U.S. Department of Education
September 2016

Focusing on Mathematical Knowledge: The Impact of Content-Intensive Teacher Professional Development

Executive Summary

Michael S. Garet
Jessica B. Heppen
Kirk Walters
Julia Parkinson
Toni M. Smith
Mengli Song
Rachel Garrett
Rui Yang
American Institutes for Research

Geoffrey D. Borman
University of Wisconsin-Madison

Thomas E. Wei
Project Officer
Institute of Education Sciences

INSTITUTE OF EDUCATION SCIENCES

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE

This page has been left blank for double-sided copying.

Focusing on Mathematical Knowledge: The Impact of Content-Intensive Teacher Professional Development

Executive Summary

September 2016

Michael S. Garet
Jessica B. Heppen
Kirk Walters
Julia Parkinson
Toni M. Smith
Mengli Song
Rachel Garrett
Rui Yang
American Institutes for Research

Geoffrey D. Borman
University of Wisconsin-Madison

Thomas E. Wei
Project Officer
Institute of Education Sciences

NCEE 2016-4009
U.S. DEPARTMENT OF EDUCATION

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

This page has been left blank for double-sided copying.

U.S. Department of Education
John King
Secretary

Institute of Education Sciences
Ruth Curran Neild
Deputy Director for Policy and Research
Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance
Joy Lesnick
Acting Commissioner

September 2016

This report was prepared for the Institute of Education Sciences under Contract ED-IES-12-C-0080. The project officer is Thomas E. Wei 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 report.

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., Garrett, R., Yang, R., & Borman, G. D. (2016). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development, executive summary* (NCEE 2016-4009). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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

Alternate Formats: 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 202-260-9895 or 202-205-8113.

This page has been left blank for double-sided copying.

Acknowledgments

This study represents a collaborative effort among school districts, schools, teachers, researchers, and professional development providers. We appreciate the willingness and commitment of the school districts, schools, and teachers to volunteer for the study, participate in the professional development activities, and respond to requests for data, feedback, and access to classrooms. We benefited from advice provided by the study's expert technical working group. Members included Sybilla Beckmann, University of Georgia; Linda Davenport, Boston Public Schools; Skip Fennell, McDaniel University; Drew Gitomer, Rutgers University; Patrick Meyer, University of Virginia; Christopher Rhoads, University of Connecticut; Jeffrey Smith, University of Michigan; Jon Star, Harvard University; and Suzanne Wilson, University of Connecticut. We also received important feedback related to classroom video analysis from Catherine McClellan, Clowder Consulting.

We appreciate the dedication and close collaboration of our partners at Harvard University: Heather Hill, Daniel McGinn, Corinne Herlihy, and their team of classroom observation raters. We would also like to thank the many individuals who supported or provided the professional development during the study, including the Intel Math instructors and the Intel Math project manager, the Mathematics Learning Community trainers, the Harvard team, and the district-based facilitators. We are also grateful to the AIR staff who closely coordinated the study professional development activities in participating districts: Rachel Croninger, Jennifer Ford, Arayle Freels, Jeanette Moses, Cheryl Pruce, and Abigail Stein; the local videographers who participated in training and collected the classroom videos used in the study; AIR senior staff, Anja Kurki and Nicholas Sorensen, who led the study's recruitment efforts, and Seth Brown, who served as a senior recruiter; and AIR staff, Edith Tuazon and Zipporah Sanders, who provided administrative and communications assistance throughout the project. Finally, we would like to thank our report editor, Linda Schuch, who helped make the report usable and understandable.

This page has been left blank for double-sided copying.

Disclosure of Potential Conflicts of Interest

The research team for this study consisted of a prime contractor, American Institutes for Research (AIR), and two subcontractors, Harvard University and Measured Decisions Inc. None of these organizations or their key staff has financial interests that could be affected by findings from the Impact Evaluation of Math Professional Development. No one on the 9-member Expert Advisory Panel, convened by the research team twice to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

This page has been left blank for double-sided copying.

