent)	51.0	49.6	1.4	0.38
English as a Second Language (percent)	17.9	7.9	10.0*	<0.01
Special Education Status (percent)	11.3	7.4	3.8*	<0.01
Sixth-Grade Mathematics Scores on State Accountability Assessment (standardized, in effect size)	0.19	0.05	0.14*	0.02
Baseline Student Mathematics Achievement				
NWEA Total Score (scale score)	215.98	213.57	2.41*	0.01
<i>Corresponding Percentile Rank</i>	22	18		
Fractions and Decimals Score (scale score)	215.28	212.26	3.02*	<0.01
Ratio and Proportion Score (scale score)	216.54	214.72	1.81*	0.04

Sample Size:^a N = 4,528 students (2,203 in two-year districts, 2,325 in one-year districts).

SOURCE: Fall 2007 NWEA Rational Number Test; Study District Records.

NOTES: ^a Age was calculated as the age (in years) of a student as of September 1, 2007.

The analyses are based on a three-level model.

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

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

Student Sample

This section describes the selection of students to take the fall 2008 and spring 2009 NWEA rational number tests as well as the construction of the second-year student impact sample.^{96, 97} The focus of the second year student impact analysis was all students who, in the spring of 2009, were enrolled in the teachers' eligible seventh-grade classes. However, because of logistical and budgetary constraints, it was not possible for the study to administer the computer-based NWEA rational number test to all students in each classroom. Instead, the study team drew random samples of students to take the NWEA test in fall and spring. The random selection procedure, together with student mobility during the school year, caused the student sample at baseline to be somewhat different from the sample measured at the end of the year.

Random Selection of Students to Take the Fall and Spring NWEA Tests

The sampling procedures for fall and spring testing differed slightly.

Fall Sampling Process

In the fall, the study sampled 8 students per class roster.⁹⁸ To select the fall sample, an ordered sample list of 16 students was constructed by making 16 sequential draws from each class roster. These 16 students were assigned line numbers 1 through 16, and the study team began testing from line number 1 and moved down the list until a sample of 8 students had been achieved. A student on the list might not be eligible to be tested for several reasons:

- The student had a disability or was an English language learner (ELL) and was identified for exclusion on the basis of school review.⁹⁹
- The student had withdrawn or was otherwise ineligible (i.e., student was not in any of the eligible classes).
- The student or parent refused testing.
- The student was absent on the day of testing.
- The student was an alternate who was not needed because 8 students with lower line numbers were successfully tested.

The students listed, up to and including the last tested student, minus any excluded, withdrawn, or otherwise ineligible students, constituted the fall "attempted sample." Response rates were calculated as the percentage of the attempted sample tested, and makeup sessions were held for any schools in which the overall response rate (across all class rosters) was less than 80 percent.

⁹⁶ Scores on the spring 2009 NWEA test were the student achievement outcome measure for the second year of the study. Scores on the fall 2008 NWEA test were included as covariates in the impact analyses of student achievement.

⁹⁷ The procedures described here are the same as those applied to the student sample in the first year of study implementation.

⁹⁸ Because the fall class rosters were obtained early in the school year, there was some movement of students between classes between the time of rostering and the time of testing. The samples were based on the class rosters, rather than the classrooms, meaning that a sampled student who had changed from one eligible classroom to another prior to the test date was still eligible for testing and was tested with the rest of the students on the same class roster unless excluded, refused, or absent.

⁹⁹ When school personnel determined that particular students with disabilities or ELL status could not participate meaningfully under the conditions offered by the study, the students were removed from the sample. Approximately 5 percent of students in the fall sample and 3 percent of students in the spring sample were excluded on this basis.

Spring Sampling Process

As shown in Exhibit A-1, by spring 2009, about 16 percent of the students on the rosters of classes participating in the study were “incoming” students who had entered the study after the fall 2008 class rosters were created. The remaining 84 percent of the students on the spring rosters were “continuing” students who had been on one of the fall class rosters (but not necessarily in the same class in which they were enrolled in the spring).

In spring 2009, the study sampled 9 students per classroom using a procedure that took account of the mix of continuing and incoming students in the classrooms. Using the spring class rosters as the basis for sampling, the study team constructed an ordered list of at least 16 students for each participating classroom and tested the first 9 students from the list who were eligible and present for testing.¹⁰⁰ Students who were already known to be ineligible were removed from the class roster before sampling. Each ordered list included a combination of incoming and continuing students, with a portion of the continuing student slots assigned with certainty to students who were in the fall attempted sample.¹⁰¹

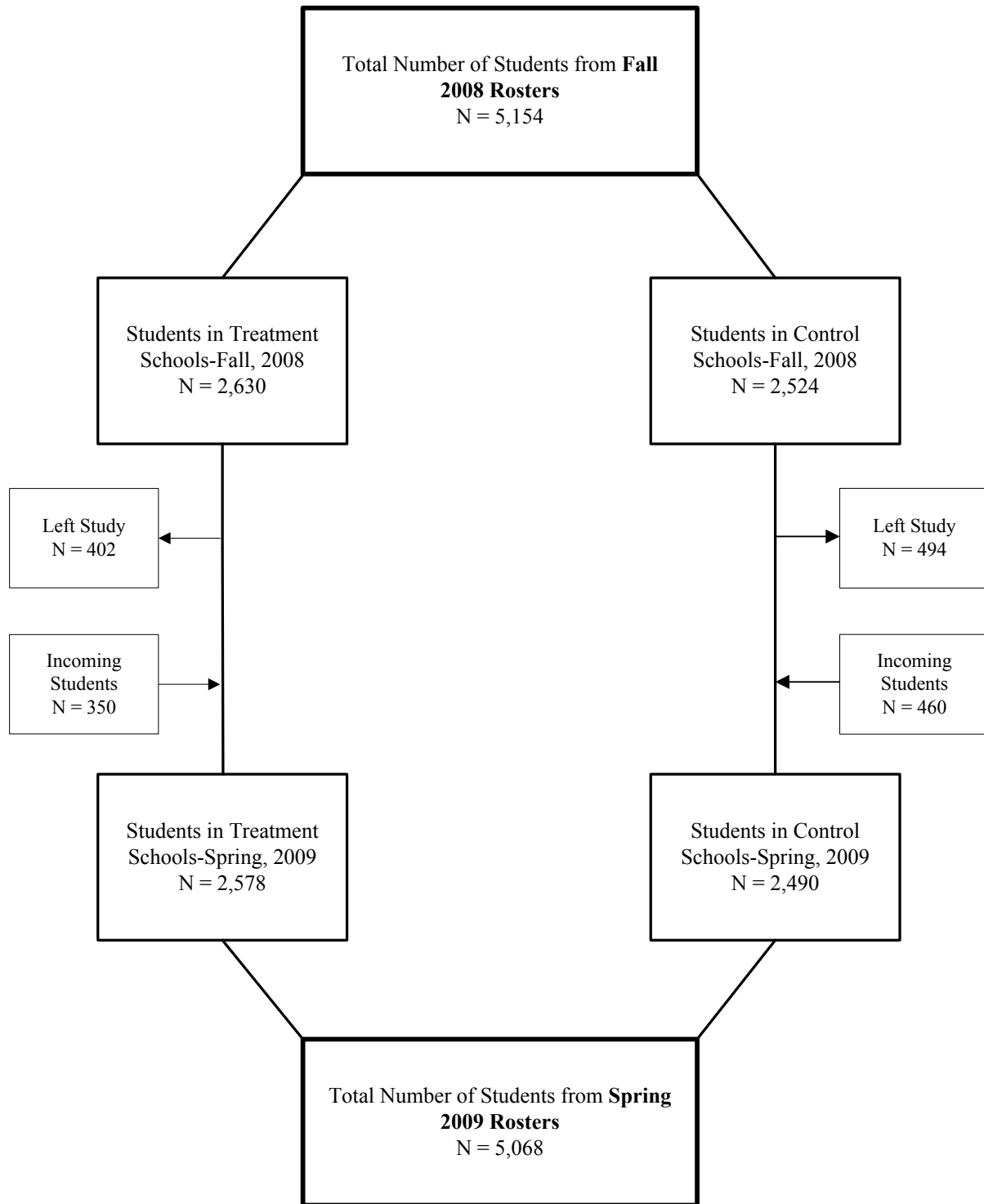
The ordered lists were constructed as follows:

- The first 9 students on the list were divided between continuing and incoming students in the same proportion as the class overall. All the continuing student slots among these first 9 students were allocated to students who were in the fall attempted sample (unless there were insufficient numbers of such students). If the spring class roster included more students from the fall attempted sample than could be accommodated in the first nine slots, sequential random draws from among the students who had been in the fall attempted sample were used to fill the slots. Similarly, sequential random draws from among the incoming students were used to fill the incoming student slots.
- The remainder of the list was also divided between continuing and incoming students in the same proportion as the class overall. In drawing students to fill the continuing student slots in positions 10 through 16+, students from the fall attempted sample were given a higher sampling probability than other continuing students, but the other continuing students still had a probability greater than zero of being drawn into the spring sample.
- Once the list of 16 students for each classroom was formed, the study team began testing from line number 1, moving down through the list until a sample of 9 students had been achieved. If 1 of the first 9 students in the spring sample was absent or refused, the team tested the next eligible student. Otherwise-eligible students on the list who were absent or refusing (including any fall refusing students who fell within the spring attempted sample) counted against the spring participation rates. As in the fall, response rates were calculated as the percentage of the (spring) attempted sample tested, and makeup sessions were held for any schools in which the overall response rate (across all classrooms) was less than 80 percent.

¹⁰⁰ If the first 16 draws included students who were known to be refusing, additional students were added to the list to ensure that the overall list was large enough to yield a sample of 9 students for testing. However, the refusing students were included in the calculation of the response rate.

¹⁰¹ The category of continuing students included students who stayed in the same school but switched from one eligible mathematics class to another. This definition was chosen to maximize the number of fall-tested students in the spring testing and to simplify the sampling strategy.

Exhibit A-1. Student Turnover During the Second Year of the Study



Characteristics of the Student Sample

The second-year *student impact analysis sample* consisted of 2,132 eligible students who were on the spring 2009 class rosters, whose parents consented to the data collection requests, and who were successfully tested using our spring 2009 student sampling procedures. About 51 percent (1,083 students) were from treatment schools, and the remaining 49 percent (1,049 students) were from control schools. Exhibit A-2 demonstrates how this sample was constructed.

The second-year impact analysis sample is a subset of the second-year spring *expanded student sample*, which consisted of 4,859 students (2,364 treatment students and 2,495 control students) who were on the spring 2009 student rosters, whose parents had consented to the study, and for whom the district was able to provide student records containing demographic information and scores on district-administered mathematics tests.

Table A-4 compares the characteristics of students in the student impact analysis sample and students who appear only in the spring expanded student sample. Test results presented in the table demonstrate that the two samples are statistically different in the following respects: students included in the second-year impact analysis sample were younger, were less likely to have special education status, and had higher sixth-grade state mathematics test scores than students who appear only in the expanded sample. These differences between the impact sample and other students from the same classes suggest that one should exercise caution in interpreting impact analysis findings.

Exhibit A-2. Construction of the Second-Year Student Impact Analysis Sample

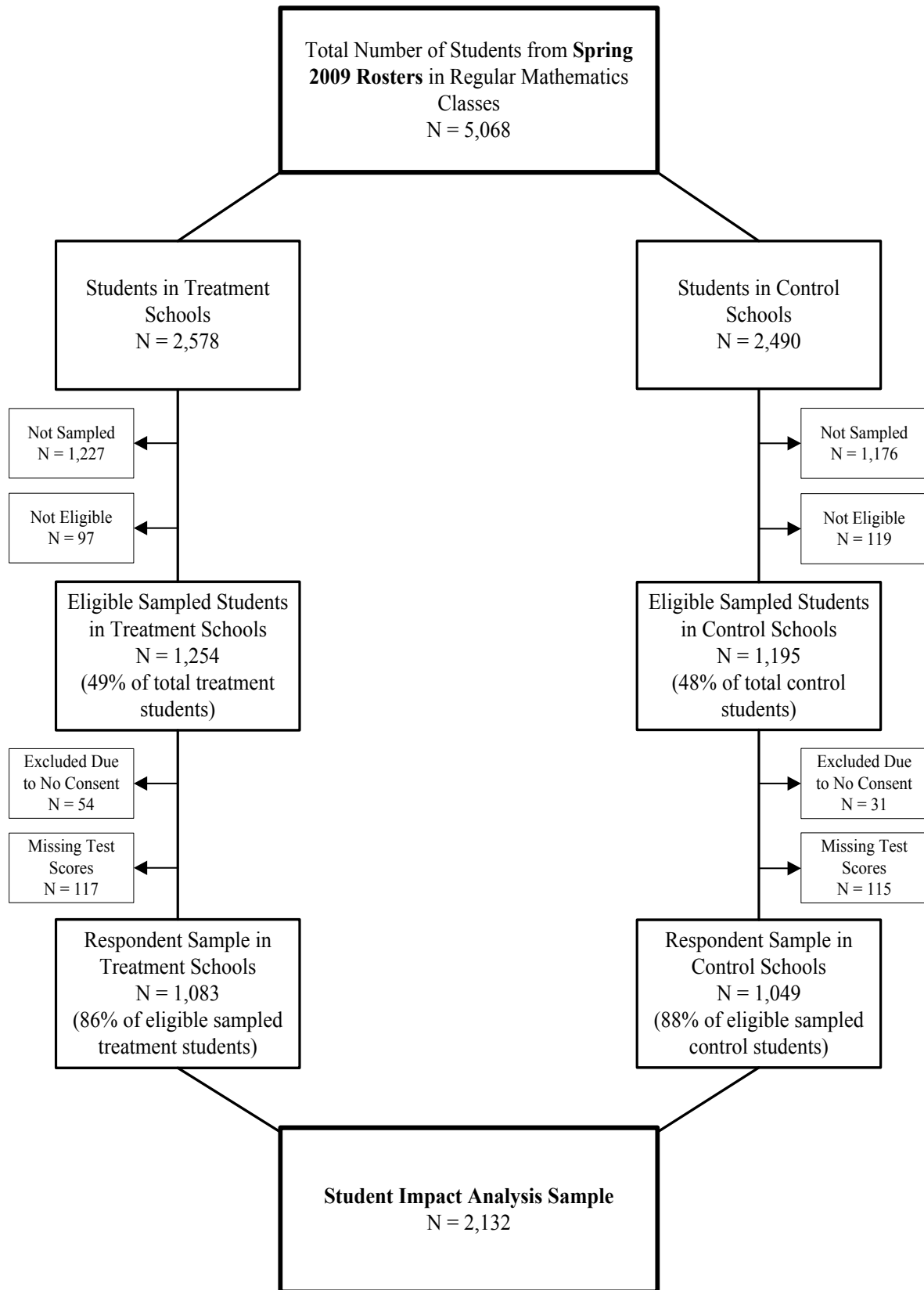


Table A-4. Characteristics of Students Included and Not Included in the Second-Year Student Impact Analysis Sample: Second-Year Spring Expanded Student Sample

Student Characteristics	Student Impact Analysis Sample	Students Not in Impact Analysis Sample	Estimated Difference	P-value for Estimated Difference
Age (years) ^a	12.6	12.7	-0.1*	<0.01
Students Eligible for Free or Reduced-Price Lunch (percent)	74.1	73.9	0.1	0.91
Race/Ethnicity (percent)				
White, non-Hispanic	32.7	33.9	-1.2	0.38
Black, non-Hispanic	36.0	37.0	-1.0	0.48
Hispanic	26.8	24.8	2.0	0.13
Asian/Pacific Islander	2.4	2.2	0.1	0.81
Other	2.0	1.9	0.1	0.79
Male (percent)	50.9	51.4	-0.5	0.73
English as a Second Language (percent)	15.5	16.2	-0.7	0.55
Special Education Status (percent)	10.4	18.1	-7.8*	<0.01
Sixth-Grade Mathematics Scores on State Accountability Assessment (standardized, in effect size)	0.16	0.01	0.15*	<0.01

Sample Size: N = 4,859 students (2,364 treatment, 2,495 control).

SOURCE: Study District Records.

NOTES: ^a Age was calculated as the age (in years) of a student as of September 1, 2008.

The analyses are based on a three-level model controlling for random assignment block.

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

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

APPENDIX B
DETAILS OF DATA COLLECTION AND
ANALYTICAL APPROACHES

APPENDIX B

DETAILS OF DATA COLLECTION AND ANALYTICAL APPROACHES

Details of Data Collection

This appendix provides additional detail on the data collection activities and analyses described in Chapter 1. Information is provided on the five instruments used in the second year of the study: implementation form, coach log, teacher knowledge test, teacher survey, and student achievement test.¹⁰² Procedures for developing the instruments and for training the data collectors are also described, as are the response rates achieved.

Implementation Form

To gauge the implementation of the professional development (PD), a member of the study team attended each institute or seminar day and completed a form on which he or she tracked the amount of time devoted to each instructional segment as well as the use of intended instructional materials. The forms for each institute or seminar day were customized on the basis of the planned agenda, PowerPoint slides, and handouts so that the observer was able to cross check the presentation against the plan. In addition, the observer noted (1) whether the facilitator “closed” each segment by reviewing the main learning goals or mathematical points of the segment and (2) the extent to which the facilitator created links to the curriculum by discussing the district’s text and/or standards. Because the observers were not experts in PD or mathematics, they were not asked to judge the quality of the presentation, but simply to record what was done.

The implementation form was developed on the basis of a similar form used in an earlier study of PD in reading conducted by AIR and MDRC (Garet et al. 2008). The form was pilot tested during the study’s planning year, when the PD was also undergoing pilot testing. Observers were provided with one hour of training and detailed written guidance on how to implement the form. The reliability of the implementation form was not formally assessed.

Coach Log

After each coaching event, coaches completed logs in which they recorded the amount of contact time with each teacher and the kinds of coaching activities pursued. Over the course of the coaching event, the coaches were expected to record the starting and stopping time of each separate coaching activity and the names of the teachers participating in that activity. They then checked precoded boxes to indicate the nature of the activity (e.g., planning a lesson, co-teaching a lesson, conducting a peer observation), the mathematical focus of the activity, and the pedagogical focus of the activity.

Like the implementation form, the coach log was developed on the basis of a similar form used in the earlier study of PD in reading conducted by Garet et al. (2008). Research staff provided instructions for completing the logs and discussed the definitions of the code categories with the

¹⁰² The first year of the study used the same instruments plus a teacher observation protocol.

coaches. As the coach logs were received from the field, they were reviewed for completeness and clarity, and research staff followed up with coaches as necessary.

Teacher Knowledge Test

The teacher knowledge test was developed by AIR specifically for this study. The test addresses 12 key understandings related to positive rational numbers. These 12 key understandings, which informed the development of the PD program as well as the development of the teacher knowledge test, are listed in the text box below. Six of the key understandings are in the general area of fractions and decimals, and six are in the general area of ratios, rates, proportions, and percents.

Key Understandings Measured in Teacher Knowledge Test

- Defining fraction as a number and the meaning of the numerator and denominator
- Equivalent fractions and the role of a/a
- Adding, subtracting, and ordering fractions and the role of common denominators
- Multiplying and dividing fractions and the role of the reciprocal and inverse operations
- Decimals are an extension of place value
- Rational numbers can be expressed as fractions, decimals, percents, or ratios
- Ratios are comparisons by division
- Rates are special cases of ratios
- Percents are ratios
- Additive vs. multiplicative relationships
- Proportions are equivalent ratios
- Direct and inverse proportional relations

Every item in the teacher knowledge test was designed to measure either common knowledge of mathematics (CK) or specialized knowledge of mathematics for teaching (SK) associated with one of these key understandings. CK items address the teacher's ability to understand concepts and carry out operations in the area of positive rational numbers, as typically taught in seventh grade. SK items are intended to measure the more specialized knowledge required to successfully teach positive rational numbers content at this grade level, including knowledge associated with planning instruction, delivering instruction, and assessing student understanding. Sample CK and SK items are presented below. At baseline, the average teacher in the study sample had an approximately three-quarters chance of answering each of these sample items correctly.¹⁰³

¹⁰³ The probability for a given teacher to answer a particular item correctly was computed as follows: probability = $\text{EXP}(D)/(1+\text{EXP}(D))$, where D is the difference between the teacher's knowledge score and the item's difficulty derived from IRT scaling.

Example Teacher Knowledge Items

Common Knowledge (CK): Equivalence among rational numbers

1. Which number represents a point on a number line different from the other three?
A 1.2
*B $\frac{5}{4}$
C 120%
D $2 - \frac{4}{5}$

Specialized Knowledge (SK): Rates are special cases of ratios/Assessing student understanding

2. Your class is grappling with a situation in which you can buy three grapefruit for \$5. Gina says, "That's easy. Just think about one grapefruit for $\frac{\$5}{3}$."
Which BEST describes Gina's understanding?
*A She understands how to find a unit rate.
B She understands that all ratios are rates.
C She understands how to solve rate problems.
D She understands that she has to divide 3 by 5 to make the problem easier.

A total of 72 items were developed so that each teacher could complete a different 24-item form at each of the three scheduled administrations. Each form included 12 CK items and 12 SK items, equally distributed across the 12 key understandings. The 72 items were divided into six 12-item half-forms that could be administered in different combinations to facilitate scaling.

During development, several types of information were used to refine the test. First, test items were administered to samples of volunteer teachers in one-on-one cognitive (think-aloud) interviews. These interviews enabled us to refine the items by providing insight into how teachers understood the items and the mathematical or reasoning processes that were accessed in answering them. Second, all test items were reviewed for accuracy and relevance by mathematicians familiar with teacher education. Third, to obtain rough estimates of item difficulty, four pilot forms were created, and each was administered to nine volunteer teachers.

For operational use, the teacher knowledge test was administered as an untimed, proctored test. Teachers in the treatment group took the baseline test at the start of the first day of PD provided by the study. Teachers in the control group took the baseline test at their school, sometime within the first 10 weeks of the fall semester during the year in which they entered the study. Teachers in both groups took a different form of the teacher knowledge test at their school, sometime within the last 8 weeks of the spring semester, during each year they participated in the study. Thus, a given teacher could take the teacher knowledge test up to three times.

Item Response Theory (IRT) analysis procedures were used to produce three knowledge scores (a total score, a CK score, and an SK score) for each teacher, at each point in time, on a common scale. The IRT item parameters were calculated based on all fall 2007 and spring 2008 responses to the teacher knowledge test. The same item parameters were then used to score the teacher knowledge tests in the second year of the study.

IRT assumes a mathematical model for the probability that an examinee will respond correctly to a specific test question, given the examinee's overall performance and the characteristics of the question. When different examinees complete different blocks of items, IRT scoring accounts for the relative difficulty of the items. Individual teacher scores are thus not affected by variations in the average difficulty of the items on a given test form.

Classical test theory-based reliability indices, such as Cronbach's alpha, are not appropriate for the teacher knowledge test given the spiraling of forms. (As noted above, at each administration, each teacher was administered a subset of the total item pool based on his or her assigned test form.) The reliability coefficient for the instrument was, therefore, calculated as the marginal reliability, $\bar{\rho}$ (Sireci, Thissen, and Wainer 1991). This statistic is equivalent in interpretation to classical internal consistency estimates of reliability, an upper-bound estimate of true reliability. Based on the combined fall and spring sample from the first year of the study, the marginal reliability estimates for the teacher knowledge test were $\bar{\rho} = 0.74$ (Overall), $\bar{\rho} = 0.65$ (CK), and $\bar{\rho} = 0.56$ (SK).¹⁰⁴ These levels of reliability are sufficient for the purposes of this study, which draws comparisons between groups of teachers (i.e., between treatment and control teachers).

Although there is no normative information on how a representative sample of teachers would perform on the teacher knowledge test, there is some limited evidence that the test is measuring the intended constructs. First, when the test was administered to the facilitators responsible for delivering the PD program, the facilitators scored significantly higher than the teachers participating in the first year of the study, as reported in Chapter 2. Second, for teachers in the first year of the study, baseline scores on the teacher knowledge test were significantly correlated ($r = 0.256$; $p = .0004$) with teachers' self-reports of number of mathematics courses taken.

Teacher Surveys

Teacher surveys were administered four times during the two-year study. During the first year, surveys were administered at baseline (fall 2007) and at the end of the school year (spring 2008) to all teachers who were present at the time of the survey. During the second year, surveys were administered in fall 2008 to teachers who were new to the study in the second year and in spring 2009 to all teachers who were present at the end of the study.

The survey questions addressed two major constructs:

- The characteristics of the teachers that might affect their baseline knowledge of mathematics for teaching and/or their ability to benefit from the PD program
- The nature and extent of the mathematics-related PD received by treatment and control teachers during the time period of the study

¹⁰⁴ Reliabilities for study instruments were not re-estimated in the second year of the study.

To measure the latter construct, several different survey questions were combined into scales using exploratory factor analysis of survey data from the first year of the study. The reliability of the scales was evaluated using Cronbach's alpha. Exhibit B-1 shows the reliability and contributing items for each scale.¹⁰⁵

¹⁰⁵ Scale reliabilities were calculated using data from fall 2007 and spring 2008 surveys administered to treatment and control teachers from all 12 districts that participated in the first year of the study. The fall 2007 survey included questions about PD during the initial months of the study, when the treatment teachers were participating in the summer institutes. The spring 2008 survey included questions about PD during the remainder of the first study year.

Exhibit B-1. PD Characteristics Scales Used in Analysis of Service Contrast

Scale	Contributing Items
<p>PD Emphasis on Mathematics Content</p> <p>Emphasis on fractions and decimals Reliability = 0.89</p> <p>Emphasis on percent, ratio, rate, and proportion Reliability = 0.85</p> <p>Emphasis on whole numbers/integers, algebra, geometry, probability, and statistics Reliability = 0.72</p>	<p>Emphasis on fractions</p> <p>Emphasis on decimals</p> <p>Emphasis on percents</p> <p>Emphasis on ratios, rates, and proportional reasoning</p> <p>Emphasis on whole numbers</p> <p>Emphasis on algebra</p> <p>Emphasis on geometry</p> <p>Emphasis on probability and statistics</p>
<p>PD Emphasis on Pedagogic Content</p> <p>Emphasis on pedagogical topics intervened upon Reliability = 0.79</p> <p>Emphasis on pedagogical topics not intervened upon Reliability = 0.69</p>	<p>How students think about and learn mathematics (including common student difficulties)</p> <p>How to plan and structure lessons</p> <p>How to use representations to convey mathematical concepts</p> <p>How to ask students questions and provide feedback</p> <p>How to use your mathematics curriculum/textbook</p> <p>How to interpret and use assessment data to guide instruction</p> <p>How to organize and manage a classroom</p> <p>How to teach students with diverse needs</p> <p>How to use technology in mathematics instruction</p>
<p>Active participation in PD Reliability = 0.74</p>	<p>Practiced what you learned and received feedback</p> <p>Led group discussions</p> <p>Conducted a demonstration of a lesson, unit, or skill</p> <p>Developed student materials and practiced using them</p>
<p>Collective participation in PD Reliability NA (single item)</p>	<p>Did you participate with most or all of the mathematics teachers from your department or grade level?</p>
<p>Relevance of the PD to my own teaching Reliability = 0.81</p>	<p>Consistent with your own goals for your professional development</p> <p>Aligned with state or district standards and/or assessments</p> <p>Supportive of the use of district-adopted curricular materials</p> <p>Relevant to the mathematics you taught this year</p> <p>Focused on material at the right level of difficulty, given your prior knowledge of mathematics and mathematics teaching</p>

Exhibit continues on next page

Exhibit B-1. PD Characteristics Scales Used in Analysis of Service Contrast (continued)

Scale	Contributing Items
Clarity of purpose of the PD Reliability = 0.79	Logically connected from one day or session to the next Clear about what you should learn from the PD experience Clear about how you could use what you learned from the PD experience in your classroom
Use of plan-observe-debrief coaching cycle in PD Reliability = 0.90	Planning lessons with your coach or mentor Being observed in your classroom by your coach or mentor Debriefing lessons with your coach or mentor
Observing coaches and/or other teachers as part of PD Reliability = 0.71	Observing OTHER TEACHERS in their classrooms with your coach or mentor Co-teaching lessons, or watching demonstration lessons led by your coach or mentor

SOURCE: Teacher Surveys (fall 2007 and spring 2008 respondents).

