Professional Development for Teachers

What Two Rigorous Studies Tell Us

Janet Quint

July 2011

Funding for this report was provided by the Carnegie Corporation of New York.

Dissemination of MDRC publications is supported by the following funders that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Ambrose Monell Foundation, The Annie E. Casey Foundation, Carnegie Corporation of New York, The Kresge Foundation, Sandler Foundation, and The Starr Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

The findings and conclusions in this report do not necessarily represent the official positions or policies of the funders.

For information about MDRC and copies of our publications, see our Web site: www.mdrc.org.

Overview

Professional development — formal in-service training to upgrade the content knowledge and pedagogical skills of teachers — is widely viewed as an important means of improving teaching and learning. While many interventions *include* professional development, professional development *was* the central intervention of the two recent research and demonstration projects — the Professional Development in Reading Study (the "Reading PD study," for short) and the Middle School Mathematics Professional Development Impact Study (the "Math PD study") — whose findings are synthesized in this report. The studies were carried out by the American Institutes for Research and MDRC for the U.S. Department of Education. The professional development that was provided went far beyond the "one-shot" workshop approach that has been widely criticized; it instead included intensive summer institutes, follow-up group sessions, and coaching of individual teachers. The evaluations of the interventions employed random assignment design, and, as a result, they supply unusually rigorous evidence about the effects of the professional development that was offered both on instruction and on student achievement.

The impacts of both interventions were substantially less positive than had been hoped. The Reading PD study increased teachers' content knowledge; the Math PD study did not. In both studies, the professional development had positive effects on some targeted instructional practices but not on others. Most critically, students of teachers who received the training scored no higher on subject-matter achievement tests than students of teachers who did not receive the training. Moreover, in the reading study, professional development that included one-on-one coaching as well as group workshops did not lead to significantly larger impacts than professional development involving just the workshops; in the mathematics study, receiving two years of professional development did not lead to better results than receiving just one year.

A number of factors likely reduced the effectiveness of the professional development and the researchers' ability to measure that effectiveness. For example, teacher turnover in the Math PD study meant that many teachers did not receive the full dose of professional development that had been planned. And the two-year time frames of the two studies may not have allowed enough time for major changes in teaching and learning to take hold.

Nonexperimental analyses that were conducted as part of these two studies, along with other research, suggest that the theory of change underlying the studies is correct: professional development of the type that was delivered is associated with increased teacher knowledge and that teacher knowledge and improved instruction is associated with higher student test scores. But changes in teacher-related variables must be substantial — considerably larger than they were in these studies — to move the needle on student achievement even a small amount.

By themselves, the findings of the two studies do not mean that professional development efforts *cannot* work. New thinking emphasizes a broader conception of teacher learning that involves all teachers in a school in a professional learning community that is engaged in a continuous and collegial cycle of learning, practice, reflection, and improvement. Randomized trials to test professional development that is reinforced within professional learning communities are in order. At the same time, in-service training should not be the only vehicle for improving student achievement.

Contents

List of Exhibits

Table

Figure

Preface

Professional development for teachers — in-service training for the teaching force that is already in place — has become a widely accepted approach for improving teaching and learning in America's schools. But there have been few rigorous large-scale evaluations of the effectiveness of this strategy. The Professional Development in Reading Study and the Middle School Mathematics Professional Development Impact Study discussed in this report are exceptions to this rule.

Together, the studies included almost 170 elementary schools and middle schools, which were randomly assigned to treatment and control conditions. Second-grade reading teachers and seventh-grade math teachers in the studies' treatment group schools received intensive professional development related to these subjects, while their counterparts received the professional development usually offered by their districts. The random assignment helped to ensure that the studies would provide the strongest possible evidence about the role of professional development in improving instructional practices and boosting student achievement.

This report reviews the findings of the two studies and reflects on their meaning. It offers an important caution to educators and policymakers: Professional development cannot be counted on to improve outcomes for students. In both the studies examined here, the professional development — which went far beyond the "one-shot" approach that has been widely decried — had only limited effects on teachers' knowledge and instruction and did not have an impact on student test scores.