Executive Summary

Improving math achievement among U.S. students remains a high priority as results from recent math assessments continue to show room for improvement. For example, 60 percent of fourth-graders scored below the proficient level on the 2015 National Assessment of Educational Progress. On the most recent Program for International Student Assessment's math problem-solving test, U.S. 15-year-olds outperformed students in only 6 of the 34 participating countries.

In an era of increasingly rigorous state standards, teachers at all grade levels face heightened expectations to deepen their students' understanding of mathematical concepts. Teachers may thus benefit from professional development (PD) that deepens their own conceptual understanding of math. Elementary school teachers may especially benefit from content-focused PD because they are less likely to formally study math in college than secondary teachers, who tend to specialize in the subject matter they teach. Unfortunately, there is limited convincing evidence to date on the effectiveness of content-focused PD.

This report examines the impact of content-intensive PD on teachers' math content knowledge, their instructional practice, and their students' achievement. The study's PD had three components, totaling 93 hours. The core of the PD was *Intel Math*, an intensive 80-hour workshop delivered in summer 2013 that focused on deepening teachers' knowledge of grades K–8 mathematics. Two additional PD components totaling 13 hours were delivered during the 2013–14 school year: the *Mathematics Learning Community*, a series of five 2-hour collaborative meetings focused on analyzing student work; and *Video Feedback Cycles*, a series of three one-on-one coaching sessions where teachers' lessons were observed and critiqued. The purpose of these two components was to reinforce the math content in Intel Math and help teachers apply the content to improve their instruction.

Grade 4 teachers from 94 schools in six districts and five states participated in the study and were randomly assigned within schools to either a treatment group that received the study PD or a control group that did not receive the study PD. The key findings on the impact of the study PD on teacher knowledge, practice, and student achievement include:

- **The PD had a positive impact on teacher knowledge.** On average, treatment teachers' math knowledge scores on a study-administered math assessment were 21 percentile points higher than control teachers' scores in spring 2014, after the PD was completed.
- **The PD had a positive impact on some aspects of instructional practice, particularly *Richness of Mathematics.*** We assessed teachers on three dimensions of practice: *Richness of Mathematics*, which emphasizes the conceptual aspects of math, such as the use and quality of mathematical explanations; *Student Participation in Mathematics*, which focuses on student mathematical contributions, explanations, and reasoning; and *Errors and Imprecision*, which focuses on incorrect, unclear, and imprecise use of math. On average, treatment teachers had *Richness of Mathematics* scores that were 23 percentile points higher than the scores of control teachers in spring 2014, after the PD was completed. Although treatment teachers also had better average scores for *Student*

Participation in Mathematics and *Errors and Imprecision* than did control group teachers, these differences were not statistically significant.

- **Despite the PD's generally positive impact on teacher outcomes, the PD did not have a positive impact on student achievement.** On average, treatment teachers' students scored 2 percentile points lower than control teachers' students in spring 2014 on both a study-administered math assessment aligned with the content of the PD and the state math assessment. This difference was statistically significant for the state math assessment but not for the study-administered assessment. However, the state math assessment difference was not statistically significant in any of our sensitivity analyses.

Study overview

The study addressed the following research questions:

1. Was the study PD implemented with fidelity? What were the features of the PD as implemented? To what extent did teachers participate in the PD?
2. What was the impact on teachers' content knowledge, teachers' classroom practices, and student achievement, of offering content-focused PD relative to business-as-usual PD?

Study design and samples

The study sample included 221 grade 4 teachers from 94 schools who agreed to participate in the study. The schools were diverse, situated in urban, suburban, and rural settings across six districts and five states, and served students from a range of racial and socioeconomic backgrounds. The schools had self-contained classes in which the teachers taught multiple subjects, including math.