NOTE: Reliabilities are based on Cronbach's alpha.

Student Achievement Test

A customized, computer-adaptive student achievement test was constructed for the study by the Northwest Evaluation Association (NWEA). The test was restricted to positive rational number content and drew on a customized item base that contained nearly 1,200 positive rational number items abstracted from the larger NWEA item bank of scaled, operational mathematics items.¹⁰⁶

Each student was presented with 30 items from the customized item base, selected adaptively from the topic areas of fractions, decimals, percents, and ratios/proportions.¹⁰⁷ Specifically, each student was presented with items matching the distribution shown in Table B-1. The order of presentation of items was designed to ensure that items from a given content area or a given cognitive dimension were distributed across the test session. Within the constraints imposed by this ordering, however, the adaptive process was continuous; that is, each new item was chosen from the available pool on the basis of the current best estimate of the student's achievement level. The test algorithm prevented the same student from seeing a given item more than once—either during a single test session or across time (baseline and outcome testing).

¹⁰⁶ There is no single version of the NWEA computer-adaptive tests. Customers typically have the tests that they purchase customized to reflect their own state or district standards.

¹⁰⁷ We decided on 30 items to ensure that the test could be administered in a single class period, which simplified logistics and decreased the impact on instructional time.

Table B-1. Distribution of Items on NWEA Rational Number Test

	Fractions	Decimals	Percent	Ratio/Proportion	Total
Concepts	3	2	2	2	9
Operations	4	1	1	2	8
Applications	4	1	1	7	13
Total	11	4	4	11	30

The text box below presents example items from the customized test. The first two items were easy for most students in our study population; the last two were challenging.

Example Items From NWEA Rational Number Test

Example 1:

1. What is $6/12$ in simplest form?

- *A. $1/2$
- B. $12/24$
- C. $2/4$
- D. $1/6$
- E. $1/12$

Example 2:

2. $0.32 \div 8 =$

- A. 4.3
- B. 0.15
- *C. 0.04
- D. 280
- E. 43.75

Example 3:

3. What is $2 \frac{1}{8}$ written as a decimal?

- A. 2.25
- B. 2.1
- *C. 2.125
- D. 2.13
- E. 2.5

Example 4:

4. 8 is what % of 32?

- A. $1/4$
- B. 4%
- C. 20%
- *D. 25%
- E. 2.56%

The NWEA test was not intended to be a timed test, and students were allowed to take as much time as they needed to complete the test. However, the test software did not allow students to skip items. Table B-2 provides information on the mean test duration for students in the treatment and control conditions, in fall 2007 and spring 2008.

Table B-2. Average Test Duration (Minutes) for NWEA Rational Number Test, by Treatment Status and Test Wave in the First Year of the Study: First-Year Student Baseline and Impact Analysis Samples

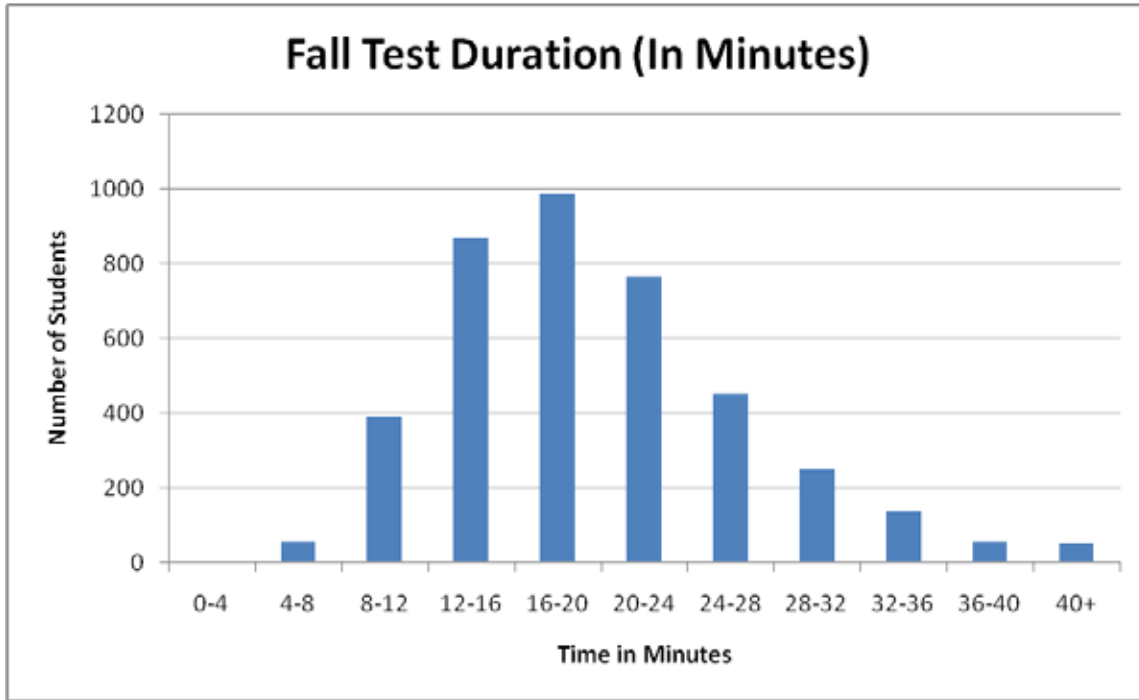
	Treatment Group Mean (S.D.)	Control Group Mean (S.D.)
Fall 2007 (baseline)	20.28 (7.46)	20.31 (6.91)
Spring 2008 (impact)	18.30 (7.26)	17.56 (6.92)

Sample Size: N = 4,211 students (2,178 treatment, 2,033 control).

SOURCE: Administration Records for Fall 2007 and Spring 2008 NWEA Rational Number Tests.

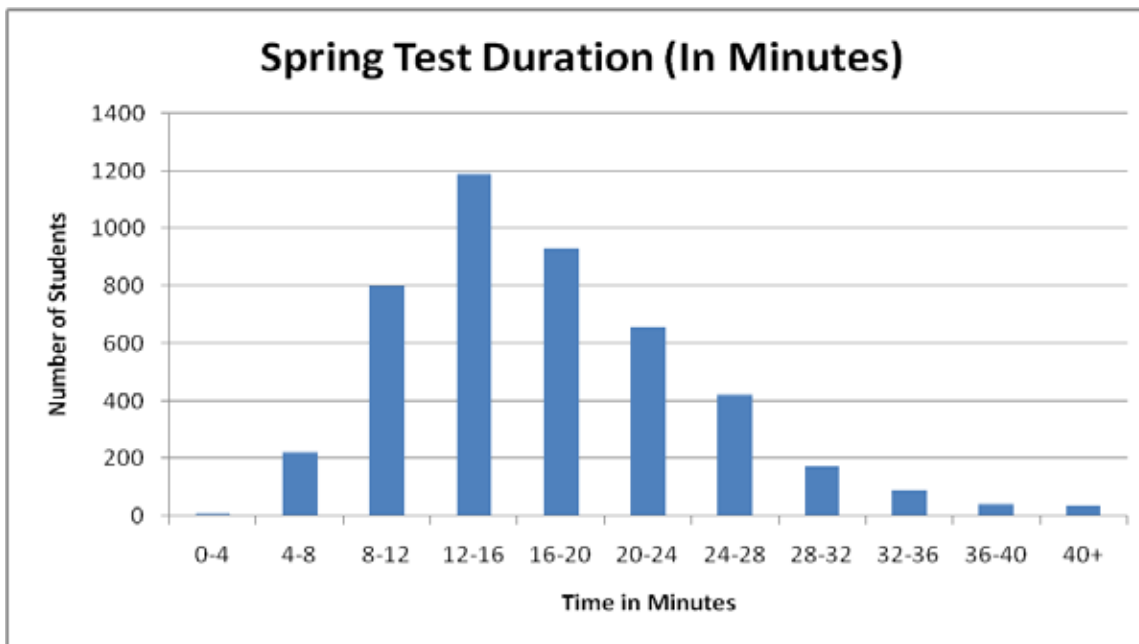
To assess the extent to which students made a serious effort to complete the test, we examined the distribution of test durations in the first year of the study. Figure B-1 shows the distribution of test durations for the fall 2007 and spring 2008 administrations of the student test (all students combined). As can be seen, a small number of students took less than four minutes to complete the test. This may indicate that they were not attending to the content of the test items for some or all items, which could invalidate their scores. However, because we did not intend to use the test data to evaluate individual students, but only to estimate group performance, we decided to leave the test scores for such students in the analysis files for both the first- and second-year impact analyses.

Figure B-1. Test Duration for Student Test Administration, by Test Wave in the First Year of the Study: First-Year Student Baseline and Impact Analysis Samples



Sample Size: N = 4,211 students (2,178 treatment, 2,033 control).

SOURCE: Administration Records for Fall 2007 NWEA Rational Number Test.



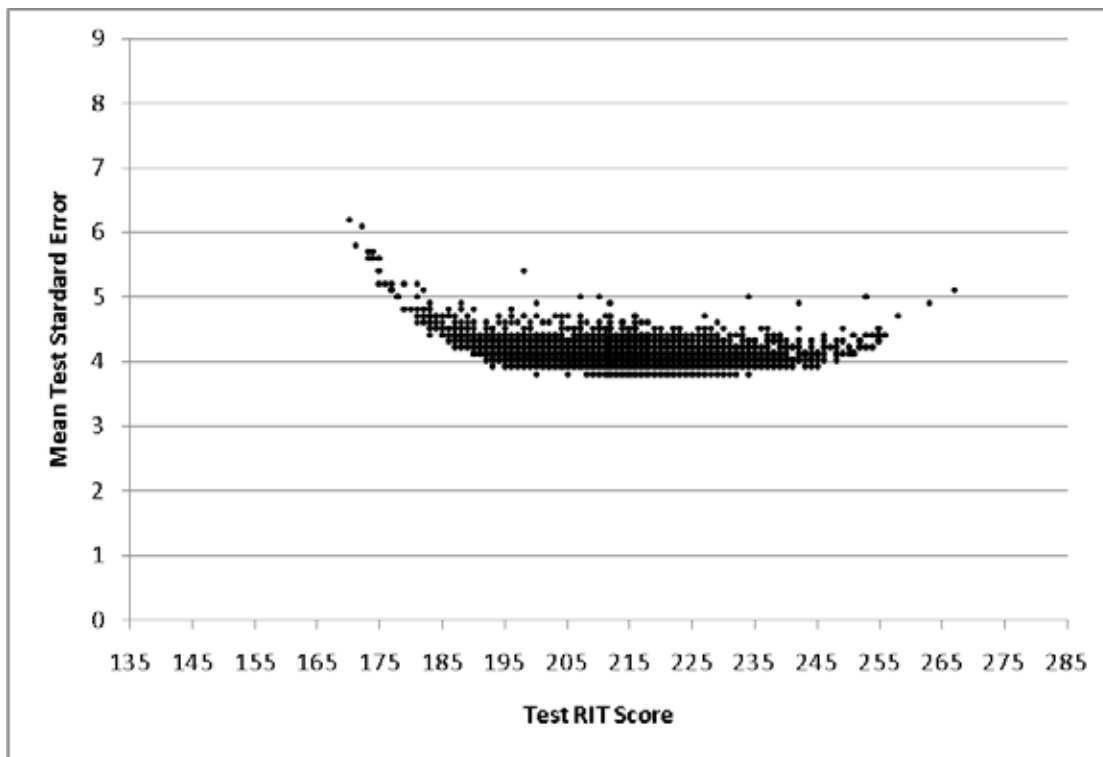
Sample Size: N = 4,528 students (2,336 treatment, 2,192 control).

SOURCE: Administration Records for Spring 2008 NWEA Rational Number Test.

Each NWEA assessment provides an estimate of a student’s position on an underlying Rasch-model scale of achievement, which NWEA calls a RIT scale.¹⁰⁸ For this study, the regular item parameters used for NWEA operational testing were used to place students on the scale. Details on the item parameters and scaling methods used by NWEA can be found in the NWEA technical manual and in a special NWEA report on test reliability and validity estimates (NWEA 2003, 2004).

For the customized test used in the study, each student received a total score, a fractions and decimals subscore, and a ratio and proportion subscore that also included performance on the percent items. The average standard error for the total score on the student test was 4.08 at baseline of the first year. Figure B-2 provides more detail on the relationship between student test scores (expressed as values on the RIT scale) and standard errors. The standard error curve remains relatively flat until close to the ends of the distribution. For reference, note that the RIT scores corresponding to the 25th, 50th, and 75th percentiles for our 2007 fall testing were 206, 215, and 223, respectively.

Figure B-2. Distribution of Standard Errors by Total RIT Score on Fall 2007 NWEA Rational Number Test: First-Year Student Baseline Analysis Sample



Sample Size: N = 4,211 students (2,178 treatment, 2,033 control).

SOURCE: Fall 2007 NWEA Rational Number Test.

To aid with the interpretation of the total score results, NWEA also constructed customized, seventh-grade norms by reanalyzing data from its Growth Research Database—a large database

¹⁰⁸ NWEA (2003) explains that RIT is its shorthand for “Rasch unit.” Each student’s RIT score is 200 plus the product of 10 times his or her logit score. NWEA derives logit scores from a one-parameter IRT model (i.e., a Rasch model).

compiled from operational NWEA testing.¹⁰⁹ The data set represents students from a wide range of school districts in many states, but it is not specifically tailored to be nationally representative. For the customized norms, test records from 704,750 seventh-grade students who had attempted three or more rational number items and had answered at least one of them correctly were rescored using only the rational number items.¹¹⁰ This norming sample had a mean scale score of 228.2 for fall and 232.6 for spring, with standard deviations of 17.17 and 18.28, respectively. By comparison, the first-year study sample had a mean scale score of 214.6 (standard deviation 13.07) for fall 2007, which placed them at the 20th percentile, and 217.0 (standard deviation 14.70) for spring 2008, which placed them at the 21st percentile.

Response Rates

Table B-3 shows the response rates in the two-year districts for each of the teacher and student instruments described in this appendix, separately by treatment status. None of the differences between the treatment and control groups is statistically significant. For teachers, the response rates for the fall 2007 instruments are calculated as the percentage of responses received from eligible teachers who were teaching regular seventh grade mathematics classes during the first 10 weeks of the fall semester. The response rates for the spring 2008 instruments are calculated as the percentage of responses received from eligible teachers who were teaching regular seventh-grade mathematics classes during the last 8 weeks of the spring semester.¹¹¹ Response rates for the teacher instruments used in the second year were calculated in the same manner except that the fall 2008 instruments were applicable only to teachers who joined the study at the beginning of the 2008–2009 school year. The response rates for students are based on students in the attempted second-year baseline sample and the attempted second-year impact sample. Appendix A explains the manner in which the attempted samples for students were constructed.

¹⁰⁹ NWEA reported that its Growth Research Database contained more than 115 million scores at the time at which the customized, seventh-grade norms were constructed. For a description of the Growth Research Database, see NWEA (2003).

¹¹⁰ Because the NWEA test is adaptive, the number of rational number items presented to seventh-grade students varies from individual to individual. Among 735,930 students in the Growth Research Database who had attempted at least three rational number items, 704,750 (96 percent) answered at least one of rational number item correctly and were included in the norming sample. Among the norming sample students, the number of rational number items attempted ranged from 3 to 19, with a median of 5 and an average of 6.3.

¹¹¹ Only one teacher per teaching position was eligible to be included in a particular sample. If teacher turnover occurred during the time window that defined the sample, data from the teacher who was active for the greater part of the window were included.

Table B-3. Response Rates for All Teacher and Student Measures, by Treatment Status: Two-Year Districts^a

Data Source	Overall	Treatment Group	Control Group
Teachers			
Fall 2007 Teacher Survey (percent)	95.8	>=93.3	>=94.0
Fall 2007 Teacher Knowledge Test (percent)	>=96.8	>=93.3	>=94.0
Fall 2007 Teacher Sample Size ^b	95	45	50
Spring 2008 Teacher Survey (percent)	>=96.8	93.2	>=94.0
Spring 2008 Teacher Knowledge Test (percent)	>=96.8	93.2	>=94.0
Spring 2008 Teacher Sample Size ^c	94	44	50
Fall 2008 Teacher Survey (percent)	>=92.0	100.0	>=83.3
Fall 2008 Teacher Knowledge Test (percent)	>=92.0	100.0	>=83.3
Fall Teacher Sample Size ^d	37	19	18
Spring 2009 Teacher Survey (percent)	96.7	>=93.3	>=93.6
Spring 2009 Teacher Knowledge Test (percent)	96.7	>=93.3	>=93.6
Spring Teacher Sample Size ^e	92	45	47
Students			
Fall 2008 NWEA Rational Number Test (percent)	89.1	88.8	89.4
Fall 2008 Student Sample Size (attempted sample) ^e	2,105	1,063	1,042
Spring 2009 NWEA Rational Number Test (percent)	87.0	86.4	87.6
Spring 2009 Student Sample Size (attempted sample) ^f	2,449	1,254	1,195

SOURCE: Fall 2007, Spring 2008, Fall 2008, and Spring 2009 Teacher Surveys; Fall 2007, Spring 2008, Fall 2008, and Spring 2009 Teacher Knowledge Tests; Fall 2008 and Spring 2009 NWEA Rational Number Tests.

NOTES:

^aWhere the number of non-responding teachers was <3, the exact response rate is suppressed. Instead the response rate is reported as being larger than or equal to the percentage represented by the number of potential respondents minus 3.

^bThe sample size for teachers is based on the number of eligible teachers teaching regular seventh-grade mathematics classes during the first 10 weeks of the school year. Only one teacher per teaching position was eligible to be included in a particular sample. If teacher turnover occurred during the time window that defined the sample, data from the teacher who was active for the greater part of the window were included.

^cThe sample size for teachers is based on the number of eligible teachers teaching regular seventh-grade mathematics classes during the last eight weeks of the school year. Only one teacher per teaching position was eligible to be included in a particular sample. If teacher turnover occurred during the time window that defined the sample, data from the teacher who was active for the greater part of the window were included.

^dThe sample size for teachers is based on the number of new, eligible teachers teaching regular seventh-grade mathematics classes during the first 10 weeks of the school year. Only one teacher per teaching position was eligible to be included in a particular sample. If teacher turnover occurred during the time window that defined the sample, data from the teacher who was active for the greater part of the window were included.

^eThe students tested with the NWEA instrument were chosen from an ordered list of random draws for each eligible class, as listed on the fall student rosters. The sample size reported here includes all students attempted in the fall. The response rate is calculated as the number of tested students divided by the number of attempted students.

^fThe students tested with the NWEA instrument were chosen from an ordered list of random draws for each eligible class, as listed on the spring student rosters. The sample size reported here includes all students attempted in the spring. The response rate is calculated as the number of tested students divided by the number of attempted students.

Technical Notes on Analytic Approaches

This part of the appendix provides two sets of technical notes that accompany the Analytic Approaches section in Chapter 1 of the report. The first section describes the statistical model used to estimate the impacts of the PD program on teacher and student outcomes. The second section addresses issues related to tests of impacts on multiple outcome measures and subgroups.

Statistical Models for Estimating Impacts

The second year of the study focused on the impact of professional development on two types of outcomes: teacher knowledge and student achievement. All teachers and students with available outcome measures were included in the impact analysis.

The basic approach for the impact analyses was a pooled-sample approach, which combined the data from all six districts in the study sample, using dummy variables to control for district and block differences as fixed effects. This approach used the whole data set in a single analysis and allowed us to see how the impact of the PD program differed across districts and whether those differences are statistically significant. Separate analyses were conducted for the first year impacts (spring 2008) and the second year impacts (spring 2009) for each type of outcome. We specify the models as follows:

Teacher Knowledge Impact

The Model

$$Y_{jk} = \sum_m \sum_n \gamma_{0mn} B_{mnk} + \sum_m \gamma_{1m} T_k D_{mk} + \gamma_2 Y_{-1jk} + \sum_l \gamma_{3l} Z_{jkl} + \mu_k + \nu_{jk} \quad (\text{B-1})$$

Where:

- Y_{jk} = outcome measurement for teacher j from school k,
- B_{mnk} = one if school k is in block n (n = 1 to 11) in district m (m = 1 to 6) and zero otherwise,
- D_{mk} = one if school k is in district m (m = 1 to 6) and zero otherwise,
- T_k = one if school k is assigned to receive the treatment and zero otherwise,
- Y_{-1jk} = fall teacher knowledge test total score for teacher j from school k,
- Z_{jkl} = background characteristic l for teacher j from school k,
- μ_k, ν_{jk} = a school-level and a teacher-level random error, respectively, assumed to be independently and identically distributed.

This model reflects the hierarchical structure of the dataset with teachers nested within schools and is estimated as a multilevel model using the MIXED procedure in SAS. The weighted average γ_1 of the estimated γ_{1m} coefficients for the six districts (using the number of treatment

schools in each district as weight) is the estimated program effect on teacher knowledge for the average treatment school in the study sample. A two-tailed t-test is used to assess whether γ_1 differs from zero. For both the first-year and second-year impacts, we also report the estimate γ_1 as an effect size, based on the standard deviation for the control group (pooled across all 12 districts in the first-year study sample) from the spring 2008 data collection. In addition, to help readers interpret the findings, we report the impact on teacher knowledge in terms of the estimated probability of getting the average item correct on the test.

Covariates in the Model

Other than the block indicators and the treatment indicator, we included a set of teacher-level covariates in the model to improve the precision of the estimates. To serve this purpose, we selected variables that we anticipated would be correlated with the outcome measure. For teacher knowledge outcomes, in addition to the baseline teacher knowledge total scores, we included the following teacher characteristics: total teaching experience; teaching experience in middle school mathematics; teacher's education level (master's degree or not); mathematics major or not; and number of postsecondary mathematics courses taken.¹¹²

Student Achievement Impact

The Model

$$Y_{ijk} = \sum_m \sum_n \gamma_{0mn} B_{mnk} + \sum_m \gamma_{1m} T_k D_{mk} + \gamma_2 Y_{-1ijk} + \gamma_3 Y_{-1k} + \sum_l \alpha_l X_{lijk} + \mu_k + \nu_{jk} + \varepsilon_{ijk} \quad (\text{B-2})$$

Where:

- Y_{ijk} = achievement measurement for student i from class j in school k,
- B_{mnk} = one if school k is in block n (n = 1 to 11) in district m (m = 1 to 6) and zero otherwise,
- D_{mk} = one if school k is in district m (m = 1 to 6) and zero otherwise,
- T_k = one if school k is assigned to receive the PD treatment and zero otherwise,
- Y_{-1ijk} = baseline score for student i from teacher j in school k,
- Y_{-1k} = average baseline NWEA score for school k,
- X_{lijk} = demographics for student i from teacher j in school k,
- $\mu_k, \nu_{jk}, \varepsilon_{ijk}$ = a school-level, class-level, and student-level random error, respectively, assumed to be independently and identically distributed.

¹¹² The measures of teacher experience included in the model were indicator variables coded 1 = less than three years of teaching experience and 0 = three or more years of teaching experience.

The error term structure reflects the hierarchical or nested structure of the data, which has students nested within classes and classes nested within schools. The model is estimated as a three-level hierarchical model using the MIXED procedure in SAS.

The weighted average γ_1 of the estimated γ_{1m} coefficients for the six districts (using the number of treatment schools in each district as weight) is the estimated program effect on student achievement for the average treatment school in the study sample. A two-tailed t-test is used to assess whether γ_1 differs from zero. Impact results are reported both in terms of scaled scores and effect sizes.¹¹³ We also report the mean outcome levels for the treatment and control groups in terms of percentile ranks based on the customized norming sample of the NWEA rational numbers test to provide context for the findings.

Covariates in the Model

The covariates in the regression model include school average NWEA test scores from fall 2008, student-level achievement scores from fall 2008, and student-level demographic information such as gender, age, race/ethnicity, English as a second language (ESL)/Limited English proficiency (LEP) status, and/or free or reduced-price lunch status from student record data. They are included in the model to improve the precision of the impact estimates.

Table B-4 provides information on the number of cases for which covariate data were missing. For all but one covariate, the number of cases of missing data was zero or negligible. The relatively large number of missing cases for student baseline NWEA scores is the result of including in the second-year impact analysis sample a mix of “continuing” students—most of whom were members of the second-year student baseline analysis sample—and “incoming” students who were not in the baseline analysis sample.

¹¹³ We use the control group standard deviation on the spring 2008 NWEA Rational Number Test (pooled across the 12 districts in the first-year study sample) to calculate effect size for both the spring 2008 and spring 2009 outcomes. This approach was chosen to be consistent with the way teacher outcome effect sizes were calculated.

Table B-4. Missing Data for Teacher and Student Characteristics Used as Covariates in the Impact Models, Second-Year Teacher and Student Impact Analysis Samples

Characteristics	Number Missing	Percent Missing
Covariates for Teacher Knowledge		
Baseline Teacher Knowledge Test Total Score	6	3.5
Mathematics Major	0	0.0
Educational Level: M.A. and Above	0	0.0
Years of Teaching Experience	0	0.0
Years of Teaching Experience in Middle School Mathematics	0	0.0
Number of Postsecondary Mathematics Courses Taken	0	0.0
Number of Postsecondary Mathematics Education Courses Taken	0	0.0
Sample Size: N = 89 teachers.		
Covariates for Student Achievement		
Age	0	0.0
Eligible for Free or Reduced-Price Lunch	<4	<0.2
Race/Ethnicity	13	0.6
Male	0	0.0
English as a Second Language	0	0.0
Special Education Status	9	0.4
Fall 2008 NWEA Total Score	712	33.4
Sample Size: N = 2,132 students.		
SOURCE: Fall 2007 Teacher Knowledge Test; Fall 2008 Teacher Knowledge Test; Teacher Survey; Fall 2008 NWEA Rational Number Test; Study District Records.		