This does not mean that professional development cannot work, but only that the professional development tested here did not work. As the field advances, new approaches to promoting professional learning among teachers and learning among students must continue to be developed and tested, in the continuing search to improve the educational prospects of America's children. After all, for the next decade or more, our children will go to school with the teaching force that is in place now. Given the central role that high-quality teaching must play in our efforts to make the nation's schools more effective, new strategies for improving teacher quality will be essential.

Gordon L. Berlin
President

Acknowledgments

The comments of several reviewers on earlier drafts of this report greatly strengthened the final product. The author is especially grateful to Michael S. Garet of the American Institutes for Research (AIR) for his painstaking and detailed critique. Within MDRC, Gordon Berlin, William Corrin, Fred Doolittle, John Hutchins, and Robert Ivry made many helpful suggestions. Robert Weber edited the report with his usual tact and skill. Mario Flecha was responsible for initial versions of the tables and figure, and David Sobel prepared the report for publication.

The Author

Introduction

There is broad agreement that the educational prospects of America's children largely depend on the quality of the nation's teaching force. There is less consensus, however, on how to ensure that the best teachers are teaching our children. Some experts advocate the use of financial and other incentives and of alternative teacher preparation pathways to attract bright young people and those interested in making a midlife career change into teaching. Some call for improving the preservice training that would-be teachers receive in colleges and universities, and especially in these institutions' schools of education. Some administrators favor the dismissal of teachers who are no longer effective (or perhaps were never effective to begin with). And many policymakers and practitioners support the use of professional development — formal in-service training, often delivered by outside experts — to upgrade the content knowledge and pedagogical skills of the teaching force that is now in place.[1]

Two recent evaluations — the Professional Development in Reading Study (referred to here as the "Reading PD study") and the Middle School Mathematics Professional Development Impact Study ("Math PD study," for short) — supply unusually rigorous evidence about professional development for teachers as a strategy for improving teaching and learning in these two areas.[2] In many studies, professional development has been an important accompaniment to the main intervention being tested (a new curriculum, for example, or a change in school structure); in these two evaluations, professional development *was* the intervention tested. The evaluations were conducted by the American Institutes for Research (AIR) and MDRC for the Institute of Education Sciences (IES) in the U.S. Department of Education. A distinctive feature of the evaluations is their use of random assignment experiments — the gold standard of research designs — to provide highly credible findings about the impacts of this professional development in increasing the reading achievement of second-graders and the math achievement of seventh-graders.

IES contracted with the two organizations because prior studies of the effects of professional development had yielded unreliable and/or inconclusive results. Although literally hundreds of such studies had been conducted, a comprehensive literature review found that only five of these had employed robust random assignment designs that yielded unequivocal findings about the causal effects of professional development on student outcomes.[3] All five studies

[1]Of course, these strategies are not mutually exclusive. Rather, they represent different foci of attention.

[2]The full reports from these studies may be found at http://ies.ed.gov/ncee and at www.mdrc.org. Please see Garet et al. (2008); Garet et al. (2010); and Garet et al. (2011).

[3]Yoon et al. (2007). A sixth study also involved a randomized controlled trial in which five teachers were randomly assigned; two received the professional development, while three did not. However, because students were not randomly assigned across the teachers' classrooms, there was no way of assuring that students were similar across the treatment and control conditions at baseline — a precondition for a rigorous test. The (continued)

focused on students in the elementary grades; the amount of professional development tested ranged between 3 and 40 hours. Collectively, the studies included 14 measures of student achievement in reading and math. Along all 14 measures, students whose teachers had received the professional development had higher test scores than students whose teachers had not gotten this training, but in the large majority of cases (9 of the 14), these differences are not statistically significant — that is, they could have arisen by chance.[4] The studies also provided few clues about the characteristics of effective professional development, although they did suggest that interventions that provided less than 30 hours of training did not affect student learning.