Random assignment of grade 4 teachers occurred separately within each school, generating a treatment group of 104 teachers and a control group of 117 teachers. As expected, there were no statistically significant differences between teachers and students in the two groups on any measured baseline characteristics. The final analysis sample included 165 of these 221 randomly assigned teachers from 73 schools, who remained in their schools teaching grade 4 through the study year, provided all outcome data described in the section on data sources, and taught in a school where at least one grade 4 teacher in the opposite treatment condition also had no missing outcome data. The percentage of randomly assigned teachers who were in the analysis sample was similar by condition (76 percent of treatment teachers, 74 percent of control teachers), and there were no statistically significant differences between teachers in the final analysis sample and teachers not in the final analysis sample on various baseline characteristics, including baseline teacher knowledge, years of teaching experience, level of education, certification status, and number of math courses taken. The grade 4 students in the classes of teachers in the analysis sample were the basis for the student samples.

Description of the PD program

The 93-hour PD program had three interrelated components:

- *Intel Math* (Intel Foundation, 2009), a widely used, 80-hour professional development workshop designed to promote deep understanding of the conceptual foundations and interconnectedness of grades K–8 mathematics topics through solving and discussing math problems. Intel Math is often used by school districts in federally funded Math-Science Partnership programs.
- *Mathematics Learning Community* (Regional Science Resource Center at the University of Massachusetts Medical School, 2011), a series of five collaborative meetings (10 hours total) in which teachers analyzed student work on topics covered in Intel Math.
- *Video Feedback Cycles,* three rounds of individualized, video-based coaching (3 hours total) that provided feedback to teachers on the quality and clarity of their mathematical explanations.

Intel Math was delivered to treatment teachers by experienced instructors (one mathematician and one math educator in each district) in summer 2013. In each district, the grade 4 treatment teachers were joined by about 10 teachers in grades K–3 and 5–8, to emulate typical implementation of Intel Math in which participants span across grades K–8. The five Mathematics Learning Community meetings were led by two trained, district-based facilitators in each district during the 2013–14 school year. Participants in the Mathematics Learning Community included the grade 4 treatment teachers and the grades 3 and 5 teachers who took part in Intel Math. Approximately 5 of the 10 grades K–3 and 5–8 teachers who participated in Intel Math taught grades 3 and 5. The district-based facilitators also delivered the Video Feedback Cycles, three one-on-one feedback sessions with grade 4 treatment teachers during the 2013–14 school year. The feedback was based on video excerpts of teachers' lessons on topics covered in Intel Math and the Mathematics Learning Community, coded with the Mathematical Quality of Instruction (MQI) instrument. The MQI focuses on three dimensions of instructional practice: *Richness of Mathematics* emphasizes the conceptual aspects of math, such as the use and quality of mathematical explanations; *Student Participation in Mathematics* focuses on student mathematical contributions, explanations, and reasoning; and *Errors and Imprecision* focuses on incorrect, unclear, and imprecise use of math.

Data sources

Data on the implementation of the PD program. We documented the delivery of each PD component with activity logs completed by the study team or PD facilitators, and we documented teacher participation in PD sessions with detailed attendance records. To describe the features of the PD components as delivered, we video-recorded all group sessions, including the 80 hours of Intel Math and the 10 hours of Mathematics Learning Community meetings in each district. We coded the videos to describe the activities in which instructors/facilitators and teachers engaged, and we coded a subsample of the videos using the MQI to assess the mathematical quality of the discussions. To describe the features of the individually delivered Video Feedback Cycles, we used the feedback forms completed jointly by the MQI raters and the district-based facilitators.

Data on outcome measures. We measured teacher knowledge at three time points: in the summer (at baseline), in the fall (after Intel Math), and in the spring (after the full PD program). We measured teacher knowledge using an adaptive assessment provided by the Northwest Evaluation Association (NWEA). Because the NWEA assessment was customizable, we were able to ensure that the content of the assessment aligned with the content of the PD. In particular, the assessment covered five mathematical domains emphasized by the PD: whole numbers; fractions; rational numbers; ratio, proportion, and rate; and linear equations and functions.

We measured instructional practice at two time points: in the fall (after Intel Math) and in the spring (after the full PD program). We measured instructional practice by video-recording participating teachers' lessons and using established procedures to score them on the three MQI dimensions of instructional practice. The first two dimensions, *Richness of Mathematics* and *Student Participation in Mathematics*, were scored on a four-point scale, with a score of 1 indicating no evidence of the practice. The other three possible scores ranged from 2 (low) if the practice was evident and had at least a basic level of quality to 3 (mid) and 4 (high) if the practice occurred with greater intensity and/or at a higher level of quality. The third dimension, *Errors and Imprecision*, was reverse coded on the same four-point scale, with the lowest score (1) being the most desirable because it indicated no errors and imprecision.