Addressing Risks Associated With Multiple Hypothesis Tests

When making judgments about statistical significance, it is important to recognize potential problems associated with conducting multiple hypothesis tests. Specifically, when multiple tests are conducted, the problem of making a Type I error (falsely concluding there is an impact when there is no true effect) rises; but efforts to control for this problem may reduce statistical power.

To control the Type I error rate while maintaining power insofar as possible, we used a two-step approach to address the multiple hypothesis testing issue. The first step in this process is to divide the impact analyses into two tiers: confirmatory analyses, which provide answers to our key research questions; and exploratory analyses, which facilitate a deeper analysis of our key findings and what they mean. In Chapter 3, all the analyses of teacher knowledge (*Total Score*, *CK Score*, and *SK Score*) and student achievement (*Total Score*, *Fractions and Decimals Score*, and *Ratio and Proportion Score*) at the end of the second year of implementation are the focus of this report and are considered confirmatory. All the analyses of teacher knowledge and student achievement measures at the end of the first year of implementation using the sample of teachers or students in the one-year districts or the two-year districts are included as supplemental information and are considered exploratory. All the analyses discussed in Chapter 4 are non-experimental and are therefore considered exploratory analyses.

The second step, which is applied only to the confirmatory analyses, involves using composite “qualifying” tests to assess the overall statistical significance of a set of confirmatory impact estimates within a measurement domain. The qualifying test uses a composite index averaging the individual measures included in a domain. When a qualifying test indicates a statistically significant difference between groups, it suggests that there are in fact statistically significant findings in one or more of the individual tests included and hence adds confidence to the interpretation of the individual findings. However, when a qualifying test does not indicate a statistically significant difference between groups, it calls into question the interpretation of specific findings within that domain. Therefore, when reporting significant findings for any of the confirmatory analyses, we also report the results of the associated qualifying test and indicate whether or not the qualifying test results are supportive of the specific findings.

The qualifying tests were specified as follows, for the confirmatory analyses in the two domains—teacher knowledge and student achievement—on which the impact analyses focus:

- For the teacher knowledge domain, we treated the *Total Score* as a qualifying test for its subscores: *CK Score* and *SK Score*.
- For the student achievement domain, we treated the *Total Score* as a qualifying test for its subscale scores: *Fractions and Decimals Score* and *Ratio and Proportion Score*.

APPENDIX C
SUPPLEMENTAL INFORMATION ON THE
DESIGN AND IMPLEMENTATION OF THE PD
PROGRAM

APPENDIX C

SUPPLEMENTAL INFORMATION ON THE DESIGN AND IMPLEMENTATION OF THE PD PROGRAM

This appendix supplements the description of the professional development (PD) program and its implementation given in Chapter 2. The first section describes the scheduled coverage of seventh-grade mathematics topics in each district participating in the study, a key context for the study's PD program. The second and third sections provide details of each PD provider's approach and a detailed list of each PD provider's summer institute and seminar day topics, respectively. The fourth section describes supplemental PD implementation results separately for each PD provider. The fifth section presents two supplemental views of the PD *participation* results: (a) separately for each PD provider and (b) based on teachers' dates of entry into the study. The final section presents the service contrast for each PD provider.

Scheduled Coverage of Mathematics Topics

During the two school years in which the PD was implemented, the topics in rational numbers that were the focus of the study's PD program accounted for 36 percent of the curriculum covered in seventh-grade mathematics in the two-year study districts. Time explicitly devoted to fractions and decimals ranged from zero to 18 percent. Time explicitly dedicated to ratio, rate, proportion, and percent ranged from 10 percent to 37 percent.¹¹⁴

The timing of this instruction also varied. Some districts completed all of their scheduled instruction on rational numbers prior to the winter break, while others spread this instruction throughout the school year or concentrated the instruction in the final weeks of the school year. PD seminars and coaching were scheduled to coincide with planned rational number instruction to the extent possible. Despite this customization of PD schedules, variation in the extent and timing of scheduled rational number topics may still have moderated the potential impact of the PD program provided by the study. For example, in districts where most of the rational number instruction occurred early in the school year, teachers would have had less time to practice and apply lessons learned from the PD.

Detailed Specifications of Each PD Provider's Approach to Institutes and Seminars and to Coaching

Institutes and Seminars

America's Choice

As noted in Chapter 2, in America's Choice PD segments, teachers were asked to solve sets of mathematics problems. Teachers worked on the problem sets individually or in small groups, and these activities were followed by structured discussions led by the facilitator. The problem sets were

¹¹⁴ Of course, teachers may also have addressed students' understanding of rational number topics at other points during the school year, in the context of providing instruction on other mathematics topics.

designed to lay the groundwork for or reinforce definitions of key rational number concepts and to illustrate common student misconceptions. The facilitator guide and training provided explicit guidance about how to direct the discussions of the problem sets (including examples of questions to ask teachers) and identified the mathematics concepts to emphasize during the summary portion of each PD segment.

America's Choice also introduced several specific representations designed to help teachers convey rational number topics, including the number line, the double number line, ratio tables, area models, set models, and strip diagrams. The facilitators explained how teachers should use these representations with students, and the problem sets offered opportunities for teachers to practice using them.

In addition, America's Choice introduced "questioning strategies" that teachers could use to elicit student thinking. These questioning strategies included asking a student to restate another student's reasoning, asking a student to apply his or her own reasoning to another student's conclusions, and using the "Say More" technique, in which teachers asked individual students to say more about an answer or explanation.

Finally, within each seminar day, the design allowed teachers (individually and with feedback from the facilitator) to work on a rational number lesson, linked to their textbook, that was then to be implemented on a coaching day following the seminar.

Pearson Achievement Solutions

Pearson Achievement Solutions used a single problem or task to structure each PD segment. Each task was designed to elicit multiple approaches and to fuel extended discussions about the core ideas, common student approaches, and potential student misconceptions associated with the task. Like the America's Choice facilitators, the Pearson Achievement Solutions facilitators had guidance regarding the types of questions to ask participants and the key ideas to emphasize during these discussions. However, the Pearson Achievement Solutions tasks were more open-ended, and facilitators were told to use their expertise to determine how to structure the discussions and whether to extend the length of any given PD segment. For institute days, the Pearson Achievement Solutions facilitator guide and training provided a summary statement to be used for each PD segment but did not specify how much time should be devoted to the segment summaries. For seminar days, the facilitator guide did not explicitly specify segment summaries.

For some segments in each study year, the problem used to structure the segment was designed to elicit multiple representations of rational number concepts, and these problems provided the basis for the discussion of the representations. Facilitators were expected to address the number line, ratio table, area model, and set model.

Pearson Achievement Solutions organized its coverage of the rational numbers content by focusing the summer institutes in both years on deepening teachers' understanding of three "big ideas" about rational numbers: (1) "numbers represent quantities," (2) "rational numbers are about division," and (3) "a ratio shows a comparison by division." To structure each seminar day, facilitators gave teachers a problem or used a problem nominated by the teachers that was designed to elicit multiple student approaches to a particular rational number concept and to reveal potential student misconceptions. After teachers worked on the problem and considered various ways their

students were likely to approach the problem, they collaboratively planned how they would teach the lesson, which they were to insert into the curriculum during a subsequent coaching visit. The lesson was planned using a lesson format that had four sections: identifying the mathematical goal(s) of the lesson, monitoring and classifying student approaches to the task, providing a summary statement of the core mathematics of the lesson, and developing formative assessment questions.

Coaching

America's Choice

During coaching visits, America's Choice facilitators were expected to work with teachers on whatever lesson the teachers were planning to teach according to the district pacing plan. Although the coach visits were scheduled to occur when teachers were teaching rational number content, variation in teachers' progress through the curriculum meant that some of the America's Choice coaching days could take place when teachers were teaching other topics.

America's Choice planned to engage in different individual and group coaching activities on each of the two-day coaching visits. In the first year, during the first two-day coaching visit, the facilitator observed a teacher teaching a typical lesson, modeled a lesson for the teacher, and then met with the teacher to discuss the strengths and weaknesses of both lessons. On the second coaching visit, the facilitator worked with the teacher to practice using the mathematical discussion techniques, first with small groups of students and then with the whole class. The third coaching visit emphasized the teachers' use of a tool that was designed to help teachers monitor students' understanding of the main mathematics ideas in a lesson. The fourth coaching visit revolved around peer observations, in which one or more teachers used an observation tool that focused on a prespecified set of student behaviors as they observed another teacher. The fifth and final coaching visit was designed to have pairs of teachers co-plan and co-teach a lesson and debrief with the facilitator afterward.

During the second year, all coaching visits included "fishbowl" peer observation activities, whereby teachers used a structured observation protocol to observe one another teach and then reviewed the lessons using a peer-directed, collaborative discussion format, during pre- and postlesson conferences. The final two-day coaching visit included time for teachers to discuss ways to sustain collegial activities after the America's Choice coaching program was completed.

Pearson Achievement Solutions

The Pearson Achievement Solutions approach during the first year focused each coaching visit on a rational number lesson that used a problem introduced by the PD provider during the preceding seminar. Each lesson was planned collaboratively during the seminar, and teachers were asked to insert it into the curriculum for the coaching visit. During the second year, Pearson Achievement Solutions followed a similar model, but some of the lessons that were planned collaboratively and then used as the basis for coaching were built around problems that the teachers created or that appeared in their textbooks.

On the first day of each two-day coaching visit at each school, the facilitator was expected to observe each teacher as he or she taught the planned lesson. Then, in an after-school group meeting, the facilitator led a discussion about how the lesson was implemented across the classrooms. The collegial meeting was designed to focus on how the content was presented, how students responded

to end-of-lesson assessment questions, and what the next lesson should look like after the teacher had reflected upon the current lesson. On the second day of the two-day coaching visit, the teachers implemented the lesson they planned during the after-school meeting. They also participated in a shorter group debriefing meeting with the facilitator, who summarized the main ideas in the lesson taught during the second day of the visit and encouraged teachers to think about how they could use the material discussed during the two-day visit in future lessons. The Pearson Achievement Solutions coaching plan emphasized observation, collaborative planning, and group debriefing focused on common lessons; it did not emphasize having the facilitator model instruction or co-teach in the classroom.

Content and Structure of Institutes and Seminars

America's Choice

Summer Institute—First Year

The following outline indicates the segment topics for each of the three summer institute days conducted by America's Choice in the first year of the study:

Summer Institute Day 1: Introduction to Fractions

- Introduction to the study
- Representing fractions on the ruler or number line
- Conceptual background for representing fractions
- Recognizing and representing fraction situations
- Equivalent fractions
- Representing fractions using a card sort
- Daily wrap up, reflections, and evaluations

Summer Institute Day 2: Compare and Order Numbers

- Welcome, goals, and parking lot
- Defining decimals
- Zooming in on the number line
- Matching fractions and decimals that are close to each other
- Ordering a mixed set of fractions and decimals
- Planning for effective mathematical discussions
- Multiplying and dividing with decimals
- Daily wrap up, reflections, and evaluations

Summer Institute Day 3: Multiply and Divide Fractions

- Welcome, goals, and parking lot
- Representing multiplication of fractions
- What does division of fractions mean?
- Homework discussion
- Two types of division
- Developing action plans
- Why does “invert and multiply” work?
- Daily wrap up, reflections, and evaluations

Seminars—First Year

The following outline indicates the segment topics for each of the five seminar days conducted by America’s Choice in the first year of the study:¹¹⁵

Seminar Day 1: Ratio Tables

- Welcome and goals
- Representing ratios
- Introducing ratio tables
- Connecting ratio tables and fractions
- Lesson planning
- Connecting to algebra
- Closing the seminar day

Seminar Day 2: Strip Diagrams and Scale Factor

- Welcome and goals
- Introducing strip diagrams
- Applying strip diagrams
- Homework discussion
- Scale Factor
- Applying scale factor
- Closing the seminar day

Seminar Day 3: Rate

- Welcome and goals
- Applying unit rate
- Is it really addition of fractions?
- Homework discussion
- Do all rate problems involve proportions?
- Closing the seminar day

Seminar Day 4: Percent

- Welcome and goals
- Developing number sense for percents
- Lesson planning: What’s the math?
- Three kinds of percent problems
- Anticipating student responses
- Applying percent
- Closing the seminar day

¹¹⁵ The five daylong seminars were reordered in each district so that each seminar was scheduled when the topics covered by that seminar were expected to be taught, according to the district’s curriculum pacing guide. For the America’s Choice district that used *Glencoe*, it was not possible to schedule all three of the ratio, rate, and proportion seminars when these topics were expected to be covered in the schools because ratio, rate, and proportion are covered in a single chapter in the *Glencoe* curriculum.

Seminar Day 5: Add and Subtract Fractions

- Welcome and goals
- On rulers and number lines
- Using shaded area models
- Teaching rational numbers and ratio
- Mathematical justification
- Closing all the seminars

Summer Institute—Second Year

The following outline indicates the segment topics for each of the four summer institute days conducted by America's Choice in the second year of the study (two days for the makeup institute for teachers new to the study in 2008–2009, and two days for the institute for new and returning teachers):

New Teacher (Makeup) Summer Institute Day 1: Fractions, Decimals, and Percents

- Introduction to the study
- Conceptual background for representing fractions
- Representing fractions using a card sort
- Defining decimals
- Zooming in on the number line
- Matching fractions and decimals that are close to each other
- Developing number sense for percents
- Daily wrap-up, reflections, and evaluations

New Teacher (Makeup) Summer Institute Day 2: Ratios

- Welcome, goals, and parking lot
- Introducing ratio tables
- Introducing strip diagrams
- Applying strip diagrams
- Connecting to algebra
- Daily wrap-up, reflections, and evaluations

New and Returning Teacher Institute Day 1: Multiply and Divide Fractions

- Welcome, goals, and parking lot
- Assessing understanding of first-year PD content
- Representing the multiplication and division of rational numbers
- Meaning of division of fractions
- Understanding the two types of division
- Long-term unit planning
- Daily wrap-up, reflections, and evaluations

New and Returning Teacher Institute Day 2: Planning Ratio Lessons

- Welcome, goals, and parking lot
- Planning for effective closings

- Mistakes and misconceptions with ratios and rates
- Lesson planning
- Sharing ratio lesson plans
- Daily wrap-up, reflections, and evaluations

Seminars—Second Year

The following outline indicates the segment topics for each of the three seminar days conducted by America’s Choice in the second year of the study:

Seminar Day 1: Ratios and Rates

- Welcome and goals
- Look alike rectangles
- Similar triangles
- Homework review
- Pre-lesson conferencing: “Fishbowl” process
- Combining rates
- Daily wrap-up, reflections, and evaluations

Seminar Day 2: Strip Diagrams

- Welcome and goals
- Extended practice with strip diagrams
- Homework review
- Looking at student work
- Reading, writing, and talking mathematics
- Crafting strip diagram problems
- Daily wrap-up, reflections, and evaluations

Seminar Day 3: Ratio Tables and Percent

- Welcome and goals
- Applications of percent
- Ratio tables revisited
- Homework review
- Planning for presentations
- Rational number lesson segment presentations
- Daily wrap-up, reflections, and evaluations

Pearson Achievement Solutions

Summer Institute—First Year

The following outline indicates the segment topics for each of the three summer institute days conducted by Pearson Achievement Solutions in the first year of the study:

Summer Institute Day 1: Numbers Represent Quantities

- Introduction to the study and the summer institute
- Alternate representations of numbers and the concept of number
- Decimal notation and place value
- Numbers as points on the number line
- Number systems studied in k–8 mathematics
- Identify, create, and use situations that require partitive or measurement division
- Closing and evaluations

Summer Institute Day 2: Rational Numbers Are About Division

- Opening and review
- Use division to generate fractions
- Interpretations of rational numbers written in fraction form
- Use divisions and subdivisions to show that different numerical representations of rational numbers are equivalent
- Comparing and ordering rational numbers
- Explore operations with rational numbers
- Fraction and decimal unit planning
- Closing and evaluations

Summer Institute Day 3: A Ratio Shows a Comparison by Division

- Opening and review
- Describe the relationship between two numbers
- What types of comparisons can we make?
- Compare two part problems: one that describes a part : part comparison and one that describes a part : whole comparison
- Use ratio tables to examine multiplicative relationships
- Examine student work on proportional reasoning problems
- Closing and evaluations

Seminars—First Year

The following outline indicates the segment topics for each of the five seminar days conducted by Pearson Achievement Solutions in the first year of the study:¹¹⁶

Seminar Day 1: Fraction Foundations

- Introductory activity
- Working through the task: comparing fractions
- Introduce lesson overview
- Developing learning goals and writing formative assessment
- Create lesson flow chart
- Introduce lesson plan structure
- Plan “introduce the task”
- Plan “students work on the task”
- Plan “public discussion of the task”
- Plan “direct instruction”
- Finalize formative assessment
- Closing and evaluation

Seminar Day 2: Fraction Follow Up

- Introductory activity
- Working through the task: operations with fractions
- Review lesson overview
- Developing learning goals and writing formative assessment
- Create lesson flow chart
- Review lesson plan structure
- Plan “introduce the task”
- Plan “students work on the task”
- Plan “public discussion of the task”
- Plan “direct instruction”
- Finalize formative assessment
- Closing and evaluation

Seminar Day 3: Ratio and Proportion Foundations

- Introductory activity
- Working through the task: ratios
- Review lesson overview
- Create lesson flow chart
- Developing learning goals and writing formative assessment
- Review lesson plan structure

¹¹⁶ The five daylong seminars were reordered in each district so that each seminar was scheduled when the topics covered by that seminar were expected to be taught, according to the district’s curriculum pacing guide. For the two Pearson Achievement Solutions districts that used *CMP*, it was difficult to align the content of seminars 1 and 2 to primary topics in the curriculum. Although most of the units in the seventh-grade *CMP* curriculum include fraction review problems, none of the units or lessons made fractions a primary focus. The content of seminars 3–5 was more closely aligned with the primary topics in other *CMP* units, two of which focused in-depth on ratio, proportion, and percent.

- Plan “introduce the task”
- Plan “students work on the task”
- Plan “public discussion of the task”
- Plan “direct instruction”
- Finalize formative assessment
- Closing and evaluation

Seminar Day 4: Ratio and Proportion Follow Up

- Introductory activity
- Working through the task: proportions
- Review of what two features of teaching help students understand mathematics
- Review lesson overview and lesson plan structure review lesson plan structure
- Plan “introduce the task”
- Plan “students work on the task”
- Plan “public discussion of the task”
- Plan “direct instruction”
- Finalize formative assessment
- Closing and evaluation

Seminar Day 5: Connections

- Introductory activity
- Review of what two features of teaching help students understand mathematics
- Working through the task: sharing pizza
- Review lesson overview and lesson plan structure
- Developing learning goals and writing formative assessment
- Plan “introduce the task”
- Plan “students work on the task”
- Plan “public discussion of the task”
- Plan “direct instruction”
- Finalize formative assessment
- Closing and evaluation

Summer Institute—Second Year

The following outline indicates the segment topics for each of the four summer institute days conducted by Pearson Achievement Solutions in the second year of the study (two days for the makeup institute for teachers new to the study in 2008–2009, and two days for the institute for new and returning teachers):

New Teacher (Makeup) Summer Institute Day 1: Numbers Represent Quantities and Fractions Are About Division

- Welcome and introductions
- Alternate representations of numbers and the concept of number
- Conceptual background for representing fractions
- Alternate representations of numbers and the concept of number
- Decimal notation and place value

- Numbers as points on the number line
- Number systems studied in k–8 mathematics
- Identify, create, and use situations that require partitive or measurement division
- Use division to generate fractions
- Interpretations of rational numbers written in fraction form
- Closing

New Teacher (Makeup) Summer Institute Day 2: Fractions Are About Division and a Ratio Shows a Comparison by Division

- Welcome and introductions
- Use divisions and subdivisions to show that different numerical representations of rational numbers are equivalent
- Comparing and ordering rational numbers
- Describe the relationship between two numbers
- What types of comparisons can we make?
- Compare two part problems—one that describes a part : part comparison and one that describes a part : whole comparison
- Use ratio tables to examine multiplicative relationships
- Examine student work on proportional reasoning problems
- Introduce lesson plan structure
- Closing

New and Returning Teacher Institute Day 1: Focus on Fractions

- Welcome and introductions
- Working through the task: fractions
- Review lesson overview
- Develop learning goals and formative assessment
- Plan “introduce the task”
- Plan “students work on the task”
- Plan “public discussion of the task”
- Plan “direct instruction”
- Finalize formative assessment
- Fraction unit planning
- Closing

New and Returning Teacher Institute Day 2: Focus on Ratio/Proportion

- Welcome and introductions
- Use ratio tables to examine multiplicative relationships
- Working through the task: ratios
- Review lesson overview
- Develop learning goals and formative assessment
- Plan “introduce the task”
- Plan “students work on the task”
- Plan “public discussion of the task”
- Plan “direct instruction”

- Finalize formative assessment
- Ratio and proportion unit planning
- Closing

Seminars—Second Year

The following outline indicates the segment topics for each of the three seminar days conducted by Pearson Achievement Solutions in the second year of the study:

Seminar Day 1: Fractions

- Welcome and introductions
- Analyze and present findings from lesson taught during coaching
- Review and revise fraction unit plan
- Lesson planning: fractions
- Closing

Seminar Day 2: Ratio and Proportion

- Welcome and introductions
- Analyze and present findings from lesson taught during coaching
- Discuss math teaching research article
- Working through the task: ratios
- Working through the task: proportions
- Working through the task: percents
- Closing

Seminar Day 3: Fractions and Ratios

- Welcome and introductions
- Analyze and present findings from lesson taught during coaching
- Review and revise fraction and ratio unit plans
- Lesson planning: ratio and proportion
- Closing

Supplemental PD Implementation Results by PD Provider

This fourth section describes supplemental PD implementation results separately for each PD provider.

Institutes and Seminars

Across the eight institute and seminar days in the first year of the study, the program was designed to provide equal coverage to fractions and decimals (four days) and ratio, rate, proportion, and percent (four days). Based on the PD providers' reflections on the first year of the program, the providers selected material for the second year that they believed would reinforce and deepen the teachers' understanding, particularly in areas in which the teachers had seemed weakest during the first year. PD provider America's Choice designed its five days of institutes and seminars in the second year to allocate more coverage to ratio, rate, and proportion (four days) than to fractions and decimals (one day), and Pearson Achievement Solutions designed its second year institutes and

seminars to provide equal coverage to fractions and decimals (two and a half days) and ratio, rate, and proportion (two and a half days).

The results presented in Table C-1 reflect the different choices of the two providers in terms of the content coverage planned for each year of the study.

Table C-1. Teacher Institutes and Seminars—Approximate Hours of Implemented Time Covering Specific Content Areas, by PD Provider: Two-Year Districts

Content Area	America's Choice		Pearson Achievement Solutions	
	Mean Actual Hours	S.D.	Mean Actual Hours	S.D.
First Year				
Fractions, Decimals	23.4	0.33	23.0	1.28
Percent, Ratio, Rate, Proportion	22.2	1.69	20.9	0.77
Second Year				
Fractions, Decimals	10.4	0.17	18.5	1.06
Percent, Ratio, Rate, Proportion	30.3	1.08	18.8	2.56

Sample Size: N = 6 two-year districts.

SOURCE: 2007–2008 Institute and Seminar Implementation Form; 2008–2009 Institute and Seminar Implementation Form.

NOTES: Hours per topic are an approximation based on the primary focus of each agenda section.

Second-year estimates include hours for the makeup institutes that were offered to teachers who joined the study after the first-year summer institutes.

Table C-2 indicates that America's Choice, which used a prescriptive plan that stressed coverage of all segments, reallocated 0.6 and 0.5 hours of planned segments each day, respectively, in the first and second years of the program. Pearson Achievement Solutions, whose planned flexibility allowed some segments to run long and others to be omitted, reallocated an average of 1.5 hours per day during the implementation of each year of the institute and seminars.

Table C-2. Teacher Institutes and Seminars—Mean Reallocated Hours and Percent of Planned Segments Omitted and Abbreviated, by PD Provider: Two-Year Districts

	America's Choice	Pearson Achievement Solutions
First Year		
Mean Hours Reallocated	0.6	1.5
Percent of Segments Omitted	0.0	4.3
Percent of Segments Highly Abbreviated (lasted 50 percent or less of intended time)	9.4	21.5
Percent of Segments Abbreviated (lasted 51–75 percent or less of intended time)	11.9	15.0
Second Year		
Mean Hours Reallocated	0.5	1.5
Percent of Segments Omitted	0.0	10.5
Percent of Segments Highly Abbreviated (lasted 50 percent or less of intended time)	2.5	10.5
Percent of Segments Abbreviated (lasted 51–75 percent or less of intended time)	9.4	21.6

Sample Size: N = 393 planned PD segments in the first year (160 for America's Choice, 233 for Pearson Achievement Solutions); 312 planned PD segments in the second year (159 for America's Choice, 153 for Pearson Achievement Solutions); 48 institute and seminar days in the first year (24 for America's Choice, 24 for Pearson Achievement Solutions); 42 institute and seminar days in the first year (21 for America's Choice, 21 for Pearson Achievement Solutions).

SOURCE: 2007–2008 Institute and Seminar Implementation Form; 2008–2009 Institute and Seminar Implementation Form.

NOTES: The results were calculated across PD days, with each PD day weighted by the number of planned PD segments. Reallocated hours include the intended duration for omitted segments and the difference between the intended and actual duration for abbreviated segments (i.e., segments that did not last for the intended duration). Minutes reallocated from one segment may have been shifted to another segment or skipped and never delivered. Results presented elsewhere indicate that the majority of the reallocated hours were shifted to other segments rather than skipped entirely.

Second-year estimates include hours for the makeup institutes that were offered to teachers who joined the study after the first-year summer institutes.

Table C-3 summarizes implementation results on the use of multiple delivery formats, planned materials, and other planned features of the institute and seminar days.

Each PD day was designed to include a combination of individual, small-group, and whole-group activities, as well as teacher presentations. We assessed the percent of days on which all four types of activities occurred. On average, all four planned types of activities occurred on 79 percent of the first-year institute and seminar days and on 52 percent of the second-year institute and seminar days.

Each provider's PD plan described a set of participant materials to be used each day, including problem sets, worksheets, charts, readings, and other materials. Observers recorded whether each of these planned materials was used. On average, 80 percent or more of the planned materials were used on 40 percent of the first-year institute and seminar days and 74 percent of the second-year institute and seminar days. These figures reflect the abbreviation or omission of segments described earlier, facilitators' decisions about the best use of time within particular segments, and the development of more focused sets of planned materials for the second year. Differences by provider were more marked in the first year, when America's Choice implemented 71

percent of planned materials and Pearson Achievement Solutions implemented 8 percent. In the second year, the two providers implemented 81 percent and 67 percent of planned materials, respectively.

According to the providers' PD plans, the main ideas were to be summarized at the end of each segment. On average, the main ideas were summarized at the end of at least 80 percent of the day's segments for 58 percent of the institute and seminar days in the first year and 64 percent of the institute and seminar days in the second year. In the first and second years, respectively, 75 and 86 percent of institute and seminar days met this criterion for America's Choice, which provided its facilitators with explicit guidance about summary segments and allocated substantial time for summary segments. The corresponding percentages for Pearson Achievement Solutions were 42 and 43 percent.