Mindful of these issues, IES funders and AIR-MDRC evaluation team members were guided by two primary considerations. First, they wanted to ensure that the professional development would be intensive and well designed and that it would be implemented as designed. Second, they wanted the evaluations to yield hard evidence about the causal role of professional development in improving student achievement.

The findings that emerge from the Reading PD and Math PD studies are, however, mixed at best. They suggest that professional development is not necessarily the "royal road" (or an easy path) to better student outcomes. Most critically, the interventions did not achieve their ultimate goals: Students of teachers who received the training scored no higher on subject-matter achievement tests than students of teachers in the control group. More proximal impacts were also limited: In only one of the two studies did the group of teachers who received the professional development ("program group teachers") have a significantly higher overall score on a test of content knowledge than teachers who did not receive the professional development ("control group teachers"), and, in both studies, the professional development had positive effects on some targeted instructional practices but not on others. Moreover, in the reading study, professional development that included intensive one-on-one coaching as well as group workshops did not lead to better results than professional development involving the workshops alone. And in the mathematics study, receiving two years of professional development did not lead to better results than receiving just one year.

While the findings themselves are straightforward, their interpretation is much less clear. Although the evaluations rank among the most carefully executed studies of the effects of professional development that have been conducted to date, high levels of staff and student turnover and other issues may have lessened the likelihood of detecting statistically significant effects. The findings also raise questions about whether the underlying theory of change is the

---

literature review also identified three additional studies that were judged to have reasonably strong research designs that, however, lack the rigor of true experiments.

[4]One study yielded some results that are statistically significant and others that are not, depending on the measure used.

right one and whether a different approach to professional development might make more of a difference.

Two related points should be emphasized at the outset. First, the evaluations tested the effects of a particular kind of professional development. That professional development, in format and intensity, went well beyond the "one-shot" lectures and workshops that are widely decried (but that nonetheless continue to be offered because of their low cost). In several respects, the professional development differed from what some experts have come to believe is a more effective model, as discussed below. Second, while the results indicate that much has yet to be learned about how best to deliver professional development and to measure its effects, by themselves the findings of the two studies do *not* mean that professional development efforts cannot work and should be abandoned as a means of improving teaching and student achievement. While other strategies toward that end should be deployed and tested, professional development might instead be reconceptualized as part of a larger professional learning strategy.

The rest of this synthesis report explores these ideas. The next section reviews the assumptions underlying the demonstration and the theory of change guiding the evaluation. Then the section "Testing the Theory" compares and contrasts the research designs of the studies, and that is followed by a section describing the professional development that teachers in each study received. The section "Summary of the Impacts" reviews the findings and is followed by a section that focuses on possible explanations for the results. The report's final section considers the possible implications for providers of professional development and for program evaluators.