We measured student achievement in spring 2014 using an adaptive assessment provided by NWEA. We were able to customize the assessment to ensure that it focused on the mathematical domains covered in grade 4 and emphasized by the PD: namely, whole numbers, decimals, and fractions. All students took the same NWEA assessment, thus ensuring comparability in student outcomes. However, because the state mathematics assessment may be a policy-relevant outcome, we also collected and analyzed these scores for students in spring 2014.

Methods

We conducted descriptive analyses to assess the fidelity of PD implementation, examine the features of the PD components as implemented, and document treatment teachers' participation. We also compared treatment teachers and control teachers in their self-reported, math-related PD experiences during the study year, to determine the contrast in the amount and type of math-related PD experienced by the two groups of teachers.

We assessed the impact of the PD by comparing teacher and student outcomes between the treatment and control groups. Because the study used random assignment, any differences in teacher or student outcomes between the treatment and control groups can be attributed to the study PD and not some other characteristic of the districts, schools, or teachers.

Detailed summary of findings

The PD was well implemented with mathematical instructional quality evident most of the time, based on MQI scores. All three components of the PD were implemented with high fidelity. On average, 96 percent of the expected 80 hours of Intel Math, and 100 percent of the planned Mathematics Learning Community and Video Feedback Cycle hours were delivered. Mathematical instructional quality was evident during most of the whole-group discussion of math content and solution strategies. For example, MQI scores indicated that *Richness of Mathematics* was evident at a low, mid, or high level in 94 percent of the whole-group discussion in Intel Math (53 percent at a mid or high level), and in 93 percent of the whole-group discussion in the Mathematics Learning Community (45 percent at a mid or high level).

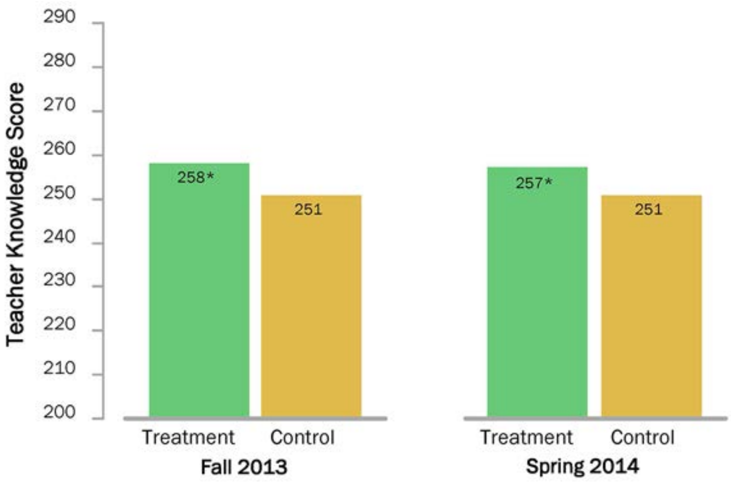
The PD provided extended time for teachers to solve math problems, analyze student work, explain their solutions to math problems, share their analyses of student work, and receive feedback. For example, 88 percent of the time in Intel Math and 84 percent of the time in the Mathematics Learning Community were spent in individual or small-group table work and whole-group discussions of the table work, where the above activities would occur. The video-based feedback provided to teachers emphasized the richness of mathematical presentations and discussions, with 82 percent of the feedback focused on this area. The feedback also focused to a lesser extent on identifying and addressing errors and instances of imprecision and lack of clarity, with the remaining 18 percent of feedback focused on this area.