The PD providers planned to make explicit links on each institute and seminar day between the PD content and the specific seventh-grade mathematics curriculum used in the study schools. The percentage of institute and seminar days on which at least 15 minutes were devoted to making explicit links to the curriculum was, on average, 63 percent for the first year and 76 percent for the second year.¹¹⁷ The 15-minute criterion for explicit links was met on 88 percent of first-year seminar days and 81 percent of second-year seminar days for America's Choice, which drew on lessons from the existing curriculum. Pearson Achievement Solutions met the criterion on 38 percent of first-year seminar days and 71 percent of second-year seminar days, reflecting the fact that Pearson's plan for the first year of PD made extensive use of inserted lessons that were specially developed for the PD program and not part of the existing curriculum.

Finally, we examined the overall level of teacher engagement for each day of the PD.¹¹⁸ On 98 percent of the first-year institute and seminar days and 100 percent of the second-year days, at least 80 percent of the participating teachers were engaged in the PD, as measured by the study observers.

¹¹⁷ The extent to which the PD made explicit links to the curriculum materials, standards, or assessments used by teachers in the study districts was determined on the basis of the cumulative total time spent per day on these links. The Institute and Seminar Implementation Form asked observers to indicate whether no time, less than 5 minutes, between 5 and 15 minutes, between 15 and 30 minutes, or more than 30 minutes were spent making such links during the day.

¹¹⁸ The Institute and Seminar Implementation Form included an item on teacher engagement that had five possible responses: 20 percent or less, 40 percent, 60 percent, 80 percent, or 100 percent of participating teachers were actively engaged for the majority of the day. Observers were to record teacher engagement at least four times across the day. Teachers were to be counted as actively engaged if they were watching the facilitator, working problems, or listening to or contributing to the discussion. To be actively engaged, teachers did not need to be enthusiastic, just attentive.

Table C-3. Percent of Teacher Institute and Seminar Days on Which Features of the PD Matched the Plan, Overall and by PD Provider: Two-Year Districts

	Total		America's Choice		Pearson Achievement Solutions	
	Mean Percent	S.D.	Mean Percent	S.D.	Mean Percent	S.D.
First Year						
<i>Percentage of PD days on which:</i>						
Delivery formats matched plan	79.2	0.41	87.5	0.34	70.8	0.46
Participant materials essentially matched plan	39.6	0.49	70.8	0.46	8.3	0.28
Main ideas were summarized	58.3	0.50	75.0	0.44	41.7	0.50
Links were made to curriculum, standards, or assessment during at least 15 minutes of the day	62.5	0.49	87.5	0.34	37.5	0.49
80 percent or more of participating teachers were engaged	97.9	0.14	100.0	0.00	95.8	0.20
Second Year						
<i>Percentage of PD days on which:</i>						
Delivery formats matched plan	52.4	0.51	90.5	0.30	14.3	0.36
Participant materials essentially matched plan	73.8	0.45	81.0	0.40	66.7	0.48
Main ideas were summarized	64.3	0.48	85.7	0.36	42.9	0.51
Links were made to curriculum, standards, or assessment during at least 15 minutes of the day	76.2	0.43	81.0	0.40	71.4	0.46
80 percent or more of participating teachers were engaged	100.0	0.00	100.0	0.00	100.0	0.00

Sample Size: N = 48 institute and seminar days in the first year (24 for America's Choice, 24 for Pearson Achievement Solutions); 42 institute and seminar days in the second year (21 for America's Choice, 21 for Pearson Achievement Solutions).

SOURCE: 2007–2008 Institute and Seminar Implementation Form; 2008–2009 Institute and Seminar Implementation Form.

NOTES: Second-year estimates include days for the makeup institutes that were offered to teachers who joined the study after the first-year summer institutes.

Segments were the unit of implementation coding and are demarcated by planned transitions in agenda subtopics or activities. Delivery formats included facilitator lecture, individual activities, small-group activities, whole-group activities, and teacher presentations. Each day of institutes or seminars was to include instances of individual, small-group, and whole-group activities, as well as teacher presentations. The delivery format for a day of PD was coded as “matched plan” if all four formats were included in the day’s PD.

Participant materials included materials such as worksheets, problem sets, charts, and readings. PowerPoint slides were not included as participant materials in this analysis. The extent to which participant materials matched the plan was determined on the basis of the percentage of planned participant materials covered by the facilitator each day. Participant materials were coded as “essentially matched plan” if 20 percent or fewer of the materials were not used and as “substantially different from the plan” if more than 20 percent of the materials were not used or the segment was dropped.

The extent to which the main ideas were summarized each day was determined on the basis of the percentage of segments in which the facilitator explicitly reviewed key concepts as planned. Matching the plan required 80 percent or more of segments to have a summary of main ideas. The analysis excluded segments planned for 15 minutes or less.

The extent to which the PD made explicit links to the curriculum materials, standards, or assessments used by teachers in the study districts was determined on the basis of the cumulative total time spent per day on these links. The Institute and Seminar Implementation Form and coding guide asked observers to indicate whether no time, less than 5 minutes, between 5 and 15 minutes, between 15 and 30 minutes, or more than 30 minutes were spent making such links during the day.

The extent of teacher engagement was reported on the basis of the percentage of teachers actively engaged. The form had five possible responses: 20 percent or less, 40 percent, 60 percent, 80 percent, or 100 percent of participating teachers were actively engaged for the majority of the day. Observers received a coding guide and were trained in its use. Observers were to record teacher engagement at least four times across the day. Teachers were to be counted as actively engaged if they were watching the facilitator, working problems, or listening to or contributing to the discussion. To be actively engaged teachers did not need to be enthusiastic, just attentive.

Coaching

We now move to an examination of each PD provider’s implementation of the coaching component of the PD program. As noted above, the Pearson Achievement Solutions coaching plan relied heavily on the group delivery format in both years, and this allowed more hours of coaching per teacher. The America’s Choice plan for the first year emphasized individual teacher coaching, except for the final session in which pairs of teachers were coached. As shown in table C-4, the America’s Choice facilitators reported an average of 3.1 hours per teacher per first-year coaching visit and 5.3 hours per teacher per second-year coaching visit, when their coaching format was more group oriented. Pearson Achievement Solutions facilitators reported providing an average of 5.5 hours of coaching per teacher per visit in the first year and 5.2 hours per teacher per visit in the second year.

Table C-4. Coaching—Percent of Intended Time Implemented and Mean Actual Hours per Teacher per Visit, by PD Provider: Second-Year Teacher Impact Analysis Sample

Coaching Time per Visit	America’s Choice			Pearson Achievement Solutions		
	Percent of Intended Hours Implemented	Mean Actual Hours	S.D.	Percent of Intended Hours Implemented	Mean Actual Hours	S.D.
First Year						
Hours Coached per Teacher per Visit	78.5	3.1	1.88	136.8	5.5	2.33
Second Year						
Hours Coached per Teacher per Visit	132.9	5.3	2.28	130.9	5.2	2.27

Sample Size: N = 85 two-day coaching visits for America’s Choice; 40 two-day coaching visits for Pearson Achievement Solutions available to teachers in the second-year impact sample during the first year; 112 two-day coaching visits for America’s Choice; 68 for Pearson Achievement Solutions available to teachers in the second-year impact sample during the second year.

SOURCE: 2007–2008 Coach Log; 2008–2009 Coach Log.

Features of the two-day coaching visits, as reported by the coaches, are described in Table C-5.¹¹⁹ Separate results for each PD provider are presented in Table C-6.

On average, coaches covered topics in rational numbers in 81 percent and 93 percent of the first-year and second-year coaching visits, respectively. For each two-day coaching visit, teachers received an average of 3.4 hours of rational number content and 1.0 hour of other mathematical content during the first year. During the second year, the average hours per visit were 4.7 for rational number content and 0.9 hour for other mathematical content.

America’s Choice, which planned to adapt the coaching to whatever topics the teachers were teaching at the time of the coaching visit, focused on rational number content during 72 percent of first-year visits and 89 percent of second-year visits. Pearson Achievement Solutions, which focused coaching visits on specially developed rational number lessons that facilitators asked teachers to insert into the curriculum, reported that 100 percent of visits covered rational number content in both years.

¹¹⁹ The total number of two-day visits included in this analysis was 106 for the first year and 163 for the second year. Only visits in which teachers from the second-year impact sample participated enter the analysis.

Overall, “Common student misunderstandings” and “Using representations” were the most common pedagogical foci, featured in 82 percent and 80 percent of the first-year coaching visits and 90 percent and 88 percent of the second-year coaching visits, respectively.

The coaching was delivered using a mix of individual and group formats. One-on-one coaching was used in 87 percent of first-year visits and 84 percent of second-year visits. Coaching as part of a group was included in 69 percent and 86 percent of first-year and second-year visits, respectively.

On average, debriefing after a lesson, planning lessons, and observing teachers’ instruction were the most common coaching activities used in both first-year and second-year coaching; each of these activities featured in more than 75 percent of the visits. In contrast, modeling activities that involved the coach instructing students while the teacher observed or co-taught were used in 58 percent of the first-year coaching visits and 31 percent of the second-year coaching visits. America’s Choice included modeling activities in 79 percent of first-year visits and 46 percent of second-year visits, and Pearson Achievement Solutions included such activities in 14 percent of first-year visits and 7 percent of second-year visits.

Table C-5. Percent of Coaching Visits With Specified Features and Time Spent in Coaching With These Features: Second-Year Teacher Impact Analysis Sample

	Percent of Coaching Visits Covering Focus	Hours per Teacher per Visit	
		Mean	S.D
First Year			
Content Focus			
Rational numbers	81.1	3.4	2.54
Other mathematical focus	37.7	1.0	1.53
No mathematical focus	15.1	0.2	0.68
Pedagogical Focus			
Precise language	51.9	0.5 ^a	0.66
Using representations	80.2	1.3 ^a	1.06
Correcting teacher mathematics	18.9	0.1 ^a	0.28
Connections among mathematics concepts	68.9	0.8 ^a	0.86
Common student misunderstandings	82.1	1.0 ^a	0.80
Other focus	43.4	0.8 ^a	1.33
Delivery Format			
One-on-one coaching	86.8	2.5 ^a	1.77
Coached as part of a group	68.9	2.0 ^a	1.78
Activities			
Planning	80.2	0.8 ^c	0.60
Observing	76.4	1.7 ^c	1.41
Instructing	57.5	0.8 ^c	0.88
Debriefing	90.6	1.2 ^c	0.80
Second Year			
Content Focus			
Rational numbers	93.3	4.7	2.16
Other mathematical focus	30.7	0.9	1.65
No mathematical focus	14.7	0.2	0.71
Pedagogical Focus			
Precise language	69.3	0.9 ^b	0.83
Using representations	88.3	1.4 ^b	0.82
Correcting teacher mathematics	22.7	0.1 ^b	0.36
Connections among mathematics concepts	83.4	1.0 ^b	0.74
Common student misunderstandings	90.2	1.6 ^b	0.99
Other focus	27.6	0.7 ^b	1.63
Delivery Format			
One-on-one coaching	84.0	2.4 ^b	1.93
Coached as part of a group	85.9	3.5 ^b	2.31

Table continues on next page

Table C-5. Percent of Coaching Visits With Specified Features and Time Spent in Coaching With These Features: Second-Year Teacher Impact Analysis Sample (continued)

	Percent of Coaching Visits Covering Focus	Hours per Teacher per Visit	
		Mean	S.D.
Activities			
Planning	87.1	1.2 ^c	1.15
Observing	90.2	2.0 ^c	1.16
Instructing	31.3	0.5 ^c	0.99
Debriefing	96.3	1.6 ^c	0.76

Sample Size: N = 106 two-day coaching visits attended by second-year impact sample teachers in the first year; 163 two-day coaching visits attended by second-year impact sample teachers in the second year.

SOURCE: 2007–2008 Coach Log; 2008–2009 Coach Log.

NOTES: Each two-day coaching visit consisted of multiple sessions (interactions between a coach and an individual teacher or group of teachers). Hours per content focus, pedagogical focus, delivery format, and activity were determined within each session and then aggregated for each two-day coaching visit. For individual sessions that covered multiple content areas, pedagogical foci, or activities, the duration of those sessions was divided by the number of content areas, pedagogical foci, or activities covered, allocating the time equally. Sessions involving multiple delivery formats did not occur.

^aNumbers do not sum to 4.6 hours total owing to rounding. In the first year, on average, teachers in the second-year impact sample spent 4.6 hours per two-day coaching visit attended.

^bNumbers do not sum to 5.8 hours total owing to rounding. In the second year, on average, teachers in the second-year impact sample spent 5.8 hours per two-day coaching visit attended.

^cNumbers do not sum to 4.6 hours and 5.8 hours in the first year and second year, respectively, owing to “other” or unspecified activities, which are not reported in the table. “Other” activities include reviewing or reteaching mathematics content from previous PD days, examining the textbook and/or district pacing guide in depth, examining student work and/or assessment results in depth, and conducting follow-up phone calls or providing written feedback to teachers.

Table C-6. Percent of Coaching Visits With Specified Features and Time Spent in Coaching With These Features, by PD Provider: Second-Year Teacher Impact Analysis Sample

	America's Choice			Pearson Achievement Solutions		
	Percent of Coaching Visits Covering Focus	Hours per Teacher per Visit		Percent of Coaching Visits Covering Focus	Hours per Teacher per Visit	
		Mean	S.D.		Mean	S.D.
First Year						
Content Focus						
Rational numbers	71.8	2.1	1.93	100.0	6.0	1.33
Other mathematical focus	56.3	1.5	1.67	0.0	0.0	0.00
No mathematical focus	15.5	0.2	0.58	14.3	0.3	0.86
Pedagogical Focus						
Precise language	49.3	0.4 ^a	0.59	57.1	0.7 ^b	0.74
Using representations	73.2	1.0 ^a	0.99	94.3	2.0 ^b	0.88
Correcting teacher mathematics	25.4	0.1 ^a	0.30	5.7	< 0.01 ^b	0.20
Connections among mathematical concepts	53.5	0.4 ^a	0.49	100.0	1.7 ^b	0.78
Common student misunderstandings	77.5	0.7 ^a	0.85	91.4	1.6 ^b	0.97
Other focus	57.7	1.1 ^a	1.44	14.3	0.2 ^b	0.77
Delivery Format						
One-on-one coaching	80.3	1.9 ^a	1.52	100.0	3.8	1.58
Coached as part of a group	57.7	1.8 ^a	2.02	91.4	2.5	1.06
Activities						
Planning	76.1	0.6 ^c	0.54	88.6	1.0 ^c	0.66
Observing	66.2	1.0 ^c	0.99	97.1	3.1 ^c	1.07
Instructing	78.9	1.1 ^c	0.87	14.3	0.1 ^c	0.26
Debriefing	85.9	0.8 ^c	0.54	100.0	2.0 ^c	0.60
Second Year						
Content Focus						
Rational numbers	89.2	4.0	2.25	100.0	5.8	1.46
Other mathematical focus	47.1	1.4	1.90	3.3	< 0.01	0.12
No mathematical focus	21.6	0.4	0.87	3.3	< 0.01	0.11
Pedagogical Focus						
Precise language	64.7	1.0	0.89	77.0	0.8	0.72
Using representations	83.3	1.5	0.92	96.7	1.4	0.62
Correcting teacher mathematics	29.4	0.2	0.40	11.5	0.1	0.25
Connections among mathematical concepts	79.4	0.8	0.73	90.2	1.2	0.69
Common student misunderstandings	86.3	1.5	1.08	96.7	1.7	0.80
Other focus	25.5	0.8	1.76	31.1	0.6	1.40

Table continues on next page

Table C-6. Percent of Coaching Visits With Specified Features and Time Spent in Coaching With These Features, by PD Provider: Second-Year Teacher Impact Analysis Sample (continued)

	America's Choice			Pearson Achievement Solutions		
	Percent of Coaching Visits Covering Focus	Hours per Teacher per Visit		Percent of Coaching Visits Covering Focus	Hours per Teacher per Visit	
		Mean	S.D.		Mean	S.D.
Delivery Format						
One-on-one coaching	74.5	1.5	1.39	100.0	3.8	1.88
Coached as part of a group	94.1	4.3	2.32	72.1	2.0	1.41
Activities						
Planning	83.3	1.3 ^c	1.40	93.4	0.9 ^c	0.47
Observing	86.3	1.5 ^c	0.87	96.7	2.9 ^c	1.04
Instructing	46.1	0.8 ^c	1.14	6.6	0.1 ^c	0.42
Debriefing	97.1	1.6 ^c	0.79	95.1	1.5 ^c	0.72

Sample Size: N = 173 two-day coaching visits attended by second-year impact sample teachers for America's Choice (71 in the first year; 102 in the second year); 96 two-day coaching visits attended by second-year impact sample teachers for Pearson Achievement Solutions (35 in the first year; 61 in the second year).

SOURCE: 2007–2008 Coach Log; 2008–2009 Coach Log.

NOTES: Each two-day coaching visit consisted of multiple sessions (interactions between a coach and an individual teacher or group of teachers). Hours per content focus, pedagogical focus, delivery format, and activity were determined within each session and then aggregated for each two-day coaching visit. For individual sessions that covered multiple content areas, pedagogical foci, or activities, the duration of those sessions was divided by the number of content areas, pedagogical foci, or activities covered, allocating the time equally. Sessions involving multiple delivery formats did not occur.

^a Numbers do not sum to 3.8 hours total owing to rounding. In the first year, on average, teachers in the second-year impact sample served by America's Choice spent 3.8 hours per two-day coaching visit attended.

^b Numbers do not sum to 6.3 hours total owing to rounding. In the first year, on average, teachers in the second-year impact sample served by Pearson Achievement Solutions spent 6.3 hours per two-day coaching visit attended.

^c Numbers do not sum to 3.8 hours and 6.3 hours in the first year and 5.8 hours and 5.8 hours in the second year, respectively, owing to "other" or unspecified activities, which are not reported in the table. "Other" activities include reviewing or reteaching mathematics content from previous PD days, examining the textbook and/or district pacing guide in depth, examining student work and/or assessment results in depth, and conducting follow-up phone calls or providing written feedback to teachers.

Teacher Participation in the PD Program by Provider and Date of Entry

This section presents PD participation results separately for each PD provider and are based on teachers' date of entry into the study. Tables C-7a and C-7b show the hours of study-provided PD attended for the subgroup of districts served by America's Choice and for the subgroup of districts served by Pearson Achievement Solutions, respectively.

Table C-7a. Percent of Implemented PD Hours Attended by the Average Teacher: Second-Year Teacher Impact Analysis Sample—America’s Choice

PD Type (Implemented Hours)	Percent of Implemented PD Hours Attended by the Average Treatment Teacher
First Year	
All PD (63 hours)	55.9
Institute (18 hours) ^a	56.5
Seminars (27 hours)	56.1
Coaching (18 hours)	57.5
Second Year	
All PD (52 hours)	79.9
Institute (12 hours)	57.0
Seminars (17 hours)	79.3
Coaching (23 hours)	93.3
Total (First and Second Years)	
All PD (115 hours)	66.5
Institute (30 hours)	56.6
Seminars (44 hours)	65.1
Coaching (41 hours)	76.9

Sample Size: N = 11 schools; 28 teachers.

SOURCE: 2007–2008 Participation Form; 2008–2009 Participation Form; 2007–2008 Institute and Seminar Implementation Form; 2008–2009 Institute and Seminar Implementation Form; 2007–2008 Coach Log; 2008–2009 Coach Log.

NOTES: For each district, the mean total number of hours that program teachers were coached was used in the denominator when calculating the percent of implemented hours of PD attended by treatment teachers.

The row headings contain, in parentheses, the weighted average actual number of hours implemented of each type of PD across the districts. Districts are weighted by the numbers of treatment schools.

^a Teachers who did not participate in the first-year summer institute because of absence, refusal, or entry into the program subsequent to delivery of the institute were provided a two-day makeup institute in the second year of the program. Because this makeup institute addressed content covered during the first-year institute, teachers’ participation in the makeup institute is treated as participation in first-year PD.

Table C-7b. Percent of Implemented PD Hours Attended by the Average Teacher: Second-Year Teacher Impact Analysis Sample—Pearson Achievement Solutions

PD Type (Implemented Hours)	Percent of Implemented PD Hours Attended by the Average Treatment Teacher
First Year	
All PD (72 hours)	48.7
Institute (15 hours) ^{ab}	68.9
Seminars (28 hours)	41.5
Coaching (29 hours)	43.7
Second Year	
All PD (49 hours)	87.9
Institute (11 hours)	70.6
Seminars (16 hours)	90.3
Coaching (22 hours)	95.2
Total (First and Second Years)	
All PD (121 hours)	65.1
Institute (26 hours)	68.8
Seminars (44 hours)	59.2
Coaching (51 hours)	68.5

Sample Size: N = 9 schools; 17 teachers.

SOURCE: 2007–2008 Participation Form; 2008–2009 Participation Form; 2007–2008 Institute and Seminar Implementation Form; 2008–2009 Institute and Seminar Implementation Form; 2007–2008 Coach Log; 2008–2009 Coach Log.

NOTES: For each district, the mean total number of hours that program teachers were coached was used in the denominator when calculating the percentage of implemented hours of PD attended by treatment teachers.

The row headings contain, in parentheses, the weighted average actual number of hours implemented of each type of PD across the districts. Districts are weighted by the numbers of treatment schools.

^a Teachers who did not participate in the first-year summer institute because of absence, refusal, or entry into the program subsequent to delivery of the institute were provided a two-day makeup institute in the second year of the program. Because this makeup institute addressed content covered during the first-year institute, teachers' participation in the makeup institute is treated as participation in first-year PD.

Teacher turnover limited the average number of hours of PD participation. As discussed in Chapter 2, of the 45 treatment teachers teaching regular seventh-grade mathematics classes in spring 2009, 22 were not present and teaching that course at the beginning of the study and therefore were not eligible to receive the full dosage of 118 hours of study PD. Table C-8 shows that among treatment teachers teaching regular seventh-grade mathematics classes in spring 2009, the maximum possible PD dosage based on program entry dates was 87 hours on average. Therefore, the average treatment teacher received 89 percent of the maximum possible PD dosage.

Table C-8. Maximum Possible PD Dosage Based on Teacher PD Program Entry Dates: Second-Year Teacher Impact Analysis Sample

Teacher Group	PD Program Entry Date^a	Group Average Maximum Possible Hours of PD Dosage	Group Average Hours of PD Received	Group Average Percent of Maximum Possible PD Dosage Received^b
Entered at Beginning of PD Program Full Dosage (N = 21)	4/1/2007 to 8/9/2007	118.0	102.2	86.6
Entered at Start of or During First School Year Missed first summer institute and/or seminars (N = 5)	8/7/2007 to 2/1/2008	87.4	77.6	88.8
Entered at Start of Second School Year Missed first summer institute, seminars, and coaching (N = 13)	6/30/2008 to 8/25/2008	62.0	52.4	84.5
Entered During Second School Year Missed all first-year PD and second-year summer institute and/or seminars (N = 6)	8/18/2008 to 1/5/2009	38.2	43.1	112.8
Unweighted Average (N = 45)		87.8	77.2	87.9
Weighted Average (N = 45)		87.1	77.2	88.7

Sample Size: N = 20 treatment schools; 45 treatment teachers.

SOURCE: PD Participation Records.

NOTES: ^a Program entry dates for groups overlap because start dates for districts varied.

^b The percentage of maximum possible PD dosage received can exceed 100 percent because actual coaching hours could exceed the intended dose.

^c The weighted average is obtained by weighting the average dosage of a district's set of teachers by the number of treatment schools in the district.

Supplemental Data on Service Contrast

This final section presents supplemental information on the treatment and control group service contrasts. Tables C-9a and C-9b present the service contrasts in hours of mathematics-related PD, separately for each PD provider. (Table C-9a describes the service contrast for teachers in districts served by America's Choice, and Table C-9b describes the service contrast for teachers in districts served by Pearson Achievement Solutions.)

Table C-9a. Treatment and Control Group Contrast in Hours of Mathematics-Related PD, for Districts Served by America’s Choice: Second-Year Teacher Impact Analysis Sample

Type of Mathematics-Related PD	Treatment Group Weighted	Control Group Weighted	Estimated Difference	Standard Error of the Estimated Difference	Estimated Difference Effect Size	P-value
First Year						
Summer 2007						
Institutes or Seminars ^a (hours)	11.4	5.0	6.4	3.86	0.77	0.12
2007–2008 School Year						
Institutes or Seminars ^a (hours)	22.5	6.9	15.6*	6.72	1.45	0.04
Coaching (hours)	10.4	5.0	5.4	4.48	0.57	0.25
Other PD (hours)	4.0	5.9	–1.8	3.75	–0.15	0.63
Summer 2007, 2007–2008 School Year						
TOTAL PD (hours)	48.4	22.8	25.6*	10.20	0.98	0.03
Second Year						
Summer 2008						
Institutes or Seminars ^a (hours)	11.9	9.2	2.7	3.98	0.12	0.51
2008–2009 School Year						
Institutes or Seminars ^a (hours)	18.3	9.3	9.0*	4.05	0.70	0.04
Coaching (hours)	14.5	7.1	7.4*	2.78	0.70	0.02
Other PD (hours)	5.1	6.6	–1.5	4.02	–0.09	0.72
Summer 2008, 2008–2009 School Year						
TOTAL PD (hours)	49.8	32.2	17.6	9.04	0.47	0.07

Sample Size: N = 22 schools and 49 teachers (25 treatment, 24 control) for the first year; 22 schools and 54 teachers (27 treatment, 27 control) for the second year.

SOURCE: Fall 2007 Teacher Survey; Spring 2008 Teacher Survey; Fall 2008 Teacher Survey; Spring 2009 Teacher Survey.

NOTES: ^a Institutes or seminars are defined as PD sessions lasting one-half day or longer but excluding courses such as college courses that last for several weeks.

The analyses are based on a two-level model controlling for random assignment block.

Effect sizes were calculated using the control group standard deviation, combined across all two-year districts. The control group standard deviations were 8.3 for summer 2007 institutes or seminars; 10.7 for 2007–2008 institutes or seminars; 9.5 for coaching; 12.0 for other PD; and 26.1 for the total institutes, seminars, coaching and other PD during summer 2007 and school year 2007–2008. The control group standard deviations were 21.7 for summer 2008 institutes or seminars; 12.8 for 2008–2009 institutes or seminars; 10.6 for coaching; 15.7 for other PD; and 37.2 for the total institutes, seminars, coaching and other PD during summer 2008 and school year 2008–2009.

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

Table C-9b. Treatment and Control Group Contrast in Hours of Mathematics-Related PD, for Districts Served by Pearson Learning Solutions: Second-Year Teacher Impact Analysis Sample

Type of Mathematics-Related PD	Treatment Group Weighted	Control Group Weighted	Estimated Difference	Standard Error of the Estimated Difference	Estimated Difference Effect Size	P-value
First Year						
Summer 2007						
Institutes or Seminars ^a (hours)	14.8	-0.1	14.9	6.98	1.80	0.06
2007–2008 School Year						
Institutes or Seminars ^a (hours)	25.6	1.9	23.7	12.20	2.21	0.09
Coaching (hours)	10.8	4.0	6.8	5.15	0.71	0.23
Other PD (hours)	5.9	-0.4	6.3	4.17	0.52	0.17
Summer 2007, 2007–2008 School Year						
TOTAL PD (hours)	57.1	4.5	52.6*	18.05	2.02	0.02
Second Year						
Summer 2008						
Institutes or Seminars ^a (hours)	16.2	8.7	7.5	14.90	0.35	0.63
2008–2009 School Year						
Institutes or Seminars ^a (hours)	32.6	3.3	29.3*	7.97	2.29	0.01
Coaching (hours)	14.5	0.0	14.5*	5.47	1.37	0.03
Other PD (hours)	6.7	12.1	-5.4	5.68	-0.34	0.37
Summer 2008, 2008–2009 School Year						
TOTAL PD (hours)	70.0	24.0	46.0	24.81	1.24	0.10

Sample Size: N = 16 schools and 34 teachers (15 treatment, 19 control) for the first year; 16 schools and 35 teachers (16 treatment, 19 control) for the second year.

SOURCE: Fall 2007 Teacher Survey; Spring 2008 Teacher Survey; Fall 2008 Teacher Survey; Spring 2009 Teacher Survey.

NOTES: ^a Institutes or seminars are defined as PD sessions lasting one-half day or longer but excluding courses such as college courses that last for several weeks.

The analyses are based on a two-level model controlling for random assignment block.

Effect sizes were calculated using the control group standard deviation, combined across all two-year districts. The control group standard deviations were 8.3 for summer 2007 institutes or seminars; 10.7 for 2007–2008 institutes or seminars; 9.5 for coaching; 12.0 for other PD; and 26.1 for the total institutes, seminars, coaching and other PD during summer 2007 and school year 2007–2008. The control group standard deviations were 21.7 for summer 2008 institutes or seminars; 12.8 for 2008–2009 institutes or seminars; 10.6 for coaching; 15.7 for other PD; and 37.2 for the total institutes, seminars, coaching and other PD during summer 2008 and school year 2008–2009.

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

Table C-10 presents the treatment and control group contrasts for other features of the PD that the teachers experienced. These features include type of content emphasis, type of pedagogical emphasis, active participation, collective participation, relevance to own teaching, and clarity of

purpose. The scales that measure these features were derived through factor analysis from a series of 4-point Likert-scale items on the teacher surveys.

Table C-10. Treatment and Control Group Contrasts for Features of Mathematics-Related PD: Second-Year Teacher Impact Analysis Sample (unstandardized)

PD Features	Sample Size (N)	Treatment Group Mean	Control Group Mean	Estimated Difference	Standard Error of the Estimated Difference	P-value
First Year						
Summer 2007						
Content Emphasis						
Fractions, Decimals	26	2.85	3.35	-0.50	0.49	0.35
Percent, Ratio, Rate, Proportion	27	2.60	3.32	-0.72	0.55	0.24
Whole Numbers/Integers, Algebra, Geometry, Probability, and Statistics	28	1.54	2.80	-1.26*	0.39	0.02
Pedagogical Emphasis						
Pedagogical Topics Intervened Upon	26	2.64	3.26	-0.62*	0.20	0.03
Pedagogical Topics Not Intervened Upon	28	1.60	2.51	-0.91	0.42	0.07
Active Participation	18	1.98	2.20	-0.22	0.48	0.65
Collective Participation	19	2.67	2.72	-0.06	0.41	0.90
Relevance to My Teaching	19	3.50	3.30	0.19	0.37	0.62
Clarity of Purpose	19	3.52	3.54	-0.02	0.43	0.97
2007–2008 School Year						
Content Emphasis						
Fractions, Decimals	60	2.82	2.11	0.71	0.38	0.09
Percent, Ratio, Rate, Proportion	59	2.93	2.19	0.74	0.37	0.06
Whole Numbers/Integers, Algebra, Geometry, Probability, and Statistics	60	2.03	2.35	-0.33	0.29	0.28
Pedagogical Emphasis						
Pedagogical Topics Intervened Upon	61	3.08	2.55	0.53*	0.23	0.04
Pedagogical Topics Not Intervened Upon	61	2.10	2.46	-0.36	0.20	0.10
Active Participation	46	2.53	2.72	-0.19	0.29	0.53
Collective Participation	46	2.72	2.42	0.29	0.22	0.21
Relevance to My Teaching	46	3.50	3.47	0.03	0.19	0.89
Clarity of Purpose	46	3.26	3.12	0.14	0.27	0.61

Table continues on next page

Table C-10. Treatment and Control Group Contrasts for Features of Mathematics-Related PD: Second-Year Teacher Impact Analysis Sample (unstandardized) (continued)

PD Features	Sample Size (N)	Treatment Group Mean	Control Group Mean	Estimated Difference	Standard Error of the Estimated Difference	P-value
Second Year						
Summer 2008						
Content Emphasis						
Fractions, Decimals	36	3.18	2.28	0.90*	0.29	0.01
Percent, Ratio, Rate, Proportion	35	3.33	2.31	1.02*	0.32	0.01
Whole Numbers/Integers, Algebra, Geometry, Probability, and Statistics	36	1.83	2.41	-0.58	0.33	0.10
Pedagogical Emphasis						
Pedagogical Topics Intervened Upon	36	3.22	2.78	0.44	0.28	0.13
Pedagogical Topics Not Intervened Upon	36	2.03	2.09	-0.06	0.19	0.77
Active Participation	34	2.94	2.23	0.71*	0.28	0.02
Collective Participation	34	2.70	2.10	0.60*	0.21	0.01
Relevance to My Teaching	33	3.73	3.21	0.53*	0.17	0.01
Clarity of Purpose	33	3.88	3.26	0.63*	0.21	0.01
2008–2009 School Year						
Content Emphasis						
Fractions, Decimals	73	3.22	1.55	1.67*	0.27	<.01
Percent, Ratio, Rate, Proportion	73	3.32	1.64	1.68*	0.24	<.01
Whole Numbers/Integers, Algebra, Geometry, Probability, and Statistics	72	2.08	1.54	0.54*	0.23	0.03
Pedagogical Emphasis						
Pedagogical Topics Intervened Upon	73	3.29	2.35	0.93*	0.18	<.01
Pedagogical Topics Not Intervened Upon	73	2.24	2.26	-0.01	0.19	0.94
Active Participation	50	3.10	2.38	0.72*	0.26	0.02
Collective Participation	49	2.60	2.45	0.15	0.30	0.62
Relevance to My Teaching	50	3.60	3.12	0.47	0.24	0.07
Clarity of Purpose	50	3.67	3.17	0.50	0.26	0.08
Plan-Observe-Debrief Coaching Cycle	55	3.33	2.86	0.48*	0.21	0.04
Observing Coaches and Other Teachers	55	2.52	2.22	0.30	0.30	0.33

SOURCE: Fall 2007 Teacher Survey; Spring 2008 Teacher Survey; Fall 2008 Teacher Survey; Spring 2009 Teacher Survey.

NOTES: Sample sizes vary by feature within time period because some survey items were asked only of teachers who reported participating, during that time period, in PD sessions lasting longer than a half-day or in coaching.

The analyses are based on a two-level model controlling for random assignment block.

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

APPENDIX D
SUPPORTING TABLES AND FIGURES FOR
IMPACT ANALYSES

APPENDIX D

SUPPORTING TABLES AND FIGURES

FOR IMPACT ANALYSES

This appendix supplements the presentation of the study's main impact analyses in Chapter 3. The first section provides tests of the equivalence of treatment and control groups for the first-year impact analysis samples discussed in Chapter 3 and this appendix. The second section provides first-year impact results for the teachers and students in districts that did not participate in the second year of the study. The third section includes impact analyses using alternative models to test the robustness of the findings in Chapter 3. The fourth section displays the variation in impacts across districts, and the final section provides unadjusted means and standard deviations to supplement the impact analyses presented in Chapter 3.

Equivalence of Treatment and Control Group Characteristics

Tables D-1 through D-4 display the results of tests for the equivalence of treatment and control group participants in the first-year impact analysis samples, where equivalence is based on selected characteristics of participants at the point of entry into the study and, for teachers, the percentage of teachers in each group that entered the study in fall 2007. Tables D-1 and D-2 display the equivalence of the first-year impact analysis samples in districts that participated in both years of the study (two-year districts). Tables D-3 and D-4 show the findings for the first-year impact analysis samples in districts that participated only in the first year of the study (one-year districts). As discussed in Chapter 3, in all cases, we conducted a chi-square test of all teacher-level variables and all student-level variables for each impact analysis sample. These tests indicate that there were no overall differences in the characteristics of the treatment and control groups in any of the samples.

Table D-1. Teacher Characteristics, by Treatment Status: First-Year Teacher Impact Analysis Sample—Two-Year Districts

Teacher Characteristics	Treatment Group	Control Group	Estimated Difference	P-value for Estimated Difference
Baseline Teacher Knowledge ^a				
Total Score (logits)	0.07	0.19	-0.12	0.61
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>51.7</i>	<i>54.7</i>	<i>-3.0</i>	
Common Knowledge of Mathematics (CK) Score (logits)	0.22	0.45	-0.24	0.45
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>58.4</i>	<i>64.1</i>	<i>-5.6</i>	
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.04	0.11	-0.07	0.78
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>47.8</i>	<i>49.5</i>	<i>-1.8</i>	
Years of Teaching Experience (percent)				
3 years or fewer	28.9	38.1	-9.3	0.41
4–10 years	31.3	33.3	-2.0	0.86
More than 10 years	39.8	28.4	11.4	0.27
Years of Teaching Experience in Middle School Mathematics	6.9	6.2	0.7	0.63
Educational Level: M.A. and Above (percent)	53.9	41.5	12.4	0.28
Mathematics Major (percent)	17.0	20.7	-3.7	0.67
Number of Postsecondary Mathematics Courses Taken	6.9	7.8	-0.9	0.19
Number of Postsecondary Mathematics Education Courses Taken	1.9	2.1	-0.3	0.38
Teachers Who Entered the Study in Fall 2007 (percent) ^b	86.9	89.9	-3.0	0.69

Sample Size: N = 89 teachers (41 treatment, 48 control).

SOURCE: Fall 2007 Teacher Knowledge Test; Teacher Survey.

NOTES: ^a Sample Size: N = 82 teachers (36 treatment, 46 control).

^b Sample Size: N = 90 teachers (41 treatment, 49 control).

Baseline teacher knowledge was assessed at the beginning of the teacher's first year in the study. For treatment group teachers, teacher knowledge was assessed prior to the summer PD (or prior to the teacher's first seminar day if the teacher missed the summer PD but entered the study within the first 10 weeks of the school year). For control group teachers, teacher knowledge was assessed within the first 10 weeks of the school year. Teachers who entered the study later in the school year were not tested for baseline knowledge.

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.

Educational experience items reflect the teacher's circumstances as of the teacher's entry into the study; teaching experience was calculated as the number of years of experience at the start of the 2007–2008 school year.

Values in the columns represent unadjusted means for the groups. Percentage values for characteristics with multiple categories may not sum to 100 owing to rounding.

The analyses are based on a two-level model controlling for random assignment block.

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

Table D-2. Student Characteristics, by Treatment Status: First-Year Student Impact Analysis Sample—Two-Year Districts

Student Characteristics	Treatment Group	Control Group	Estimated Difference	P-value for Estimated Difference
Age (year) ^a	12.64	12.63	0.00	0.87
Students Eligible for Free or Reduced-Price Lunch (percent)	61.84	69.63	-7.78	0.052
Race/Ethnicity (percent)				
White, Non-Hispanic	37.52	32.69	4.83	0.25
Black, Non-Hispanic	33.28	32.19	1.10	0.82
Hispanic	23.67	30.29	-6.62*	0.03
Asian/Pacific Islander	2.87	2.46	0.41	0.61
Other	2.65	2.48	0.17	0.86
Male (percent)	50.78	50.77	0.02	0.99
English as a Second Language (percent)	17.57	18.94	-1.37	0.65
Special Education Status (percent)	13.34	10.16	3.18	0.16
Sixth-Grade Mathematics Scores on State Accountability Assessment (standardized)	0.24	0.16	0.08	0.34
Fall 2007 Student Mathematics Achievement				
NWEA Total Score (scale score)	216.16	215.59	0.57	0.67
<i>Corresponding Percentile Rank</i>	23	22		
Fractions and Decimals Score (scale score)	215.71	214.78	0.94	0.54
Ratio and Proportion Score (scale score)	216.46	216.26	0.20	0.88

Sample Size: N = 2,203 students (1,094 treatment, 1,109 control).

SOURCE: Fall 2007 NWEA Rational Number Test; Study District Records.

NOTES: ^a Age was calculated as the age (in years) of a student as of September 1, 2007.

Values in the columns represent unadjusted means for the groups. Percentage values for characteristics with multiple categories may not sum to 100 due to rounding.

The analyses are based on a three-level model controlling for random assignment block.

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

Table D-3. Teacher Characteristics, by Treatment Status: First-Year Teacher Impact Analysis Sample—One-Year Districts

Teacher Characteristics	Treatment Group	Control Group	Estimated Difference	P-value for Estimated Difference
Baseline Teacher Knowledge ^a				
Total Score (logits)	-0.42	-0.14	-0.28	0.21
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>39.7</i>	<i>46.5</i>	<i>-6.8</i>	
Common Knowledge of Mathematics (CK) Score (logits)	-0.43	0.11	-0.53	0.13
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>42.5</i>	<i>55.8</i>	<i>-13.2</i>	
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	-0.28	-0.30	0.02	0.94
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>40.0</i>	<i>39.5</i>	<i>0.4</i>	
Years of Teaching Experience (percent)				
3 years or fewer	39.7	20.2	19.5*	0.04
4–10 years	24.5	34.8	-10.3	0.31
More than 10 years	35.7	43.5	-7.8	0.56
Years of Teaching Experience in Middle School Mathematics	6.3	8.8	-2.5	0.25
Educational Level: M.A. and Above (percent)	28.3	31.7	-3.4	0.72
Mathematics Major (percent)	10.3	13.3	-3.0	0.61
Number of Postsecondary Mathematics Courses Taken	5.3	5.5	-0.2	0.68
Number of Postsecondary Mathematics Education Courses Taken	1.5	2.0	-0.5*	0.02
Teachers Who Entered the Study in Fall 2007 (percent) ^b	92.8	93.2	-0.5	0.94

Sample Size: N = 101 teachers (56 treatment, 45 control).

SOURCE: Fall 2007 Teacher Knowledge Test; Teacher Survey.

NOTES: ^a Sample Size: N = N = 90 teachers (51 treatment, 39 control).

^b Sample Size: N = 101.

Baseline teacher knowledge was assessed at the beginning of the teacher's first year in the study. For treatment group teachers, teacher knowledge was assessed prior to the summer PD (or prior to the teacher's first seminar day if the teacher missed the summer PD but entered the study within the first 10 weeks of the school year). For control group teachers, teacher knowledge was assessed within the first 10 weeks of the school year. Teachers who entered the study later in the school year were not tested for baseline knowledge.

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.

Educational experience items reflect the teacher's circumstances as of the teacher's entry into the study; teaching experience was calculated as the number of years of experience at the start of the 2007–2008 school year.

Values in the columns represent unadjusted means for the groups. Percentage values for characteristics with multiple categories may not sum to 100 owing to rounding.

The analyses are based on a two-level model controlling for random assignment block.

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

Table D-4. Student Characteristics, by Treatment Status: First-Year Student Impact Analysis Sample—One-Year Districts

Student Characteristics	Treatment Group	Control Group	Estimated Difference	P-value for Estimated Difference
Age (year) ^a	12.84	12.82	0.02	0.59
Students Eligible for Free or Reduced-Price Lunch (percent)	69.88	72.56	-2.68	0.59
Race/Ethnicity (percent)				
White, Non-Hispanic	26.29	29.65	-3.36	0.48
Black, Non-Hispanic	41.38	41.02	0.36	0.94
Hispanic	27.09	25.82	1.27	0.84
Asian/Pacific Islander	2.07	1.92	0.16	0.81
Other	3.17	1.54	1.63	0.07
Male (percent)	49.22	50.17	-0.95	0.68
English as a Second Language (percent)	10.16	6.53	3.63	0.19
Special Education Status (percent)	7.88	7.05	0.83	0.65
Sixth-Grade Mathematics Scores on State Accountability Assessment (standardized)	0.06	0.05	0.01	0.93
Fall 2007 Student Mathematics Achievement				
NWEA Total Score (scale score)	214.08	213.05	1.03	0.35
<i>Corresponding Percentile Rank</i>	<i>19</i>	<i>18</i>		
Fractions and Decimals Score (scale score)	212.74	211.66	1.08	0.34
Ratio and Proportion Score (scale score)	215.25	214.28	0.97	0.39

Sample Size: N = 2,325 students (1,242 treatment, 1,083 control).

SOURCE: Fall 2007 NWEA Rational Number Test; Study District Records.

NOTES: ^a Age was calculated as the age (in years) of a student as of September 1, 2007.

Values in the columns represent unadjusted means for the groups. Percentage values for characteristics with multiple categories may not sum to 100 owing to rounding.

The analyses are based on a three-level model controlling for random assignment block.

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

First-Year Impacts for One-Year Districts

In Chapter 3, we discussed the first-year impact findings for the teachers and students in districts that participated in the second year of the study. As with our findings for the full set of 12 districts, we did not find statistically significant impacts on either teacher knowledge or student achievement. Tables D-5 and D-6 display the first-year findings for districts that were *not* included in the second-year implementation (one-year districts). For teachers in these districts, there were statistically significant impacts on teachers' *Total Score* and *Specialized Knowledge of Mathematics for Teaching (SK) Score*. There were no statistically significant impacts on student achievement.

Table D-5. Impact of the PD Program on Teacher Knowledge at the End of the First Year: First-Year Teacher Impact Analysis Sample—One-Year Districts

Outcome Measure	Treatment Group	Control Group	Estimated Impact	Standard Error of the Estimated Impact	Estimated Impact Effect Size	P-value for the Estimated Impact
Total Score (logits)	-0.03	-0.40	0.37*	0.16	0.38	0.03
<i>Percent correctly answering items of average difficulty for the test instrument</i>	49.2	40.1	9.1			
Common Knowledge of Mathematics (CK) Score (logits)	-0.12	-0.21	0.09	0.23	0.07	0.69
<i>Percent correctly answering items of average difficulty for the test instrument</i>	50.1	47.8	2.3			
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.16	-0.41	0.57*	0.23	0.50	0.02
<i>Percent correctly answering items of average difficulty for the test instrument</i>	50.9	36.9	14.0			

Sample Size: N = 38 schools (20 treatment, 18 control); 99 teachers (55 treatment, 44 control).

SOURCE: Spring 2008 Teacher Knowledge Test.

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 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 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.

Effect sizes were calculated using the control group standard deviation. The control group standard deviation was 0.97 for the *Total Score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*.

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

Table D-6. Impact of the PD Program on Student Mathematics Achievement at the End of the First Year: First-Year Student Impact Analysis Sample—One-Year Districts

Outcome Measure	Treatment Group	Control Group	Estimated Impact	Standard Error of the Estimated Impact	Estimated Impact Size	P-value for the Estimated Impact
NWEA Total Score (scale score)	215.42	214.22	1.19	0.68	0.08	0.09
<i>Corresponding Percentile Rank</i>	<i>16</i>	<i>14</i>				
Fractions and Decimals Score (scale score)	213.57	212.52	1.05	0.70	0.07	0.15
Ratio and Proportion Score (scale score)	217.19	215.93	1.26	0.76	0.08	0.11

Sample Size: N= 38 schools (20 treatment, 18 control); 2,325 students (1,242 treatment, 1,083 control).

SOURCE: Spring 2008 NWEA Rational Number Test.

NOTES: The impact analyses for student mathematics achievement were conducted using scale scores. The estimated impacts are based on a three-level model controlling for random assignment block and student-level covariates.

The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the treatment group as the basis for the adjustment.

The values for the corresponding percentile rank are derived from the treatment and control group means in scale scores.

Effect sizes were calculated using the control group standard deviation. The control group standard deviation was 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for the *Ratio and Proportion Score*.

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

Robustness Checks for Impact Estimates

In Chapter 3, we presented impact estimates and group means based on models that adjusted for student and teacher or classroom characteristics. Tables D-7 and D-8 present the impact estimates and group means for teacher knowledge and student achievement without controlling for any covariates other than the random assignment blocks. Table D-9 presents the student achievement results achieved by using the same basic model that was employed in the student analysis in Chapter 3 but also incorporating the teacher-level covariates that were included in the teacher outcome models. These teacher covariates include the baseline teacher knowledge total scores, total teacher experience, teaching experience in middle school mathematics, teacher’s education level (master’s degree or not), mathematics major (or not), the number of postsecondary mathematics courses taken by the teacher, the average class size from class rosters, and the teacher’s years of experience with the current curriculum, as recorded on the Teacher Survey. Table D-10 presents student achievement results using teacher instead of classroom as the middle level of the multilevel model. As noted in Chapter 3, the results of the robustness checks were consistent with the results of the main impact analyses in all cases.

Table D-7. Impact of the PD Program on Teacher Knowledge at the End of the Second Year, Without Covariates: Second-Year Teacher Impact Analysis Sample

Outcome Measure	Treatment Group	Control Group	Estimated Impact	Standard Error of the Estimated Impact	Estimated Impact Effect Size	P-value for the Estimated Impact
Teacher Knowledge						
Total Score (logits)	1.13	1.14	-0.00	0.20	-0.00	1.00
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>75.7</i>	<i>75.7</i>	<i>0.0</i>			
Common Knowledge of Mathematics (CK) Score (logits)	1.26	1.51	-0.25	0.26	-0.19	0.34
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>79.9</i>	<i>83.7</i>	<i>-3.7</i>			
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.78	0.52	0.26	0.22	0.22	0.26
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>65.8</i>	<i>59.8</i>	<i>6.0</i>			

Sample Size: N = 38 schools (20 treatment, 18 control); 89 teachers (43 treatment, 46 control).

SOURCE: Spring 2009 Teacher Knowledge Test.

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 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 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.

Effect sizes were calculated using the control group standard deviation for the first-year teacher impact analysis sample. The control group standard deviation was 0.97 for the *Total Score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*.

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

Table D-8. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year, Without Covariates: Second-Year Student Impact Analysis Sample

Outcome Measure	Treatment Group	Control Group	Estimated Impact	Standard Error of the Estimated Impact	Estimated Impact Effect Size	P-value for the Estimated Impact
NWEA Total Score (scale score)	219.90	218.32	1.58	1.63	0.11	0.34
<i>Corresponding Percentile Rank</i>	<i>21</i>	<i>21</i>				
Fractions and Decimals Score (scale score)	218.15	216.70	1.45	1.69	0.10	0.40
Ratio and Proportion Score (scale score)	221.71	219.88	1.83	1.65	0.12	0.28

Sample Size: N = 39 schools (20 treatment, 19 control); 2,132 students (1,083 treatment, 1,049 control).

SOURCE: Spring 2008 NWEA Rational Number Test.

NOTES: The impact analyses for student mathematics achievement were conducted using scale scores. The estimated impacts are based on a three-level model controlling for random assignment block and student-level covariates.

The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the treatment group as the basis for the adjustment.

The values for the corresponding percentile rank are derived from the treatment and control group means in scale scores.

Effect sizes were calculated using the control group standard deviation from the first-year student impact analysis sample. The control group standard deviation was 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for the *Ratio and Proportion Score*.

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

Table D-9. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year, With Teacher Covariates: Second-Year Student Impact Analysis Sample

Outcome Measure	Treatment Group	Control Group	Estimated Impact	Standard Error of the Estimated Impact	Estimated Impact Effect Size	P-value for the Estimated Impact
NWEA Total Score (scale score)	219.90	219.97	-0.07	1.04	0.00	0.95
<i>Corresponding Percentile Rank</i>	<i>22</i>	<i>22</i>				
Fractions and Decimals Score (scale score)	218.15	218.29	-0.15	1.05	-0.01	0.89
Ratio and Proportion Score (scale score)	221.71	221.64	0.07	1.15	0.00	0.95

Sample Size: N = 39 schools (20 treatment, 19 control); 2,132 students (1,083 treatment, 1,049 control).

SOURCE: Spring 2009 NWEA Rational Number Test.

NOTES: The impact analyses for student mathematics achievement were conducted using scale scores. The estimated impacts are based on a three-level model controlling for random assignment block and student-level covariates.

The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the treatment group as the basis for the adjustment.

The values for the corresponding percentile rank are derived from the treatment and control group means in scale scores.

Effect sizes were calculated using the control group standard deviation from the first-year student impact analysis sample. The control group standard deviation was 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for the *Ratio and Proportion Score*.

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

Table D-10. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year, Using Teacher as Middle Level of Multilevel Model: Second-Year Student Impact Analysis Sample

Outcome Measure	Treatment Group	Control Group	Estimated Impact	Standard Error of the Estimated Impact	Estimated Impact Size	P-value for the Estimated Impact
NWEA Total Score (scale score)	219.90	219.98	-0.08	0.90	-0.01	0.93
<i>Corresponding Percentile Rank</i>	22	22				
Fractions and Decimals Score (scale score)	218.15	218.41	-0.26	0.97	-0.02	0.79
Ratio and Proportion Score (scale score)	221.71	221.56	0.16	0.95	0.01	0.87

Sample Size: N = 39 schools (20 treatment, 19 control); 2,132 students (1,083 treatment, 1,049 control).

SOURCE: Spring 2009 NWEA Rational Number Test.

NOTES: The impact analyses for student mathematics achievement were conducted using scale scores. The estimated impacts are based on a three-level model controlling for random assignment block and student-level covariates.

The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the treatment group as the basis for the adjustment.

The values for the corresponding percentile rank are derived from the treatment and control group means in scale scores.