Treatment teachers' participation in the PD was high, and the contrast between treatment and control teachers' math-related PD was considerable. On average, treatment teachers participated in more than 90 percent of the implemented hours for each component of the PD program (98 percent of Intel Math, 90 percent of the Mathematics Learning Community, and 97 percent of the Video Feedback Cycles). Treatment and control teachers differed substantially in both the amount and the type of math-related PD in which they participated during the year of the study. Overall, treatment teachers participated in 95 more hours of math-related PD than did control teachers, which is close to the approximately 93 hours of math PD provided by the study. Treatment teachers reported a greater focus on K–8 math content and student thinking in their workshop, study group, and feedback-related PD than did control teachers who reported participating in these types of math-related PD (average differences in reported focus were statistically significant, ranging from 0.7 to 1.3 on a 4-point qualitative survey scale).

The PD had a positive impact on teacher knowledge. The PD had a statistically significant impact on teachers' content knowledge in the fall, after the 80 hours of Intel Math were delivered but before the 13 hours of supports for enactment were delivered. This impact was largely sustained into the spring, after all 93 hours of the PD were delivered. Treatment teachers' average knowledge score was 7 points higher than control teachers' average score in the fall, and 6 points higher than control teachers' average score in the spring (see Exhibit ES.1). These differences correspond to an improvement of 24 percentile points in the fall and 21 percentile points in the spring.[1]

[1] More specifically, the math knowledge score for a typical control teacher would have increased from the 50th to the 71st percentile had the teacher received the study PD. This is referred to as the "improvement index," which is based on the outcome distribution within the control group (What Works Clearinghouse, 2014).

Exhibit ES.1. Teacher Knowledge Scores in Fall and Spring



Note: Sample size = 73 schools; 79 treatment teachers and 86 control teachers.

The teacher knowledge score is reported on the scale used by the test developer (Northwest Evaluation Association), which takes into account the difficulty of individual test questions in measuring teacher knowledge. The assessment is not typically given to adults; 11th graders are the oldest students for whom norming data are available. The scale shown ranges from 200, the score that corresponds approximately to the 1st percentile for 11th graders, to 290, the score that corresponds approximately to the 99th percentile for 11th graders. In the fall, the average scores correspond to the 84th percentile for treatment teachers and 74th percentile for control teachers. In the spring, the average scores correspond to the 82nd percentile for treatment teachers and 74th percentile for control teachers.

* Difference between the average treatment teacher score and the average control teacher score is statistically significant at the 0.05 level, two-tailed test.

Source: Fall 2013 and Spring 2014 Teacher Knowledge Tests.

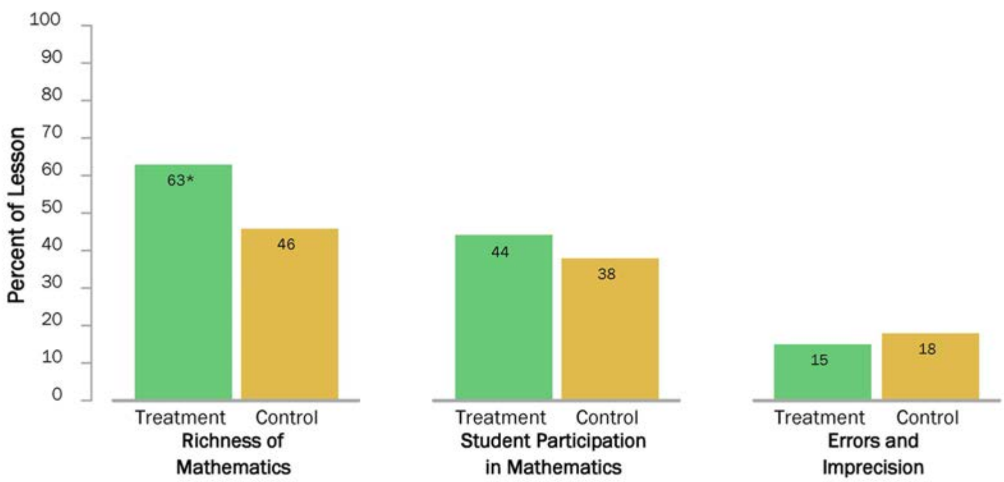
The PD's impact on teacher knowledge in the spring was larger for teachers with higher baseline knowledge. The impact of the PD on teachers' knowledge in the fall did not differ for teachers with different levels of prior math content knowledge. The estimated impact of the PD on fall teacher knowledge was positive and statistically significant for teachers with a baseline knowledge score 1 standard deviation above average (improvement index of 25 percentile points) as well as for teachers 1 standard deviation below average (improvement index of 23 percentile points).[2] However, the impact of the PD on teachers' knowledge in the spring was larger for teachers with higher baseline knowledge scores than for teachers with lower baseline scores. As in the fall, the estimated impact of the PD on spring teacher knowledge was positive and statistically significant for teachers with a baseline knowledge score 1 standard deviation above average (improvement index of 34 percentile points) but was not statistically significant for teachers with a baseline knowledge score 1 standard deviation below average (improvement index of 8 percentile points). This finding indicates that while Intel Math on average provided an initial boost to all teachers' content knowledge, the initial boost was not sustained for teachers who began the PD with lower levels of knowledge, even with the additional PD supports over the course of the school year.