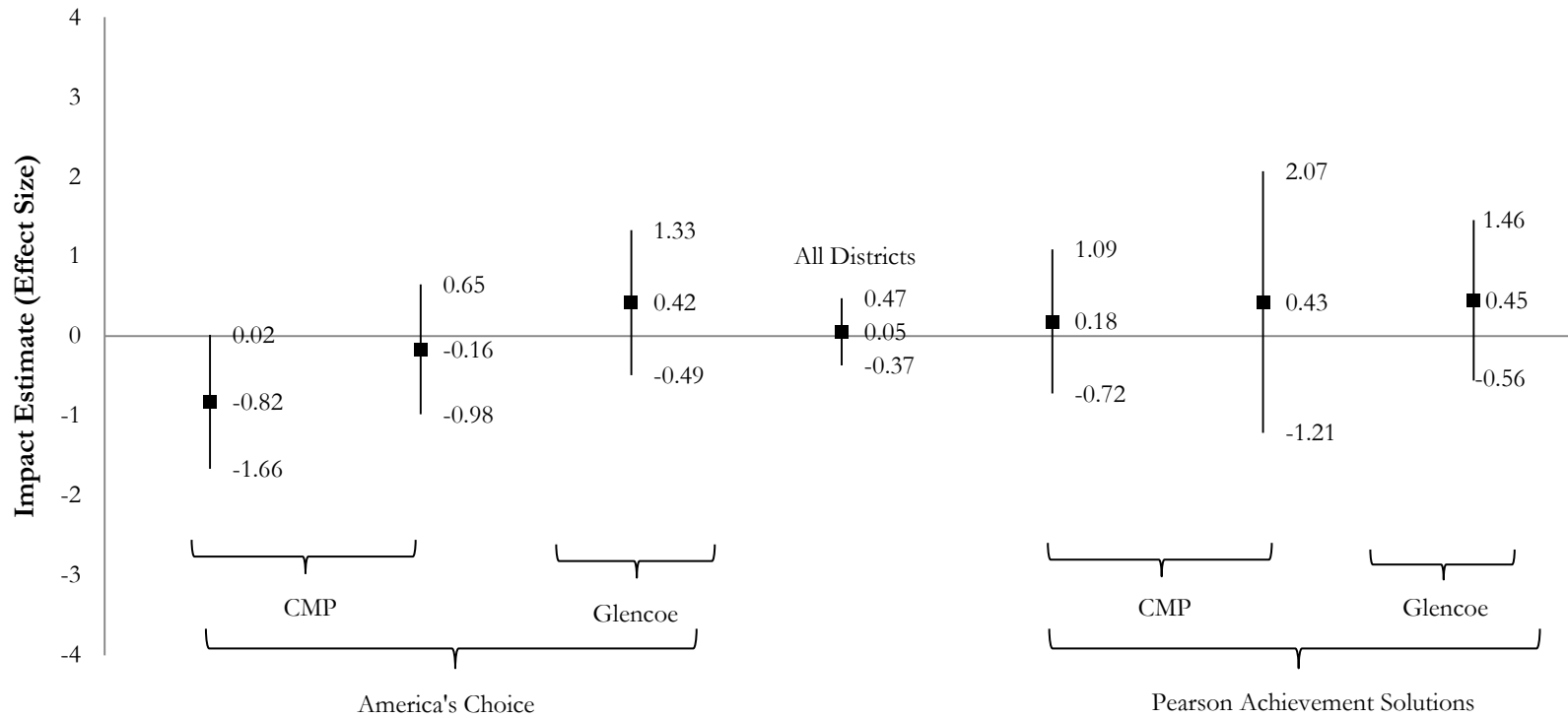
Effect sizes were calculated using the control group standard deviation from the first-year student impact analysis sample. The control group standard deviation was 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for the *Ratio and Proportion Score*.

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

Variation in the Impact of the PD Program across Districts

In the impact analyses reported in Chapter 3, the 6 two-year districts were treated as fixed effects, and separate treatment effects were estimated for each of the districts. F-tests were conducted to determine whether there was statistically significant variation in the impacts of the treatment across districts, and we found no statistically significant variation. Figures D-1 and D-2 display the estimated impacts and the upper and lower bound for the 95 percent confidence interval, by district, for each of the primary teacher and student outcome measures.

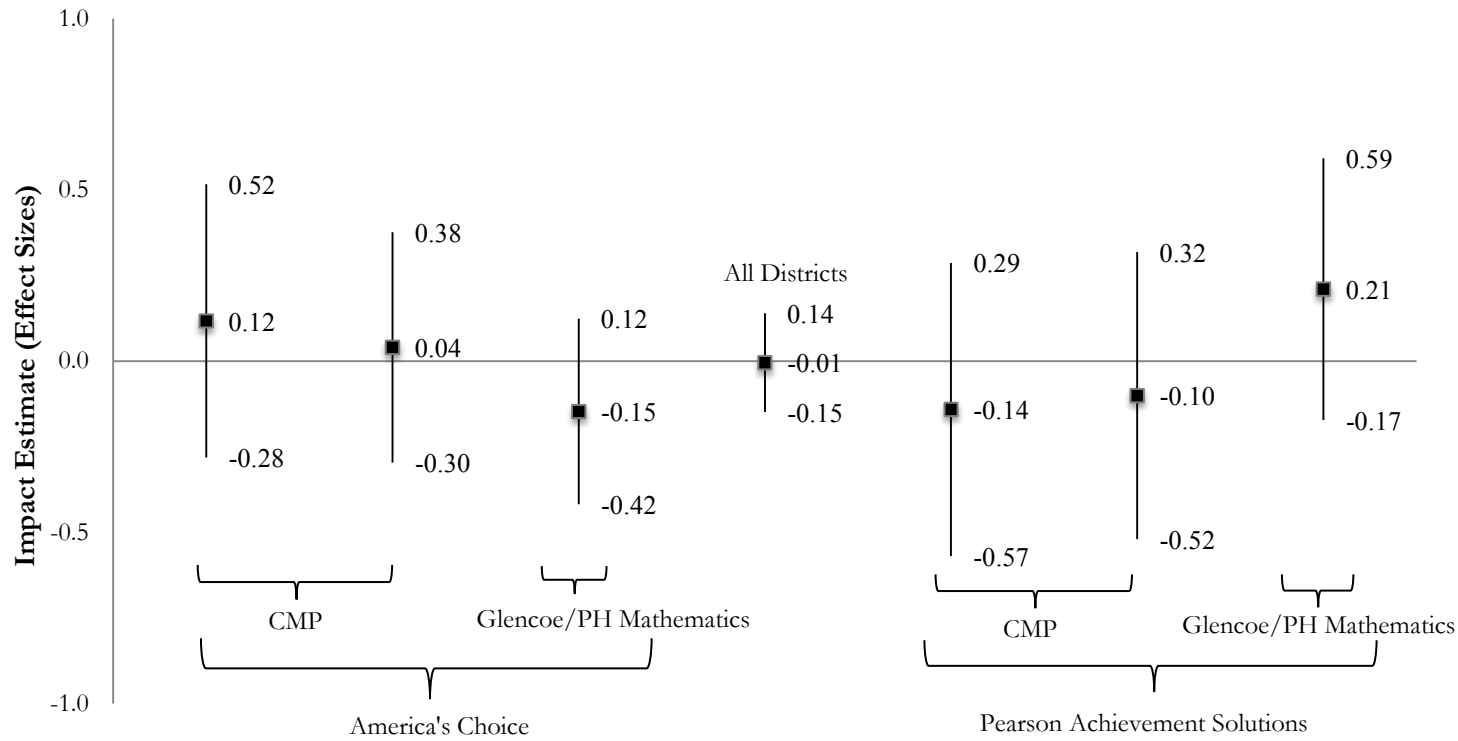
Figure D-1. Impact of the PD Program on Teacher Knowledge at the End of the Second Year: Total Score, by District: Second-Year Teacher Impact Analysis Sample



Sample Size: Total N for all districts=89 teachers, N for each district, reading from left to right=18, 19, 17, 16, 6, and 13 teachers respectively.

SOURCE: Spring 2009 Teacher Knowledge Test.

Figure D-2. Impact of the PD Program on Student Mathematics Achievement at the End of the Second Year: Total Score, by District: Second-Year Student Impact Analysis Sample



Sample Size: Total N for all districts=2132 students, N for each district, reading from left to right=244, 427, 666, 294, 224, and 277 students respectively.

SOURCE: Spring 2009 NWEA Rational Number Test.

Unadjusted Means and Standard Deviations of Second-Year Outcome Measures for Treatment and Control Groups

Table D-11 lists the unadjusted treatment and control group means and standard deviations for the teacher knowledge and student achievement outcomes in the second-year impact analysis samples. The table also includes the weighted unadjusted means for the treatment and control groups, weighted by the number of treatment group schools in each district.

Table D-11. Unadjusted Means and Standard Deviations on Teacher Knowledge and Student Mathematics Achievement: Second-Year Teacher and Student Impact Analysis Samples

Outcome Measure	Treatment Group Mean (Weighted)	Treatment Group Mean (Simple)	Treatment Group S.D. (Simple)	Control Group Mean (Weighted)	Control Group Mean (Simple)	Control Group S.D. (Simple)
Teacher Knowledge						
Total Score (logits)	1.13	1.13	0.90	1.08	1.15	0.82
<i>Percent correctly answering items of average difficulty for the test instrument</i>	75.7	75.6		74.7	76.0	
Common Knowledge of Mathematics (CK) Score (logits)	1.26	1.23	1.20	1.54	1.43	1.10
<i>Percent correctly answering items of average difficulty for the test instrument</i>	79.9	79.4		84.1	82.5	
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.78	0.79	0.98	0.37	0.63	0.91
<i>Percent correctly answering items of average difficulty for the test instrument</i>	65.8	66.1		56.2	62.4	
Sample Size: N = 38 schools (20 treatment, 18 control); 89 teachers (43 treatment, 46 control).						
Student Mathematics Achievement						
NWEA Total Score (scale score)	219.90	220.33	16.36	217.28	217.21	13.84
<i>Corresponding Percentile Rank</i>	22	23		19	19	
Fractions and Decimals Score (scale score)	218.15	218.63	17.54	215.65	215.70	14.98
Ratio and Proportion Score (scale score)	221.71	222.10	16.98	218.85	218.66	14.36
Sample Size: N = 39 schools (20 treatment, 19 control); 2,132 students (1,083 treatment, 1,049 control).						

SOURCE: Spring 2009 Teacher Knowledge Test; Spring 2009 NWEA Rational Number Test.

NOTES: The weighted means are the weighted average of the observed district means for teachers or students, weighted by the number of treatment group schools in each district. The simple means and standard deviation are the nonweighted averages and standard deviations of the teachers and students.

APPENDIX E
EXPLORATORY ANALYSES: APPROACHES AND
ADDITIONAL RESULTS

APPENDIX E

EXPLORATORY ANALYSES: APPROACHES AND ADDITIONAL RESULTS

This appendix provides the descriptions of the analytic approaches and detailed results for each of the exploratory analyses reported in Chapter 4 including the analysis of the one-year impact of the professional development (PD) program at the end of the second year, the analyses of the per-year effect of the PD program on teacher knowledge and student achievement using the pooled sample of participants from both years of the study, the analysis of program effects on teacher knowledge using gain scores rather than post scores as the dependent variable, the analysis of program effects for each PD provider, the differential effects of the PD for teachers and students of varying characteristics, and the correlation between teacher knowledge and student achievement. The appendix also includes baseline equivalence tests, and minimum detectable effect sizes (MDESs).

Analysis of the One-Year Effect of the PD Program on Teacher Knowledge at the End of the Second Year

Table E-1 presents the detailed findings for the analysis of the one-year effect of the PD program on teacher knowledge at the end of the second year of the study. In this analysis, we controlled for teacher knowledge at the end of the first year/beginning of the second year to get an estimate of the effect of the PD during the second year only. No significant effects were found.

Table E-1. One-Year Effect of the PD Program on Teacher Knowledge at the End of the Second Year: Second-Year Teacher Impact Analysis Sample

Outcome Measure	Treatment Group	Control Group	Estimated Impact	Standard Error of the Estimated Impact	Estimated Impact Effect Size	P-value for the Estimated Impact
Total Score (logits)	1.13	1.04	0.09	0.19	0.10	0.63
<i>Percent correctly answering items of average difficulty for the test instrument</i>	75.7	73.9	1.8			
Common Knowledge of Mathematics (CK) Score (logits)	1.26	1.49	-0.23	0.23	-0.17	0.34
<i>Percent correctly answering items of average difficulty for the test instrument</i>	79.9	83.3	-3.4			
Specialized Knowledge of Mathematics (SK) Score (logits)	0.78	0.39	0.39	0.23	0.34	0.11
<i>Percent correctly answering items of average difficulty for the test instrument</i>	65.8	56.6	9.2			

Sample Size: N = 38 schools (20 treatment, 18 control); 89 teachers (43 treatment, 46 control).

SOURCE: Spring 2009 Teacher Knowledge Test.

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 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 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.

Effect sizes were calculated using the control group standard deviation for the First-Year Impact Analysis Sample. The control group standard deviation was 0.97 for the *Total Score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*.

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

Analysis of the Per-Year Effect on Teacher Knowledge

As discussed in Chapter 4, to increase the precision of the estimates of the effects of teacher knowledge, we conducted an analysis of per-year effects using pooled data from the first- and second-year impact samples. The per-year effect is the effect of a single year of PD, as estimated by combining the one-year effects observed for teachers in the first year of the study and the one-year effects for teachers in the second year of the study. (Teachers who were in both years of the study entered the pooled sample twice, once with their first-year effects and once with their second-year effects.)

The pooled sample used in these analyses includes three mutually exclusive and collectively exhaustive groups of teachers: 138 teachers who were part of the first-year impact analysis sample only, 38 teachers who were part of the second-year impact analysis sample only, and 51 teachers who were in both impact analysis samples and therefore enter the sample twice. Because the third group

of teachers was a small portion of the overall sample, we did not explicitly take account of the dependence between the first- and second-year outcomes for these teachers in the analysis.

Teachers who were in both impact analysis samples and are therefore included in the pooled sample twice, appear once using their first-year outcomes, and once using their second-year outcomes. For teachers present only in the first year, we used the fall 2007 teacher knowledge score as a covariate; for teachers present only in the second year, we used the fall 2008 score as a covariate. Teachers present in both years have their fall 2007 score as a covariate when their spring 2008 score is the dependent variable and their spring 2008 score as a covariate when their spring 2009 score is the dependent variable. (The spring 2008 administration occurred after the PD in the first year was complete, but before the second year of PD began.)

The detailed results of the analysis are presented in Table E-2. Table E-3 provides a test of baseline equivalence for treatment and control teachers used in the analysis. There were two statistically significant differences: the treatment group was significantly lower than the control group on baseline common knowledge of mathematics (CK)—estimated difference = -0.43 logits—and on number of postsecondary mathematics education courses taken—estimated difference = -0.4 courses.

To assess the appropriateness of pooling the three groups of teachers, we conducted two supplementary tests. First, using a two-level model controlling for random assignment blocks, we tested whether or not the characteristics of the three types of teachers in the pooled sample differed at baseline. Table E-4 shows that there were no statistically significant differences in background characteristics among the three samples that make up the pooled sample, as evaluated by an F-test.¹²⁰ Second, we also tested whether there was an interaction between sample membership and the effect of the PD program in the pooled sample. Table E-5 displays these findings. No significant differences were found in the effects of the PD program across the three samples.

¹²⁰ Table E-4 also reports a test of baseline equivalence of treatment and control teachers within each of the three samples of teachers that are included in the pooled sample.

Table E-2. Per-Year Effect of the PD Program on Teacher Knowledge: Pooled Sample

Outcome Measure	Treatment Group	Control Group	Estimated Effect (per year)	Standard Error of the Estimated Effect	Effect Size	P-value
Total Score (logits)	0.37	0.19	0.17	0.11	0.18	0.13
<i>Percent correctly answering items of average difficulty for the test instrument</i>	59.0	54.8	4.2			
Common Knowledge of Mathematics (CK) Score (logits)	0.38	0.42	-0.04	0.16	-0.03	0.81
<i>Percent correctly answering items of average difficulty for the test instrument</i>	62.3	63.2	-0.9			
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.38	0.06	0.32*	0.14	0.28	0.02
<i>Percent correctly answering items of average difficulty for the test instrument</i>	56.3	48.3	8.0			

Sample Size: N = 76 schools (40 treatment, 36 control); 278 teachers (139 treatment, 139 control).

SOURCE: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test.

NOTES: The analyses of the effect of the PD program on teacher knowledge were conducted using measures scaled in logits. The estimated effects are based on a three-level model controlling for random assignment block and teacher-level covariates. The treatment and control group columns display regression-adjusted mean outcomes for each group, evaluated at the treatment group mean values for all covariates in the regression model.

Effect sizes were calculated using the spring 2008 Teacher Knowledge Test control group standard deviation. The control group standard deviation was 0.97 for the *Total Score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*.

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

Table E-3. Teacher Characteristics, by Treatment Status: Pooled Sample

Teacher Characteristics	Treatment Group	Control Group	Estimated Difference	P-value for Estimated Difference
Baseline Teacher Knowledge ^a				
Total Score (logits)	-0.07	0.17	-0.24	0.07
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>48.3</i>	<i>54.2</i>	<i>-5.9</i>	
Common Knowledge of Mathematics (CK) Score (logits)	0.03	0.46	-0.43*	0.03
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>54.0</i>	<i>64.3</i>	<i>-10.3</i>	
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	-0.08	0.00	-0.08	0.55
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>44.9</i>	<i>47.0</i>	<i>-2.1</i>	
Years of Teaching Experience (percent)				
3 years or fewer	33.1	27.9	5.2	0.43
4–10 years	30.3	37.9	-7.6	0.24
11–20 years	22.3	27.0	-4.6	0.44
More than 20 years	14.3	7.7	6.6	0.18
Years of Teaching Experience in Middle School Mathematics	6.6	7.4	-0.8	0.52
Educational Level: M.A. and Above (percent)	41.2	37.1	4.2	0.56
Mathematics Major (percent)	15.4	16.5	-1.1	0.81
Number of Postsecondary Mathematics Courses Taken	6.1	6.7	-0.6	0.11
Number of Postsecondary Mathematics Education Courses Taken	1.7	2.1	-0.4*	0.03
Time	25.6	23.7	1.9	0.73

Sample Size: N = 277 teachers (139 treatment, 138 control).

SOURCE: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test.

NOTES: ^a Sample Size: N = 258 teachers (128 treatment; 130 control).

For teachers present only in the first year, we used the fall 2007 teacher knowledge score as a covariate; for teachers present only in the second year, we used the fall 2008 score as a covariate. Teachers present in both years appear in the sample twice, once with their spring 2008 score as the dependent variable and their fall 2007 score as a covariate; and the second time with their spring 2009 score as the dependent variable and their spring 2008 score as a covariate. (The spring 2008 administration occurred after the PD in the first year was complete, but before the second year of PD began.)

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.

Teaching experience was calculated as the number of years of experience at the start of each of the study years. When teachers had their spring 2008 score as the dependent variable, we used their years of experience as of fall 2007. When teachers had their spring 2009 score as the dependent variable, we used their years of experience as of fall 2008.

Values in the columns represent unadjusted means for the groups. Percentage values for characteristics with multiple categories may not sum to 100 owing to rounding.

The analyses are based on a two-level model controlling for random assignment block.

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

Table E-4. Comparison of Teacher Characteristics Between Teachers in First-Year Impact Sample Only, Teachers in Second-Year Impact Sample Only, and Teachers in Both Impact Samples: Pooled Sample

Teacher Characteristic	Teachers in First-Year Impact Sample Only		Teachers in Second-Year Impact Sample Only		Teachers in Both Impact Samples		P-value for Test of Difference in Overall Sample Means ^a
	Treatment	Control	Treatment	Control	Treatment	Control	
Baseline Teacher Knowledge ^b							
Total Score	-0.25	-0.05	0.68	0.83	0.19	0.41	0.05†
Common Knowledge of Mathematics (CK) Score	-0.21	0.22	1.03	0.97	0.31	0.65	0.12
Specialized Knowledge of Mathematics for Teaching (SK) Score	-0.15	-0.19	0.18	0.43	0.14	0.36	0.88
Years of Teaching Experience (percent)							
3 years or fewer	35.6	24.6	35.0	33.3	23.9	39.3	0.75
4–10 years	27.4	36.9	45.0	33.3	32.6	39.3	0.77
More than 10 years	37.0	38.5	20.0	33.3	43.5	21.4	0.90
Years of Teaching Experience In Middle School Mathematics	6.2	8.1	3.6	5.3	8.6	6.0	0.53
Educational Level: M.A. and Above (percent)	37.0	26.2	55.0	44.4	47.8	42.9	0.52
Mathematics Major (percent)	11.0	13.8	30.0	16.7	17.4	14.3	0.47
Number of Postsecondary Mathematics Courses Taken	5.6	5.8	7.8	7.1	6.7	7.4	0.65
Number of Postsecondary Mathematics Education Courses Taken	1.6	1.8	2.2	2.0	1.8	2.3	0.74

Sample Size: N = 76 schools (40 treatment, 36 control); 277 teachers (139 treatment, 138 control).

SOURCE: Fall 2007 Teacher Knowledge Test; Spring 2008 Teacher Knowledge Test; Fall 2008 Teacher Knowledge Test; Teacher Survey.

NOTES: ^a Column displays p-values for tests of difference in overall means (treatment and control combined) among the three constituent samples. The analyses are based on a two-level model controlling for random assignment block. P-values are based on F-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

^b Sample Size: N = 258 teachers.

^c Treatment/control differences within sample were tested using t-tests. The analyses were based on a two-level model controlling for random assignment block. No significant differences were found.

For teachers present only in the first year, we used the fall 2007 teacher knowledge score as a covariate; for teachers present only in the second year, we used the fall 2008 score as a covariate. Teachers present in both years appear in the sample twice, once with their spring 2008 score as the dependent variable and their fall 2007 score as a covariate; and the second time with their spring 2009 score as the dependent variable and their spring 2008 score as a covariate. (The spring 2008 administration occurred after the PD in the first year was complete, but before the second year of PD began.) Teaching experience was calculated as the number of years of experience at the start of each of the study years. When teachers had their spring 2008 score as the dependent variable, we used their years of experience as of fall 2007. When teachers had their spring 2009 score as the dependent variable, we used their years of experience as of fall 2008.

Values in the columns represent unadjusted means for the groups. Percentage values for characteristics with multiple categories may not sum to 100 owing to rounding.

† P-value = 0.0506, which rounds to 0.05 but is not statistically significant at the 0.05 level.

Table E-5. Interaction Between Sample Membership (Teachers in First-Year Impact Sample Only, Teachers in Second-Year Impact Sample Only, and Teachers in Both Impact Samples) and Treatment Effect: Pooled Sample

Outcome Measure	P-value for Test of Interaction Between Sample Membership and Treatment Effect
Teacher Knowledge	
Total Score	0.31
Common Knowledge of Mathematics (CK) Score	0.14
Specialized Knowledge of Mathematics for Teaching (SK) Score	0.64
Sample Size: N = 76 schools (40 treatment, 36 control); 278 teachers (139 treatment, 139 control).	

SOURCE: Spring 2008 and spring 2009 Teacher Knowledge Tests.

NOTES: The estimated effects are based on a two-level model controlling for random assignment block and teacher-level covariates. P-values are based on F-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

Treatment-Control Differences in Baseline Teacher Knowledge

All samples used in analysis were tested for baseline equivalence. With regard to baseline teacher knowledge scores, there was one significant negative difference of -0.43 logits for common knowledge of mathematics (CK) in the pooled sample. The remaining treatment-control differences were negative but nonsignificant, with the exception of one positive, nonsignificant difference of 0.02 logits for specialized knowledge of mathematics for teaching (SK) in the first-year impact sample for one-year districts. The differences in baseline teacher knowledge are summarized in Table E-6 for the samples used in analyses that are included in this report.

Table E-6. Treatment-Control Difference in Baseline Teacher Knowledge for Teacher Samples Used in Specific Analyses

Time Period and Sample	Estimated Difference in Logits (treatment-control)	P-value	Sample Size
Total Score			
End-of-First-Year Impact, All 12 Districts	-0.20	0.22	172
End-of-First-Year Impact, 6 One-Year Districts	-0.28	0.21	90
End-of-First-Year Impact, 6 Two-Year Districts	-0.12	0.61	82
Cumulative End-of-Second-Year Impact, 6 Two-Year Districts	-0.26	0.27	83
End of Second-Year One-Year Impact, 6 Two-Year District	-0.23	0.30	86
Per-Year Effect, Pooled Sample, Total	-0.24	0.07	258
Pooled Sample, Teachers Present in First Year Only	-0.27	0.23	124
Pooled Sample, Teachers Present in Second Year Only	-0.02	0.97	35
Pooled Sample, Teachers Present in Both Years	-0.35	0.12	99

Table continues on next page

Table E-6. Treatment-control Difference in Baseline Teacher Knowledge for Teacher Samples Used in Specific Analyses (continued)

Time Period and Sample	Estimated Difference in Logits (treatment-control)	P-value	Sample Size
Common Knowledge of Mathematics (CK) Score			
End-of-First-Year Impact, All 12 Districts	-0.39	0.11	172
End-of-First-Year Impact, 6 One-Year Districts	-0.53	0.13	90
End-of-First-Year Impact, 6 Two-Year Districts	-0.24	0.45	82
Cumulative End-of-Second-Year Impact, 6 Two-Year Districts	-0.30	0.38	83
End of Second-Year One-Year Effect, 6 Two-Year Districts	-0.31	0.33	86
Per-Year Effect, Pooled Sample	-0.43*	0.03	258
Pooled Sample, Teachers Present in First Year Only	-0.43	0.05†	124
Pooled Sample, Teachers Present in Second Year Only	0.44	0.56	35
Pooled Sample, Teachers Present in Both Years	-0.71	0.06	99
Specialized Knowledge of Mathematics for Teaching (SK) Score			
End-of-First-Year Impact, All 12 Districts	-0.03	0.86	172
End-of-First-Year Impact, 6 One-Year Districts	0.02	0.94	90
End-of-First-Year Impact, 6 Two-Year Districts	-0.07	0.78	82
Cumulative End-of-Second-Year Impact, 6 Two-Year Districts	-0.37	0.08	83
End of Second-Year One-Year Effect, 6 Two-Year Districts	-0.28	0.19	86
Per-Year Effect, Pooled Sample	-0.08	0.55	258
Pooled Sample, Teachers Present in First Year Only	-0.07	0.66	124
Pooled Sample, Teachers Present in Second Year Only	-0.31	0.45	35
Pooled Sample, Teachers Present in Both Years	-0.27	0.22	99

SOURCE: Fall 2007 Teacher Knowledge Test; Fall 2008 Teacher Knowledge Test.

NOTES: The values for baseline teacher knowledge differ depending on the analysis. For the end of first-year impact analyses, baseline teacher knowledge was assessed in fall 2007. For the cumulative end of second-year impact analysis, baseline teacher knowledge was assessed at the beginning of the teacher's first year in the study. For the end-of-second-year one-year impact analysis, baseline knowledge was assessed in spring 2008 for teachers present in the first year and fall 2008 for teachers who entered in the second year.

For the pooled analysis, covariate values differ for different teachers. For teachers present only in the first year, we used the fall 2007 teacher knowledge score as a covariate; for teachers present only in the second year, we used the fall 2008 score as a covariate. Teachers present in both years appear in the sample twice, once with their spring 2008 score as the dependent variable and their fall 2007 score as a covariate; and the second time with their spring 2009 score as the dependent variable and their spring 2008 score as a covariate. (The spring 2008 administration occurred after the PD in the first year was complete, but before the second year of PD began.)

The analyses are based on a two-level model controlling for random assignment block.

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

† P-value = 0.0514, which rounds to 0.05 but is not statistically significant at the 0.05 level.

Analysis of the Average Annual Effect on Student Achievement

As discussed in Chapter 4, to increase the precision of the estimates of the effects of student achievement, we conducted an analysis that pooled data from the first- and second-year impact samples. Because the student samples for each year of the study comprised the students in the teachers' current seventh-grade classes, there was no overlap between these samples except to the extent that individual students from the first year of the study might have been required to repeat seventh-grade mathematics and were assigned to a study classroom for the second year.

The detailed results of the analysis are presented in Table E-7. Table E-8 provides a test of baseline equivalence for treatment and control students used in the analysis. No significant effects were found, and there were no significant differences in treatment and control students at baseline.

Table E-7. Average Annual Effect of the PD Program on Student Achievement: Pooled Sample

Outcome Measure	Treatment Group	Control Group	Estimated Effect (average annual)	Standard Error of the Estimated Effect	Effect Size	P-value
NWEA Total Score (Scale score)	217.38	216.73	0.65	0.54	0.05	0.24
<i>Corresponding Percentile Rank</i>	<i>19</i>	<i>18</i>				
Fractions and Decimals Score (scale score)	215.69	215.14	0.55	0.56	0.04	0.33
Proportion and Ratio Score (scale score)	219.05	218.33	0.72	0.59	0.05	0.23

Sample Size: N = 77 schools (40 treatment, 37 control); 6,660 students (3,419 treatment, 3,241 control).

SOURCE: Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Tests; Study District Records.

NOTES: The analyses of average annual effects on student achievement were conducted using scale scores. The estimated effects are based on a three-level model controlling for random assignment block and student-level covariates. The treatment and control group columns display regression-adjusted mean outcomes for each group, evaluated at the treatment group mean values for all covariates in the regression model.

The values for the corresponding percentile rank correspond to the treatment and control group means in scale scores.

Effect sizes were calculated using the control group standard deviation. The control group standard deviations were 13.84 for the *Total Scale Score*, 14.98 for the *Fractions and Decimals Score*, and 14.36 for the *Ratio and Proportion Score*.

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

Table E-8. Student Characteristics, by Treatment Status: Pooled Sample

Characteristics	Treatment Group	Control Group	Estimated Difference	P-value for Estimated Difference
Age (years) ^a	12.72	12.72	0.00	0.97
Students Eligible for Free or Reduced-Price Lunch (percent)	68.15	77.18	-9.03	0.13
Race/Ethnicity (percent)				
White, non-Hispanic	31.42	29.67	1.75	0.77
Black, non-Hispanic	37.53	38.28	-0.76	0.91
Hispanic	25.74	30.09	-4.35	0.51
Asian/Pacific Islander	2.49	1.52	0.97	0.36
Other	2.83	0.48	2.35	0.06
Male (percent)	50.11	49.91	0.19	0.94
English as a Second Language (percent)	12.89	11.95	0.94	0.81
Special Education Status (percent)	9.50	7.95	1.56	0.48
Sixth-Grade Mathematics Scores on State Accountability Assessment (standardized)	0.15	0.04	0.11	0.34
Fall 2008 Student Mathematics Achievement				
NWEA Total Score (scale score)	215.03	213.25	1.78	0.26
<i>Corresponding Percentile Rank</i>	22	20		
Fractions and Decimals (scale score)	214.13	212.07	2.06	0.22
Proportion and Ratio (scale score)	215.84	214.30	1.53	0.32

Sample Size: N = 6,660 students (3,419 treatment, 3,241 control).

SOURCE: Fall 2007 NWEA Rational Number Test; Fall 2008 NWEA Rational Number Test; Study District Records.

NOTES: ^aAge was calculated as the age (in years) of a student as of September 1 of his or her seventh grade school year.

Values in the columns represent unadjusted means for the groups. Percentage values for characteristics with multiple categories may not sum to 100 owing to rounding.

The analyses are based on a three-level model controlling for random assignment block.

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

Effects of the PD Program on Teacher Knowledge and Student Achievement, by Provider, for the Pooled Sample

The per-year effects of the PD program were estimated by provider using the pooled sample. The results for America's Choice are reported in Tables E-9 and E-10. The results for Pearson Achievement Solutions are reported in Tables E-11 and E-12. No significant effects were found in any of the analyses.

Table E-9. Effect of the PD Program on Teacher Knowledge: Pooled Sample, America's Choice

Outcome Measure	Treatment Group	Control Group	Estimated Effect (per year)	Standard Error of the Estimated Effect	Effect Size	P-value
Total Score (logits)	0.36	0.22	0.15	0.14	0.15	0.29
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>59.0</i>	<i>55.4</i>	<i>3.6</i>			
Common Knowledge of Mathematics (CK) Score (logits)	0.39	0.47	-0.08	0.23	-0.06	0.72
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>62.5</i>	<i>64.4</i>	<i>-1.9</i>			
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.36	0.04	0.32	0.17	0.28	0.08
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>55.9</i>	<i>47.9</i>	<i>7.9</i>			

Sample Size: N = 40 schools (20 treatment, 20 control); 155 teachers (79 treatment, 76 control).

SOURCE: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test.

NOTES: The analyses of the effect of the PD program on teacher knowledge were conducted using measures scaled in logits. The estimated effects are based on a three-level model controlling for random assignment block and teacher-level covariates. The treatment and control group columns display regression-adjusted mean outcomes for each group, evaluated at the treatment group mean values for all covariates in the regression model.

Effect sizes were calculated using the spring 2008 Teacher Knowledge Test control group standard deviation. The control group standard deviations were 0.97 for the *Total Score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*.

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

Table E-10. Effect of the PD Program on Student Achievement: Pooled Sample, America's Choice

Outcome Measure	Treatment Group	Control Group	Estimated Effect (average annual)	Standard Error of Estimated Effect	Effect Size	P-value
NWEA Total Score (Scale score)	215.59	215.16	0.43	0.66	0.03	0.52
<i>Corresponding Percentile Rank</i>	<i>16</i>	<i>16</i>				
Fractions and Decimals Score (scale score)	213.91	213.50	0.41	0.64	0.03	0.53
Proportion and Ratio Score (scale score)	217.27	216.78	0.49	0.75	0.03	0.52

Sample Size: N = 40 schools (20 treatment, 20 control); 3,971 students (2,025 treatment, 1,946 control).

SOURCE: Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Tests; Study District Records.

NOTES: The analyses of the effect of the PD program on student achievement were conducted using scale scores. The estimated effects are based on a three-level model controlling for random assignment block and student-level covariates. The treatment and control group columns display regression-adjusted mean outcomes for each group, evaluated at the treatment group mean values for all covariates in the regression model.

The values for the corresponding percentile rank correspond to the treatment and control group means in scale scores.

Effect sizes were calculated using the control group standard deviation from the First-Year Impact Analysis Sample. The control group standard deviations were 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for the *Ratio and Proportion Score*.

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

Table E-11. Effect of the PD Program on Teacher Knowledge: Pooled Sample, Pearson Achievement Solutions

Outcome Measure	Treatment Group	Control Group	Estimated Effect (per year)	Standard Error of Estimated Effect	Effect Size	P-value
Total Score (logits)	0.37	0.20	0.16	0.19	0.17	0.39
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>59.0</i>	<i>55.1</i>	<i>4.0</i>			
Common Knowledge of Mathematics (CK) Score (logits)	0.37	0.31	0.06	0.23	0.04	0.80
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>62.0</i>	<i>60.6</i>	<i>1.4</i>			
Specialized Knowledge of Mathematics for Teaching (SK) Score (logits)	0.39	0.15	0.24	0.24	0.21	0.31
<i>Percent correctly answering items of average difficulty for the test instrument</i>	<i>56.7</i>	<i>50.6</i>	<i>6.1</i>			

Sample Size: N = 36 schools (20 treatment, 16 control); 123 teachers (60 treatment, 63 control).

SOURCE: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test.

NOTES: The analyses of the effect of the PD program on teacher knowledge were conducted using measures scaled in logits. The estimated effects are based on a three-level model controlling for random assignment block and teacher-level covariates. The treatment and control group columns display regression-adjusted mean outcomes for each group, evaluated at the treatment group mean values for all covariates in the regression model.

Effect sizes were calculated using the spring 2008 Teacher Knowledge Test control group standard deviation. The control group standard deviations were 0.97 for the *Total Score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*.

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

Table E-12. Effect of the PD Program on Student Achievement: Pooled Sample, Pearson Achievement Solutions

Outcome Measure	Treatment Group	Control Group	Estimated Effect (average annual)	Standard Error of Estimated Effect	Effect Size	P-value
NWEA Total Score (scale score)	219.17	218.15	1.02	0.92	0.03	0.28
<i>Corresponding Percentile Rank</i>	22	20				
Fractions and Decimals Score (scale score)	217.48	216.64	0.84	0.97	0.03	0.40
Proportion and Ratio Score (scale score)	220.83	219.69	1.14	0.95	0.03	0.24

Sample Size: N = 37 schools (20 treatment, 17 control); 2,689 students (1,394 treatment, 1,295 control).

SOURCE: Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test; Study District Records.

NOTES: The analyses of the effect of the PD program on student achievement were conducted using scale scores. The estimated effects are based on a three-level model controlling for random assignment block and student-level covariates. The treatment and control group columns display regression-adjusted mean outcomes for each group, evaluated at the treatment group mean values for all covariates in the regression model.

The values for the corresponding percentile rank correspond to the treatment and control group means in scale scores.

Effect sizes were calculated using the control group standard deviation from the First-Year Impact Analysis Sample. The control group standard deviations were 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for the *Ratio and Proportion Score*.

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

Differential PD Effect Based on Baseline Teacher Knowledge and Years of Teaching Experience

To examine whether the effect of the PD program on teacher and student outcomes varied depending on teachers' initial level of knowledge or teachers' general teaching experience at baseline, we reestimated the model used in the pooled analysis described above, adding the interaction of baseline teacher knowledge or experience measures and the treatment indicators.¹²¹

Model 1 (main analysis model):

The following two-level hierarchical model was used to analyze whether the effect of the treatment on teacher outcomes varies with teacher knowledge or teaching experience as measured prior to the program:

$$Y_{jk} = \sum_m \sum_n \gamma_{0mn} B_{mnk} + \gamma_1 T_k + \gamma_2 K_{-1jk} + \gamma_3 (T_k * K_{-1jk}) + \sum_l \gamma_{4l} Z_{jkl} + \mu_k + \nu_{jk} \quad (\text{E-1})$$

Where:

¹²¹ For first-year outcomes, baseline teacher knowledge was measured in fall 2007; for second-year outcomes, baseline teacher knowledge was measured in fall 2008 (for teachers present in the second year only) and in spring 2008 (for teachers who were present in both years). Thus, the model tests the interaction of teacher knowledge at the start of the year and the effect of the PD program during that year. For first-year outcomes, baseline teacher experience was measured in fall 2007; for second-year outcomes, it was measured in fall 2008.

- Y_{jk} = outcome measurement for teacher j from school k,
- B_{mnk} = one if school k is in district m (m = 1 to 12) and block n (n = 1 to 20) and zero otherwise,
- T_k = one if school k is assigned to receive the treatment and zero otherwise,
- K_{-1jk} = baseline teacher knowledge test total score or general teaching experience for teacher j from school k,
- Z_{jkl} = lth baseline characteristics for teacher j from school k (same as the ones used in the impact model), and
- μ_k, U_{jk} = a school-level and a classroom-level random error, respectively, assumed to be independently and identically distributed.

The following three-level hierarchical model was used to analyze whether the effect of the treatment on student achievement varies with teacher knowledge or experience as measured prior to the program:

$$Y_{ijk} = \sum_m \sum_n \gamma_{0mn} B_{mnk} + \gamma_1 T_k + \gamma_2 K_{-1jk} + \gamma_3 (T_k * K_{-1jk}) + \gamma_4 Y_{-1k} + \gamma_5 Y_{-1ijk} + \sum_l \alpha_l X_{lijk} + \mu_k + U_{jk} + \varepsilon_{ijk} \quad (\text{E-2})$$

Where:

- Y_{ijk} = achievement measurement for student i from class j in school k,
- B_{mnk} = one if school k is in block n (n = 1 to 20) in district m (m = 1 to 12) and zero otherwise,
- T_k = one if school k is assigned to receive the PD treatment and zero otherwise,
- K_{-1jk} = baseline teacher knowledge total score or general teaching experience for teacher j from school k,
- Y_{-1ijk} = pretest score for student i from teacher j in school k,
- Y_{-1k} = average baseline NWEA score for school k,
- X_{lijk} = student-level covariate l for student i from teacher j in school k (same as the ones used in the impact model), and
- $\mu_k, U_{jk}, \varepsilon_{ijk}$ = a school-level, teacher-level, and student-level random error, respectively, assumed to be independently and identically distributed.

The coefficient γ^1 is the main estimated program effect on teacher knowledge or experience for the average treatment school in the study sample. A two-tailed t-test is used to assess whether γ^1 differs from zero. The coefficient γ^2 is the main effect of the fall teacher knowledge test total score or teacher experience on teacher outcomes, and γ^3 is the estimated coefficient for the interaction term between baseline teacher knowledge or teacher experience and treatment. Expressed as an effect size, γ^3 represents how much change in the treatment effect is associated with a one standard deviation increase in baseline teacher knowledge or a one-year change in baseline teacher experience.

Model 2:

As a check on the robustness of the analysis, we also estimated the relationships between treatment effect and teachers’ baseline knowledge level and experience using a second set of regressions. In this second approach, we estimated the regressions, allowing the treatment main effect to vary by district, and then calculated the weighted average of the treatment main effects, in the same way we estimated the treatment effects reported in Chapter 3. All other features of the model remain the same as in Equations E-1 and E-2.

Specifically, for teacher outcomes, the following two-level regression was used:

$$Y_{jk} = \sum_m \sum_n \gamma_{0mn} B_{mnk} + \sum_m \gamma_{1m} T_k D_{mk} + \gamma_2 K_{-1jk} + \gamma_3 (T_k * K_{-1jk}) + \sum_l \gamma_{4l} Z_{jkl} + \mu_k + \nu_{jk} \quad (\text{E-3})$$

And for student outcomes, the following three-level regression was used:

$$Y_{ijk} = \sum_m \sum_n \gamma_{0mn} B_{mnk} + \sum_m \gamma_{1m} T_k D_{mk} + \gamma_2 K_{-1jk} + \gamma_3 (T_k * K_{-1jk}) + \gamma_4 Y_{-1k} + \gamma_5 Y_{-1ijk} + \sum_l \alpha_l X_{lijk} + \mu_k + \nu_{jk} + \varepsilon_{ijk} \quad (\text{E-4})$$

Where:

D_{mk} = one if school k is in district m (m = 1 to 12) and zero otherwise.

All other variables are defined the same way as in Equations E-1 and E-2.

Tables E-13 and E-14 report the estimated γ^1 , γ^2 , and γ^3 coefficients for the regressions using both model 1 and model 2. In model 2, the estimated γ^1 , γ^2 , and γ^3 coefficients represent the weighted average of the corresponding coefficients for the 12 districts (using the number of treatment schools in each district as the weight). In general, results estimated from these two models exhibit similar patterns across all outcomes.

Table E-13. Effects of the Interaction Between Treatment Status and Baseline Teacher Knowledge on Teacher and Student Outcomes: Pooled Sample

Outcome Measure		Model 1			Model 2		
		Main Treatment Effect	Main Baseline Effect	Interaction Effect	Main Treatment Effect	Main Baseline Effect	Interaction Effect
Teacher Knowledge							
Total Score	estimate	0.22*	0.61*	0.01	0.27*	0.52*	0.14
	(s.e.)	(0.11)	(0.08)	(0.10)	(0.11)	(0.09)	(0.12)
	[p-value]	[0.04]	[<0.01]	[0.92]	[0.02]	[<0.01]	[0.22]
Common Knowledge of Mathematics (CK) Score	estimate	0.04	0.46*	0.05	0.05	0.37*	0.17
	(s.e.)	(0.11)	(0.08)	(0.10)	(0.12)	(0.09)	(0.12)
	[p-value]	[0.74]	[<0.01]	[0.65]	[0.69]	[<0.01]	[0.16]
Specialized Knowledge of Mathematics for Teaching (SK) Score	estimate	0.30*	0.40*	-0.05	0.35	0.31*	0.07
	(s.e.)	(0.11)	(0.10)	(0.13)	(0.12)	(0.10)	(0.14)
	[p-value]	[0.01]	[<0.01]	[0.69]	[0.01]	[<0.01]	[0.61]
Sample Size: N = 76 schools (40 treatment, 36 control); 258 teachers (128 treatment, 130 control).							
Student Mathematics Achievement							
NWEA Total Score	estimate	0.04	0.02	0.03	0.07	0.01	0.04
	(s.e.)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
	[p-value]	[0.24]	[0.53]	[0.36]	[0.10]	[0.72]	[0.23]
Fractions and Decimals Score	estimate	0.03	0.03	0.02	0.05	0.03	0.03
	(s.e.)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
	[p-value]	[0.35]	[0.27]	[0.50]	[0.14]	[0.35]	[0.39]
Ratio and Proportion Score	estimate	0.04	-0.00	0.04	0.06	-0.01	0.05
	(s.e.)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
	[p-value]	[0.25]	[0.95]	[0.28]	[0.13]	[0.77]	[0.15]

Sample Size: N = 76 schools (40 treatment, 36 control); 5,979 students (3,045 treatment, 2,934 control).

SOURCES: Fall 2007, Fall 2008, Spring 2008, and Spring 2009 Teacher Knowledge Test; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test.

NOTES: Estimates in the table are standardized regression coefficients for the interaction between the treatment indicator and baseline teacher knowledge. For teacher knowledge, the coefficients were estimated on the basis of a two-level model controlling for random assignment block and teacher-level covariates. For student mathematics achievement, the coefficients were estimated on the basis of a three-level model controlling for random assignment block and student-level covariates.

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

Table E-14. Effects of the Interaction Between Treatment Status and Years of Teacher Experience on Teacher and Student Outcomes: Pooled Sample

Outcome Measures		Model 1			Model 2		
		Main Treatment Effect	Main Baseline Effect	Interaction Effect	Main Treatment Effect	Main Baseline Effect	Interaction Effect
Teacher Knowledge							
Total Score	estimate	0.26	0.01	-0.01	0.39*	0.02	-0.02
	(s.e.)	(0.16)	(0.02)	(0.01)	(0.18)	(0.02)	(0.01)
	[p-value]	[0.11]	[0.43]	[0.68]	[0.04]	[0.13]	[0.29]
Common Knowledge of Mathematics (CK) Score	estimate	0.04	0.01	-0.00	0.06	0.01	0.00
	(s.e.)	(0.17)	(0.02)	(0.01)	(0.19)	(0.02)	(0.02)
	[p-value]	[0.83]	[0.64]	[0.94]	[0.75]	[0.47]	[0.81]
Specialized Knowledge of Mathematics for Teaching (SK) Score	estimate	0.34	0.02	-0.01	0.50*	0.03	-0.02
	(s.e.)	(0.17)	(0.02)	(0.01)	(0.19)	(0.02)	(0.02)
	[p-value]	[0.05] †	[0.34]	[0.63]	[0.01]	[0.06]	[0.19]

Sample Size: N = 76 schools (40 treatment, 36 control); 258 teachers (128 treatment, 130 control).

Student Mathematics Achievement

NWEA Total Score	estimate	0.09	-0.00	-0.00	0.10	0.00	0.00
	(s.e.)	(0.05)	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)
	[p-value]	[0.11]	[0.99]	[0.39]	[0.10]	[0.98]	[0.50]
Fractions and Decimals Score	estimate	0.08	0.00	-0.00	0.08	0.00	0.00
	(s.e.)	(0.05)	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)
	[p-value]	[0.15]	[0.70]	[0.32]	[0.15]	[0.71]	[0.47]
Ratio and Proportion Score	estimate	0.09	0.00	-0.00	0.11	0.00	0.00
	(s.e.)	(0.05)	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)
	[p-value]	[0.11]	[0.79]	[0.47]	[0.08]	[0.83]	[0.51]

Sample Size: N = 76 schools (40 treatment, 36 control); 6,371 students (3,261 treatment, 3,110 control).

SOURCES: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test; Teacher Survey.

NOTES: Estimates in the table are standardized regression coefficients for the interaction between the treatment indicator and baseline teacher experience. For teacher knowledge, the coefficients were estimated on the basis of a two-level model controlling for random assignment block and teacher-level covariates. For student mathematics achievement, the coefficients were estimated on the basis of a three-level model controlling for random assignment block and student-level covariates.

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

† P-value = 0.054, which rounds to 0.05 but is not statistically significant at the 0.05 level.

Results reported above assume a linear relationship between teacher baseline characteristics and the program effect. Without any theoretical and empirical evidence, there is no compelling reason to choose one specification over another, and the linear specification has the advantage of being the most parsimonious and interpretable form. However, we also wanted to check whether another potential specification of this relationship yields different results. Specifically, we added a quadratic term of teacher baseline characteristics (K_{-1jk}^2) and its interaction with treatment ($T * K_{-1jk}^2$) into Equations E-1 and E-2 and reestimated the models to see whether the relationship between teacher's baseline knowledge or experience and the program effect is sensitive to this model specification. Results from the augmented models are presented in Tables E-15 through E-18. None of the interaction terms between the treatment indicator and the teacher characteristics or their squared terms is statistically significant for any of the teacher or student outcome measures.

Table E-15. Effects of the Linear and Quadratic Interaction Between Treatment Status and Baseline Teacher Knowledge on Teacher and Student Outcomes: Pooled Sample, Augmented Model 1

		Model 1				
Outcome Measure		Main Treatment Effect	Main Baseline Effect	Interaction Effect	Main Baseline Quadratic Effect	Quadratic Interactive Effect
Teacher Knowledge						
Total Score	estimate	0.35*	0.62*	0.05	-0.02	-0.10
	(s.e.)	(0.12)	(0.09)	(0.11)	(0.05)	(0.07)
	[p-value]	[0.01]	[<0.01]	[0.67]	[0.76]	[0.13]
Common Knowledge of Mathematics (CK) Score	estimate	0.12	0.52*	0.04	-0.05	-0.05
	(s.e.)	(0.13)	(0.10)	(0.12)	(0.05)	(0.07)
	[p-value]	[0.36]	[<0.01]	[0.71]	[0.36]	[0.45]
Specialized Knowledge of Mathematics for Teaching (SK) Score	estimate	0.37*	0.40*	-0.01	-0.08	-0.07
	(s.e.)	(0.13)	(0.10)	(0.13)	(0.07)	(0.10)
	[p-value]	[0.01]	[<0.01]	[0.94]	[0.25]	[0.45]
Sample Size: N = 76 schools (40 treatment, 36 control); 258 teachers (128 treatment, 130 control).						
Student Achievement						
NWEA Total Score	estimate	0.06	0.02	0.03	-0.01	-0.01
	(s.e.)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)
	[p-value]	[0.17]	[0.46]	[0.39]	[0.71]	[0.58]
Fractions and Decimals Score	estimate	0.04	0.03	0.02	0.00	-0.01
	(s.e.)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)
	[p-value]	[0.28]	[0.28]	[0.52]	[0.86]	[0.66]
Ratio and Proportion Score	estimate	0.07	0.01	0.04	-0.01	-0.02
	(s.e.)	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)
	[p-value]	[0.14]	[0.79]	[0.32]	[0.63]	[0.49]

Sample Size: N = 76 schools (40 treatment, 36 control); 5,979 students (3,045 treatment, 2,934 control).

SOURCES: Fall 2007, Fall 2008, Spring 2008, and Spring 2009 Teacher Knowledge Tests; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test.

NOTES: Estimates in the table are semistandardized. The dependent variables and baseline teacher knowledge are standardized. The quadratic term is the square of the standardized baseline knowledge score, and treatment condition is coded 1/0. In addition to the linear and quadratic forms of the teacher baseline knowledge score and their interaction with the treatment indicator, the two-level model for teacher knowledge outcomes also controls for random assignment block and other teacher-level covariates used in the impact model (see Appendix B, Equation B-1 for detail). Likewise, the three-level model for student mathematics achievement also controls for random assignment block and other student-level covariates used in the impact model (see Appendix B, Equation B-2 for detail).

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

Table E-16. Effects of the Linear and Quadratic Interaction Between Treatment Status and Baseline Teacher Knowledge on Teacher and Student Outcomes: Pooled Sample, Augmented Model 2

		Model 2				
Outcome Measure		Main Treatment Effect	Main Baseline Effect	Interaction Effect	Main Baseline Quadratic Effect	Quadratic Interactive Effect
Teacher Knowledge						
Total Score	estimate	0.42*	0.51*	0.20	0.01	-0.13
	(s.e.)	(0.14)	(0.10)	(0.13)	(0.05)	(0.07)
	[p-value]	[<0.01]	[<0.01]	[0.11]	[0.90]	[0.07]
Common Knowledge of Mathematics (CK) Score	estimate	0.17	0.40*	0.20	-0.02	-0.09
	(s.e.)	(0.15)	(0.11)	(0.13)	(0.06)	(0.07)
	[p-value]	[0.24]	[<0.01]	[0.13]	[0.71]	[0.23]
Specialized Knowledge of Mathematics for Teaching (SK) Score	estimate	0.41*	0.32*	0.11	-0.08	-0.06
	(s.e.)	(0.14)	(0.10)	(0.14)	(0.07)	(0.10)
	[p-value]	[0.01]	[<0.01]	[0.44]	[0.25]	[0.53]
Sample Size: N = 76 schools (40 treatment, 36 control); 258 teachers (128 treatment, 130 control).						
Student Achievement						
NWEA Total Score	estimate	0.08	0.01	0.05	-0.01	-0.01
	(s.e.)	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)
	[p-value]	[0.07]	[0.66]	[0.25]	[0.77]	[0.52]
Fractions and Decimals Score	estimate	0.07	0.02	0.03	0.00	-0.02
	(s.e.)	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)
	[p-value]	[0.11]	[0.42]	[0.37]	[0.93]	[0.50]
Ratio and Proportion Score	estimate	0.09	0.00	0.05	-0.01	-0.02
	(s.e.)	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)
	[p-value]	[0.08]	[0.97]	[0.19]	[0.59]	[0.50]

Sample Size: N = 76 schools (40 treatment, 36 control); 5,979 students (3,045 treatment, 2,934 control).

SOURCES: Fall 2007, Fall 2008, Spring 2008, and Spring 2009 Teacher Knowledge Tests; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test.

NOTES: Estimates in the table are semistandardized. The dependent variables and baseline teacher knowledge are standardized. The quadratic term is the square of the standardized baseline knowledge score, and treatment condition is coded 1/0. In addition to the linear and quadratic forms of the teacher baseline knowledge score and their interaction with the treatment indicator, the two-level model for teacher knowledge and instructional practice outcomes also controls for random assignment block and other teacher-level covariates used in the impact model (see Appendix B, Equation B-1 for detail). Likewise, the three-level model for student mathematics achievement also controls for random assignment block and other student-level covariates used in the impact model (see Appendix B, Equation B-2 for detail).