The PD had a positive impact on some aspects of instructional practice, particularly *Richness of Mathematics*. The PD's effect on *Richness of Mathematics* in the spring was statistically significant and

[2] Teachers' baseline knowledge scores were standardized using the control group mean and standard deviation within the teacher analysis sample.

positive. An average treatment teacher demonstrated *Richness of Mathematics* at a mid or high level during 63 percent of a typical lesson, compared with 46 percent for an average control teacher (see Exhibit ES.2).[3] This 17 percent difference corresponds to an improvement of 23 percentile points. The impact on *Student Participation in Mathematics* and *Errors and Imprecision* in the spring was in the expected direction but not statistically significant (improvement of 6 and −9 percentile points; note that a negative impact on *Errors* is expected because it corresponds to a decrease in *Errors and Imprecision*). We also observed a statistically significant impact on *Student Participation in Mathematics* in the fall (after Intel Math but before the other two PD components) corresponding to an improvement index of 11 percentile points, but it was not sustained into the spring. Impacts on the other two dimensions of practice were in the expected direction in the fall but not statistically significant. In contrast to the teacher knowledge impacts, the impacts on instructional practice did not vary based on teachers' prior math knowledge.

Exhibit ES.2. Percentage of an Average Teacher's Lesson Demonstrating Three Dimensions of Mathematical Quality of Instruction in Spring



Note: Sample size = 73 schools; 79 teachers, 158 lessons, and 1,277 7.5-minute lesson segments for the treatment group; 86 teachers, 172 lessons, and 1,352 7.5-minute lesson segments for the control group.

The graph shows the percentage of a typical lesson in which an average treatment or control teacher demonstrated each of the three MQI dimensions of instructional quality in spring 2014. Demonstrating *Richness* or *Student Participation* is defined as scoring mid or high on one or more of the elements that comprise the dimension. Demonstrating *Errors* is defined as scoring present (low, mid, or high) on one or more of the elements that comprise the dimension. *Richness of Mathematics* emphasizes the conceptual aspects of math, such as the use and quality of mathematical explanations; *Student Participation in Mathematics* focuses on student mathematical contributions, explanations, and reasoning; and *Errors and Imprecision* focuses on incorrect, unclear, and imprecise use of math. Lower error and imprecision scores are desirable and indicate fewer content errors and less imprecision than higher scores.

* Difference between the average treatment teacher percentage and the average control teacher percentage is statistically significant at the 0.05 level, two-tailed test.

Source: Mathematical Quality of Instruction (MQI) scores of video-recorded lessons from spring 2014.