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

Table E-17. Effects of the Quadratic Interaction Between Treatment Status and Years of Teaching Experience on Teacher and Student Outcomes: Pooled Sample, Augmented Model 1

Outcome Measure		Model 1				
		Main Treatment Effect	Main Baseline Effect	Interaction Effect	Main Baseline Quadratic Effect	Quadratic Interactive Effect
Teacher Knowledge						
Total Score	estimate	0.26	0.00	-0.00	0.00	-0.00
	(s.e.)	(0.22)	(0.03)	(0.04)	(0.00)	(0.00)
	[p-value]	[0.25]	[0.91]	[0.93]	[0.79]	[0.93]
Common Knowledge of Mathematics (CK) Score	estimate	-0.02	-0.01	0.01	0.00	-0.00
	(s.e.)	(0.23)	(0.03)	(0.04)	(0.00)	(0.00)
	[p-value]	[0.92]	[0.87]	[0.73]	[0.64]	[0.68]
Specialized Knowledge of Mathematics for Teaching (SK) Score	estimate	0.29	0.01	0.01	0.00	-0.00
	(s.e.)	(0.23)	(0.03)	(0.04)	(0.00)	(0.00)
	[p-value]	[0.22]	[0.76]	[0.88]	[0.82]	[0.75]
Sample Size: N = 76 schools (40 treatment, 36 control); 258 teachers (128 treatment, 130 control).						
Student Achievement						
NWEA Total Score	estimate	0.05	0.01	0.00	-0.00	-0.00
	(s.e.)	(0.07)	(0.01)	(0.01)	(0.00)	(0.00)
	[p-value]	[0.46]	[0.50]	[0.75]	[0.48]	[0.66]
Fractions and Decimals Score	estimate	0.05	0.01	0.00	-0.00	-0.00
	(s.e.)	(0.07)	(0.01)	(0.01)	(0.00)	(0.00)
	[p-value]	[0.47]	[0.47]	[0.91]	[0.54]	[0.76]
Ratio and Proportion Score	estimate	0.05	0.01	0.01	-0.00	-0.00
	(s.e.)	(0.07)	(0.01)	(0.01)	(0.00)	(0.00)
	[p-value]	[0.46]	[0.49]	[0.72]	[0.42]	[0.68]

Sample Size: N = 76 schools (40 treatment, 36 control); 6,371 students (3,261 treatment, 3,110 control).

SOURCES: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test; Fall Teacher Surveys.

NOTES: Estimates in the table are semistandardized. The dependent variables are standardized. The quadratic term is the square of the teacher's baseline years of experience, and treatment condition is coded 1/0. In addition to the linear and quadratic forms of the teacher baseline experience score and their interaction with the treatment indicator, the two-level model for teacher knowledge outcomes also controls for random assignment block and other teacher-level covariates used in the impact model (see Appendix B, Equation B-1 for detail). Likewise, the three-level model for student mathematics achievement also controls for random assignment block and other student-level covariates used in the impact model (see Appendix B, Equation B-2 for detail).

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

Table E-18. Effects of the Linear and Quadratic Interaction Between Treatment Status and Years of Teaching Experience on Teacher and Student Outcomes: Pooled Sample, Augmented Model 2

Outcome Measure		Model 2				
		Main Treatment Effect	Main Baseline Effect	Interaction Effect	Main Baseline Quadratic Effect	Quadratic Interactive Effect
Teacher Knowledge						
Total Score	estimate	0.40	0.02	-0.02	0.00	0.00
	(s.e.)	(0.24)	(0.03)	(0.04)	(0.00)	(0.00)
	[p-value]	[0.10]	[0.63]	[0.65]	[0.80]	[0.96]
Common Knowledge of Mathematics (CK) Score	estimate	-0.01	-0.00	0.01	0.00	-0.00
	(s.e.)	(0.25)	(0.03)	(0.04)	(0.00)	(0.00)
	[p-value]	[0.98]	[0.97]	[0.77]	[0.64]	[0.67]
Specialized Knowledge of Mathematics for Teaching (SK) Score	estimate	0.50*	0.02	-0.02	0.00	-0.00
	(s.e.)	(0.24)	(0.03)	(0.04)	(0.00)	(0.00)
	[p-value]	[0.05]	[0.49]	[0.62]	[0.81]	[1.00]
Sample Size: N = 76 schools (40 treatment, 36 control); 258 teachers (128 treatment, 130 control).						
Student Achievement						
NWEA Total Score	estimate	0.07	0.01	0.01	0.00	0.00
	(s.e.)	(0.08)	(0.01)	(0.01)	(0.00)	(0.00)
	[p-value]	[0.40]	[0.41]	[0.70]	[0.39]	[0.66]
Fractions and Decimals Score	estimate	0.06	0.01	0.00	0.00	0.00
	(s.e.)	(0.07)	(0.01)	(0.01)	(0.00)	(0.00)
	[p-value]	[0.42]	[0.37]	[0.89]	[0.42]	[0.82]
Ratio and Proportion Score	estimate	0.07	0.01	0.01	0.00	0.00
	(s.e.)	(0.08)	(0.01)	(0.01)	(0.00)	(0.00)
	[p-value]	[0.38]	[0.42]	[0.68]	[0.36]	[0.65]
Sample Size: N = 76 schools (40 treatment, 36 control); 6,371 students (3,261 treatment, 3,110 control).						

SOURCES: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test; Teacher Surveys.

NOTES: Estimates in the table are semistandardized. The dependent variables are standardized. The quadratic term is the square of the baseline experience in years, and treatment condition is coded 1/0. In addition to the linear and quadratic forms of the teacher baseline experience and their interaction with the treatment indicator, the two-level model for teacher knowledge outcomes also controls for random assignment block and other teacher-level covariates used in the impact model (see Appendix B, Equation B-1 for detail). Likewise, the three-level model for student mathematics achievement also controls for random assignment block and other student-level covariates used in the impact model (see Appendix B, Equation B-2 for detail).

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

Differential PD Effects Based on Baseline Student Achievement

The following models were used to examine how treatment effect on student achievement varies with the students' Northwest Evaluation Association (NWEA) total test scores prior to the program. In this analysis:

Model 1

$$Y_{ijk} = \sum_m \sum_n \gamma_{0mn} B_{mnk} + \gamma_1 T_k + \gamma_2 Y_{-1ijk} + \gamma_3 (T_k * Y_{-1ijk}) + \gamma_4 Y_{-1k} + \sum_l \alpha_l X_{lijk} + \mu_k + \nu_{jk} \quad (\text{E-5})$$

As in the analyses of differential PD effects based on baseline teacher knowledge, we also conducted a robustness check in which we allowed the treatment main effect to vary by district, and then calculated the weighted average of the treatment main effects. The model used for this second analysis is:

Model 2

$$Y_{ijk} = \sum_m \sum_n \gamma_{0mn} B_{mnk} + \sum_m \gamma_{1m} T_k D_{mk} + \gamma_2 Y_{-1ijk} + \gamma_3 (T_k * Y_{-1ijk}) + \gamma_4 Y_{-1k} + \sum_l \alpha_l X_{lijk} + \mu_k + \nu_{jk} + \varepsilon_{ijk} \quad (\text{E-6})$$

All variables are defined in Equation B-2 in Appendix B.

Table E-19 reports the estimated γ_1 , γ_2 , and γ_3 from Model 1 and Model 2. Neither set of results shows significant interactions between baseline student achievement and the effects of the PD on any of the student achievement outcomes.

Table E-19. Effects of the Interaction Between Treatment Status and Baseline Student Achievement on Student Outcomes: Pooled Sample

Outcome Measures		Model 1			Model 2		
		Main Treatment Effect	Main Baseline Effect	Interaction Effect	Main Treatment Effect	Main Baseline Effect	Interaction Effect
Student Mathematics Achievement							
NWEA Total Score	estimate	0.00	0.76*	0.03	0.02	0.76*	0.03
	(s.e.)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
	[p-value]	[0.89]	[<0.01]	[0.25]	[0.52]	[<0.01]	[0.2]
Fractions and Decimals Score	estimate	-0.00	0.73*	0.04	0.01	0.73*	0.05
	(s.e.)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
	[p-value]	[1.00]	[<0.01]	[0.12]	[0.66]	[<0.01]	[0.08]
Ratio and Proportion Score	estimate	0.01	0.72*	0.01	0.03	0.71*	0.02
	(s.e.)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)
	[p-value]	[0.80]	[<0.01]	[0.56]	[0.49]	[<0.01]	[0.53]

Sample Size: N = 77 schools (40 treatment, 37 control); 4,187 students (2,154 treatment, 2,033 control).

SOURCES: Fall 2007, 2009 Fall 2008, Spring 2008, and Spring NWEA Rational Number Tests.

NOTES: Estimates in the table are standardized regression coefficients for the interaction between the treatment indicator and the baseline NWEA Rational Number Test. The coefficients were estimated based on a three-level model controlling for random assignment block and student-level covariates.

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

MDESs for Test of Differential PD Effects

Table E-20 presents the MDESs for the tests for the interactions between baseline teacher knowledge and the treatment effects and between baseline student achievement and the treatment effects.

Table E-20. Minimum Detectable Effect Sizes (MDESs) for Interaction Between Treatment Status and Baseline Teacher Knowledge, Years of Teaching Experience, and Student Achievement: Pooled Sample

Outcome Measures	MDES for Interaction Effect		
	Treatment by Baseline Teacher Knowledge	Treatment by Baseline Teacher Experience	Treatment by Baseline Student Achievement
Teacher Knowledge			
Total Score	0.30	0.04	--
Common Knowledge of Mathematics (CK) Score	0.22	0.03	--
Specialized Knowledge of Mathematics for Teaching (SK) Score	0.31	0.03	--
Sample Size: N = 76 schools (40 treatment, 36 control); 258 teachers (128 treatment, 130 control).			
Student Mathematics Achievement			
NWEA Total Score	0.09	0.01	0.07
Fractions and Decimals Score	0.09	0.01	0.07
Ratio and Proportion Score	0.09	0.01	0.07

Sample Size: N = 77 schools (40 treatment, 37 control); 6,371 students (3,261 treatment, 3,110 control).

SOURCES: Fall 2007, Fall 2008, Spring 2008, and Spring 2009 Teacher Knowledge Tests; Fall 2007, Fall 2008, Spring 2008, and Spring 2009 NWEA Rational Number Tests; Teacher Survey.

NOTES: MDESs are based on the standard errors of the interaction effect estimates.

The estimated impacts for teacher knowledge are based on a two-level model controlling for random assignment block and teacher-level covariates, and the estimated impacts for student mathematics achievement are based on a three-level model controlling for random assignment block and student-level covariates.

MDESs were calculated using the control group standard deviations. The control group standard deviations for Teacher Knowledge measures were 0.97 for the *Total Score*, 1.36 for the *CK Score*, and 1.14 for the *SK Score*. The control group standard deviations for the Student Mathematics Achievement measures were 14.27 for the *Total Score*, 15.23 for the *Fractions and Decimals Score*, and 15.06 for the *Ratio and Proportion Score*.

Relationships Between Teacher Knowledge and Student Achievement

This section provides additional detail regarding the analysis in Chapter 4 of the relationships between teacher knowledge and student achievement. The general approach employed in this analysis was to examine the extent to which students' spring seventh-grade achievement varied across teachers, after controlling for baseline student achievement and other student covariates, and then to add teacher variables to the model as predictors of achievement. This permitted us to determine the extent to which variation across teachers was reduced when teacher knowledge was included as a predictor in the model. The model was also used to estimate the magnitude of association between teacher knowledge and student achievement. In order to base the analysis on the level of teacher knowledge that the students experienced over the course of the year, the teacher knowledge measure used was the average of the fall and spring test scores for each teacher. As noted earlier, each student is in the pooled sample for only one year, and each student is paired with the teacher knowledge value for his/her teacher in that year. A hierarchical linear model (HLM) was used to estimate these relationships. The following model was used to estimate the conditional relationship of teacher knowledge to student achievement, controlling for other factors in the model. The three-level model is as follows:

Level 1: Student Level

$$Y_{ijk} = \pi_{0jk} + \varepsilon_{ijk} \quad (\text{E-7.1})$$

$$Y_{ijk} = \pi_{0jk} + \sum_s \pi_{jks} X_{ijks} + \varepsilon_{ijk} \quad (\text{E-7.2})$$

Where:

Y_{ijk} = mathematics achievement of student i in a class taught by teacher j at school k ,

X_{ijks} = s th individual student characteristic (e.g., baseline achievement, race/ethnicity poverty status) for student i in a class taught by teacher j at school k , and

ε_{ijk} = student-level random error, assumed to be independently and identically distributed across students.

Level 2: Teacher Level

$$\pi_{0jk} = \beta_{00k} + \mu_{0jk} \quad (\text{E-8.1})$$

$$\pi_{0jk} = \beta_{00k} + \sum_w \Omega_{03kw} Z_{0jkw} + \mu_{0jk} \quad (\text{E-8.2})$$

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} TK_{0jk} + \sum_w \Omega_{03kw} Z_{0jkw} + \mu_{0jk} \quad (\text{E-8.3})$$

Where

TK_{0jk} = teacher knowledge test score (average¹²²) for teacher j in school k ,

Z_{0jkw} = w^{th} teacher characteristic measure for teacher j in school k (e.g., teacher experience), and

μ_{0jk} = teacher-level random error, assumed to be independently and identically distributed across teachers.

¹²² For achievement outcomes measured in the first year, the average of the fall 2007 knowledge test score and the spring 2008 teacher knowledge test score for each teacher was used to represent the average level of teacher knowledge a student experienced during the year. For achievement outcomes measured in the second year, the average of the spring 2008 or fall 2008 and spring 2009 teacher knowledge score for each teacher was used to represent the average level of teacher knowledge a student experienced during the year. For teachers missing the relevant fall or spring teacher knowledge score, the nonmissing score was used. Teachers missing both the fall and spring score were excluded from the analysis.

Level 3: School Level

$$\beta_{00k} = \eta_{00k} \tag{E-9.1}$$

$$\beta_{00k} = \sum_m \sum_n \gamma_{0mn} B_{mn} + \eta_{00k} \tag{E-9.2}$$

$$\beta_{00k} = \sum_m \sum_n \gamma_{0mn} B_{mn} + \gamma_{001} SCH_{00k} + \eta_{00k} \tag{E-9.3}$$

Where:

B_{mn} = one if a school is in block n ($n = 1$ to 20) in district m ($m = 1$ to 12) and zero otherwise,

SCH_{00k} = school average baseline NWEA test score, and

η_{00k} = school-level random error, assumed to be independently and identically distributed across schools.

The three models at the teacher level (Equations E-8.1, E-8.2, and E-8.3) were designed to allow us to examine the sources of variation in achievement. We assessed how much of the variation in student achievement was at the school level and the teacher level by using Equation E-8.1 as an initial teacher-level model. Then other teacher characteristics were added in (Equation E-8.2) to see how much of the variation in student achievement at the teacher and school levels was explained by these teacher characteristics. Lastly, we added teacher knowledge measures (separately and jointly) into the teacher-level model (Equation E-8.3).

As an initial step in the analysis, we examined the extent to which student achievement varied across the teachers in the sample schools.¹²³ This question sets the stage for the correlational analyses, because teacher knowledge can be related to student achievement only to the extent that student achievement varies among teachers.

Table E-21 presents the variance decomposition for the standardized student spring total NWEA test scores. The table reports results based on the following models:

- With no control variables (Equations E-7.1, E-8.1, and E-9.1) (benchmark).
- Add random assignment block fixed effects (Equations E-7.1, E-8.1, and E-9.2) (model 1).
- Add school and student characteristics (Equations E-7.2, E-8.1, and E-9.3) (model 2).
- Add teacher characteristics (Equations E-7.2, E-8.2, and E-9.3) (model 3).
- Add teacher knowledge measures (Equations E-7.2, E-8.3, E-9.3) (model 4).

¹²³ The analysis is based on a three-level model, with students nested within teachers and teachers nested within schools.

The results of this analysis (reported in Table E-21) show that after controlling for student demographics and prior achievement, as well as for teacher education and experience, 3 percent of the total variance in student achievement was at the teacher level (0.03 compared with a total of 1.08), and 5 percent of the adjusted variance in student achievement was at the teacher level (0.03 compared with an adjusted total of 0.64). If we interpret the teacher-level variance as the variation among teachers in their effectiveness in raising student achievement, students with a teacher one standard deviation above average in effectiveness scored 0.17 standard deviations above average.¹²⁴

Table E-22 explores the interaction between sample membership and teacher knowledge in the regression predicting student achievement. The pooled sample includes three separate samples (teachers in the first-year impact sample only, teachers in the second-year impact sample only, and teachers in both impact samples. The latter enter into the sample twice—once with their first-year values and once with their second-year values). In the model we estimated, we treated the teachers in the first-year impact sample only as the reference group, and we included two sample membership interaction terms: one representing teachers in the second-year impact sample only versus teachers in the first-year impact sample only, and one representing teachers in both impact samples versus teachers in the first-year impact sample only. We used an F test to test the joint significance of the two interaction terms in each model. None of the joint p-values were statistically significant.

In Table 4-4, we reported the coefficients for the teacher knowledge variables, which indicate the association between the teacher outcome measures and the main student outcome, controlling for all other covariates in the model. The reported coefficient in Table 4-4 for teacher knowledge (β_{01}) comes from Equation E-8.3. The coefficient β_{01} is the conditional relationship between teacher knowledge and student achievement, holding all covariates constant. To simplify the interpretation of the estimated coefficients, all teacher knowledge and student achievement measures were standardized, and thus the estimated coefficients can be interpreted as effect sizes. Therefore, β_{01} is the effect size for the change in student test score associated with a one standard deviation increase in teacher knowledge, controlling for all other variables in the model.

¹²⁴ The value 0.17 is calculated as the square root of the variance at the teacher level as a proportion of the total variance in achievement ($0.03/1.08 = 0.03$). The estimated between-teacher variation in student achievement is similar to findings reported in the literature. For example, Rockoff (2004) looked at two school districts in New Jersey and found that “moving up one standard deviation in the teacher fixed effect distribution raises both reading and mathematics test scores by approximately 0.1 standard deviations on a nationally standardized scale.” Hanushek, Kain, O’Brien, and Rivkin (2005) put the best bounds on the standard deviation associated with teacher quality at 0.22 to 0.27. Using data from the Los Angeles Unified School District, Kane and Staiger (2008) put the estimate in the range of 0.10 to 0.25 standard deviations.

Table E-21. Variance Decomposition of Standardized Student Spring Total NWEA Test Scores by Data Structure Level: Pooled Sample

Level	Benchmark	Model 1	Model 2	Model 3	Model 4
	No Control Variables	Control for Block	Control for Block, School-, Student-Level Covariates	Control for Block, School-, Teacher-, Student-Level Covariates	Control for Block, School-, Teacher-, Student-Level Covariates, Also for TK Total Score
Total Variance	1.08	1.08	1.08	1.08	1.08
Adjusted Variance	1.08	0.99	0.65	0.64	0.64
School Level					
Variance	0.14*	0.04*	0.00	0.00	0.00
As a proportion of total variance	0.12	0.04	0.00	0.00	0.00
As a proportion of adjusted variance	0.12	0.04	0.01	0.01	0.01
P-value	<0.01	<0.01	0.37	0.40	0.46
Teacher Level					
Variance	0.08*	0.08*	0.04*	0.03*	0.03*
As a proportion of total variance	0.07	0.07	0.03	0.03	0.03
As a proportion of adjusted variance	0.07	0.08	0.05	0.05	0.05
P-value	<0.01	<0.01	<0.01	<0.01	<0.01
Student Level					
Variance	0.87*	0.87*	0.61*	0.61*	0.61*
As a proportion of total variance	0.80	0.80	0.56	0.56	0.56
As a proportion of adjusted variance	0.80	0.88	0.94	0.94	0.94
P-value	<0.01	<0.01	<0.01	<0.01	<0.01

Sample Size = 76 schools (40 treatment, 36 control); 276 teachers (138 treatment, 138 control); 6,352 students (3,240 treatment, 3,112 control).

SOURCES: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test.

NOTES: The variance components are estimated using three-level hierarchical linear models controlling for random assignment blocks and various sets of covariates as indicated in the table.

The NWEA total scale score is standardized using the distribution of the control group, which is 14.27 based on the first-year impact analysis sample.

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

Table E-22. Interaction Between Teacher Knowledge and Sample Membership (Teachers in First-Year Impact Sample Only, Teachers in Second-Year Impact Sample Only, and Teachers in Both Impact Samples) in Regression Predicting Student Achievement: Pooled Sample

Standardized Outcomes	Coefficients	Mediating Variables in Model			
		TK Total Score	CK Score	SK Score	
NWEA Total Score	Both Samples vs. Year 1 only	0.07			
	p-value	0.06			
	Year 2 only vs. Year 1 only	0.04			
	p-value	0.36			
	Joint p-value	0.15			
	Both Samples vs. Year 1 only		0.03	0.06	
	p-value		0.51	0.23	
	Year 2 only vs. Year 1 only		0.13*	-0.09	
	p-value		0.03	0.22	
	Joint p-value		0.11	0.13	
	Fractions and Decimals Score	Both Samples vs. Year 1 only	0.06		
		p-value	0.09		
Year 2 only vs. Year 1 only		0.06			
p-value		0.21			
Joint p-value		0.16			
Both Samples vs. Year 1 only			0.04	0.05	
p-value			0.44	0.33	
Year 2 only vs. Year 1 only			0.13*	-0.07	
p-value			0.03	0.29	
Joint p-value			0.08	0.24	
Ratio and Proportion Score		Both Samples vs. Year 1 only	0.07		
		p-value	0.06		
	Year 2 only vs. Year 1 only	0.02			
	p-value	0.70			
	Joint p-value	0.16			
	Both Samples vs. Year 1 only		0.02	0.07	
	p-value		0.62	0.20	
	Year 2 only vs. Year 1 only		0.11	-0.09	
	p-value		0.09	0.21	
	Joint p-value		0.23	0.10	

Sample Size = 76 schools (40 treatment, 36 control), 276 teachers (138 treatment, 138 control), 6,352 students (3,240 treatment, 3,112 control).

SOURCES: Spring 2008 and spring 2009 Teacher Knowledge Tests; Spring 2008 and spring 2009 NWEA Rational Number Tests.

NOTES: Coefficients are interactions between teacher knowledge and sample membership indicators. P-values for individual coefficients are based on t-tests, and joint tests of sets of coefficients are based on F-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

Relationships Among Teacher Knowledge and Student Achievement Using Four-Level Model

The model reported above took into account the nesting of students within teachers and teachers within schools. In this section, we examine one additional potential source of variation: variation among classes taught by the same teacher. To estimate the magnitude of the variation among classes, we estimated a four-level model, with students nested within classes, classes nested within teachers, and teachers nested within schools. First, we looked at the variance decomposition of this model. Table E-23 reports the findings from this analysis. As can be seen in this table, the class-level variance is significant and accounts for approximately 8 percent of the adjusted variance. These results suggest that students in different classes taught by the same teacher demonstrate significant differences in achievement, when prior achievement is controlled. We do not know what factors explain the variation among classes, but this variation indicates that teachers' effectiveness varied across the classes they taught. Second, we looked at the relationship between teacher knowledge and student achievement using the four-level model. Table E-24 displays these findings, which are similar to those in Table 4-4.

Table E-23. Variance Decomposition of Standardized Student Spring Total NWEA Test Scores by Data Structure Level, Four-Level Model: Pooled Sample

Level	<u>Benchmark</u>	<u>Model 1</u>	<u>Model 2</u>	<u>Model 3</u>	<u>Model 4</u>
	No Control Variables	Control for Block	Control for Block, School-, Student-Level Covariates	Control for Block, School-, Teacher-, Student-Level Covariates	Control for Block, School-, Teacher-, Student-Level Covariates, Also for TK Total Score
Total Variance	1.08	1.08	1.08	1.08	1.08
Adjusted Variance	1.08	0.99	0.65	0.64	0.64
School Level					
Variance	0.13*	0.04*	0.00	0.00	0.00
As a proportion of total variance	0.12	0.04	0.00	0.00	0.00
As a proportion of adjusted variance	0.12	0.04	0.01	0.01	0.01
P-value	<0.01	<0.01	0.30	0.33	0.39
Teacher Level					
Variance	0.04*	0.04*	0.02*	0.01*	0.01*
As a proportion of total variance	0.03	0.03	0.02	0.01	0.01
As a proportion of adjusted variance	0.03	0.04	0.03	0.02	0.02
P-value	<0.01	<0.01	0.02	0.03	0.04
Class Level					
Variance	0.10*	0.10*	0.05*	0.05*	0.05*
As a proportion of total variance	0.09	0.09	0.05	0.05	0.05
As a proportion of adjusted variance	0.09	0.10	0.08	0.08	0.08
P-value	<0.01	<0.01	<0.01	<0.01	<0.01
Student Level					
Variance	0.80*	0.80*	0.57*	0.57*	0.57*
As a proportion of total variance	0.75	0.75	0.53	0.53	0.53
As a proportion of adjusted variance	0.75	0.82	0.89	0.89	0.89
P-value	<0.01	<0.01	<0.01	<0.01	<0.01

Sample Size = 76 schools (40 treatment, 36 control); 276 teachers (138 treatment, 138 control); 6,352 students (3,240 treatment, 3,112 control).

SOURCES: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test.

NOTES: The variance components are estimated using four-level hierarchical linear models controlling for random assignment blocks and various sets of covariates as indicated in the table.

The NWEA total scale score is standardized using the distribution of the control group, which is 14.27 based on the first-year impact analysis sample.

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

Table E-24. Standardized Regression Coefficients for the Relationships Between Teacher Knowledge and Student Achievement, Using Four-Level Model: Pooled Sample

Standardized Outcomes	Mediating Variables in Model			
	TK Total Score	CK Score	SK Score	F-Test
NWEA Total Score	coefficient	0.05*		
	standard error	0.02		
	p-value	0.01		
	coefficient		0.03	0.03
	standard error		0.02	0.03
	p-value		0.21	0.32
Fractions and Decimals Score	coefficient	0.05*		
	standard error	0.02		
	p-value	0.00		
	coefficient		0.04	0.02
	standard error		0.02	0.03
	p-value		0.07	0.41
Ratio and Proportion Score	coefficient	0.04		
	standard error	0.02		
	p-value	0.05†		
	coefficient		0.02	0.03
	standard error		0.02	0.03
	p-value		0.51	0.30

Sample Size = 76 schools (40 treatment, 36 control); 276 teachers (138 treatment, 138 control); 6,352 students (3,240 treatment, 3,112 control).

SOURCES: Spring 2008 Teacher Knowledge Test; Spring 2009 Teacher Knowledge Test; Spring 2008 NWEA Rational Number Test; Spring 2009 NWEA Rational Number Test.

NOTES: Coefficients in the table are standardized regression coefficients. The coefficients were estimated based on a four-level model (students within classes within teachers within schools) controlling for random assignment blocks and school-, teacher-, and student-level covariates.

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

† P-value = 0.051, which rounds to 0.05 but is not statistically significant at the 0.05 level.