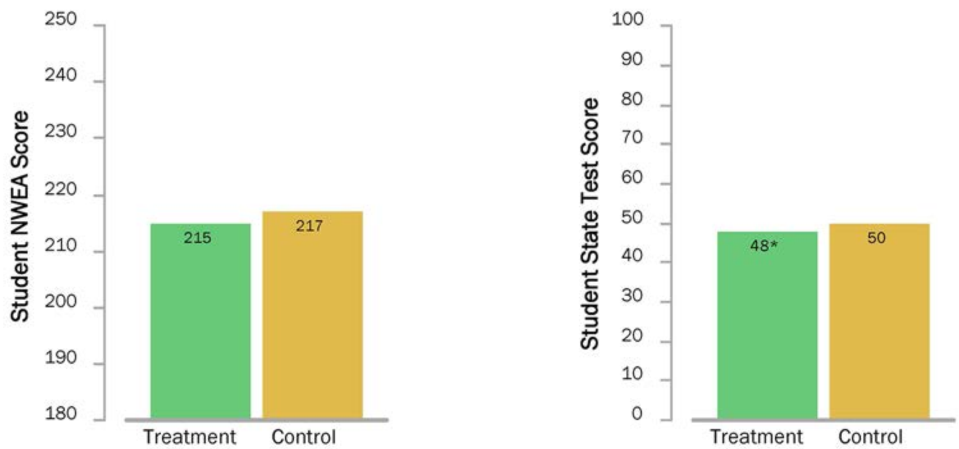
Despite the PD's generally positive impact on teacher outcomes, the PD did not have a positive impact on student achievement. On average, treatment teachers' students scored 2 percentile points lower than

---

[3] By "demonstrated *Richness of Mathematics*," we mean that the teacher was rated mid or high on one or more of the seven elements that comprise the *Richness of Mathematics* dimension.

control teachers' students on both spring 2014 student achievement measures, including the study-administered math assessment aligned with the content of the PD and the state-specific math assessment (see Exhibit ES.3). The difference between treatment and control group students was statistically significant for the state math assessment but not the study-administered assessment.[4] The results were similar for students who had higher or lower prior achievement, and were also similar for students whose teachers had higher or lower baseline knowledge or more or less teaching experience.

Exhibit ES.3. Student Math Scores in Spring



Note: Sample size for analysis of NWEA scores = 73 schools; 79 teachers and 806 students in the treatment group; 86 teachers and 891 students in the control group. Sample size for analysis of state test scores = 73 schools; 79 teachers and 1,760 students in the treatment group; 86 teachers and 1,917 students in the control group. Student sample sizes are smaller for the NWEA assessment because we administered the assessment to a random subsample of students in each class.

The NWEA score is reported on the scale used by the test developer, Northwest Evaluation Association, which takes into account the difficulty of individual test items in measuring student achievement. The scale shown ranges from 180, the score that corresponds approximately to the 1st percentile for fourth graders, to 250, the score that corresponds approximately to the 99th percentile for fourth graders.

The state score is reported using Normal Curve Equivalent (NCE) scores. NCE scores measure a student's position on the normal curve, relative to other students in their state. NCE values run from 0 to 100. They are similar to percentile ranks, but on an equal-interval scale.

* Difference between the average treatment student score and the average control student score is statistically significant at the 0.05 level, two-tailed test.

Source: District administrative records; Spring 2014 NWEA Test.

Teacher knowledge and instructional practices were generally not correlated with student achievement. The conceptual framework underlying the study PD assumed that teachers' content knowledge is related to instructional practice, which in turn is related to student achievement. Contrary to these assumptions, both knowledge and instructional practice as measured in the study were not statistically significantly associated with student achievement (estimates of association between 0.00 and −0.05). The only teacher measure associated with student achievement was the *Errors and Imprecision* dimension, which was statistically significantly related to student achievement in the expected direction (estimate of association −0.20).

[4] However, the statistically significant impact on state math assessment scores was sensitive to sample definition and the inclusion of covariates—it was not statistically significant in any of our sensitivity analyses.

Concluding thoughts

Together these results show that the study PD did change some aspects of teachers' knowledge and classroom practice, but not in a way that led to improved student achievement. This may be partially explained by our finding that the math content knowledge and dimensions of instructional practice targeted by the study PD were generally not correlated with student math achievement. The one exception was *Errors and Imprecision*, on which the study PD did not have a statistically significantly impact. Thus, future research might focus on identifying PD that will improve this aspect of practice. Future research might also seek to identify other aspects of knowledge and practice to target with PD that are more strongly related to improved student achievement.