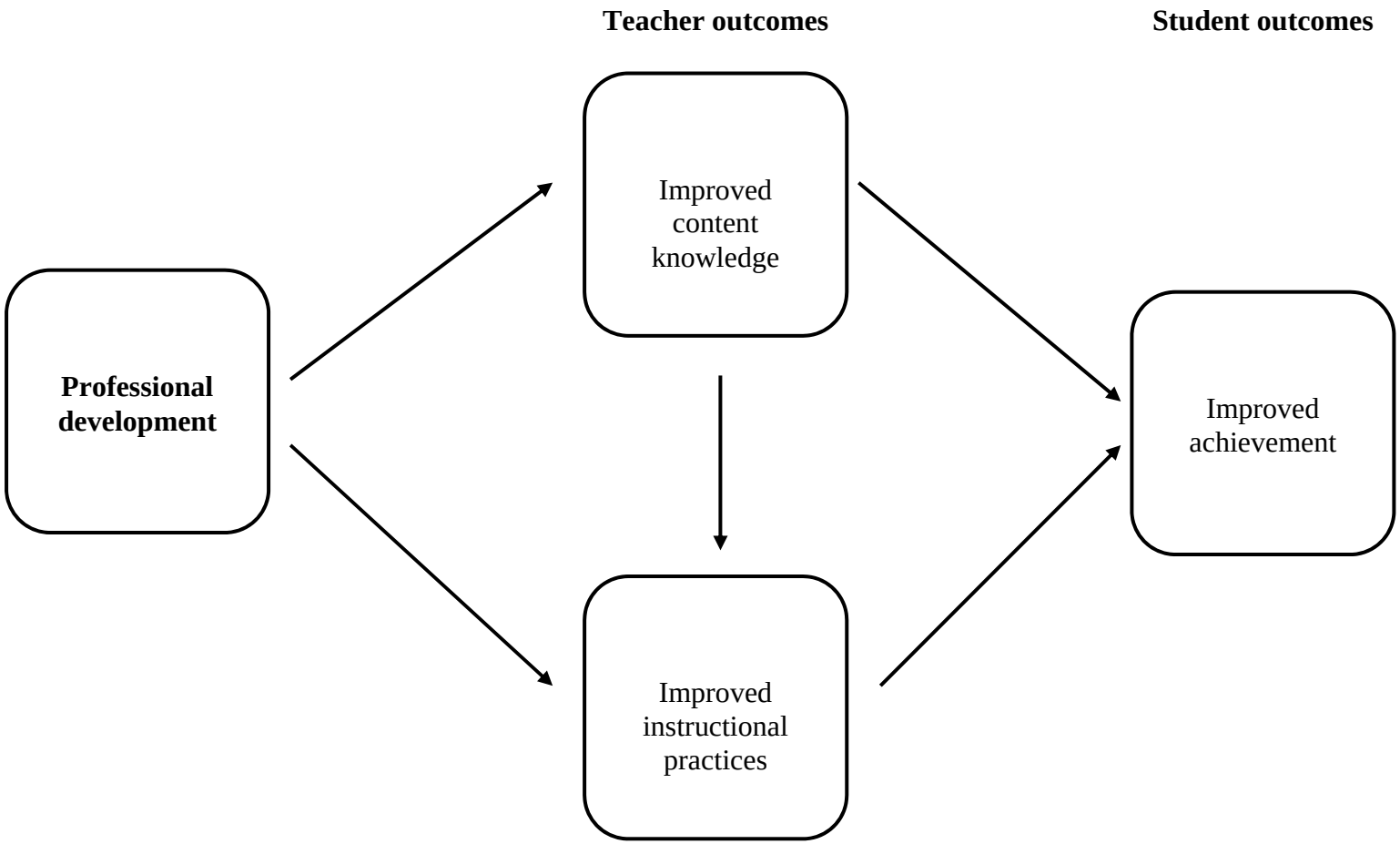
The Theory of Change and Design Choices Underlying the Demonstrations

Figure 1 depicts in simplified form the theory of action underlying the two demonstrations. The theory hypothesizes that professional development will improve both teachers' content knowledge and their instructional practices. As a result of improved instruction, students' achievement will also improve, as measured by scores on tests measuring their reading skills (in the Reading PD study) or their ability to solve mathematical problems involving rational numbers (in the Math PD study). The data collected and analyzed for the Reading PD and Math PD studies relate to the successive stages in this theory of action.

Educators have debated whether professional development should focus on increasing teachers' subject-matter knowledge or on enhancing their repertory of instructional techniques. A premise of the Reading PD and the Math PD demonstrations was that teachers needed both content knowledge and pedagogical skills to convey that content more effectively. In the Reading PD study, second-grade teachers learned about the essential components of early

Figure 1

Theory of Change Underlying the Professional Development Evaluations



reading instruction that were identified by the National Reading Panel — including phonemic awareness, phonics, fluency, vocabulary, and comprehension.[5] The second-grade reading teachers also received training on differentiating instruction and analyzing students' work. In the Math PD study, the professional development for seventh-grade teachers centered on topics in rational numbers with which students often struggle: fractions, decimals, ratios, rates, proportions, and percentages.[6] With respect to each topic, the professional development for math teachers covered two aspects of content knowledge: the understanding of rational numbers and computational skills that students should have after completing the seventh grade (referred to as "common knowledge" of mathematics, or "CK") and the specialized knowledge that could help teachers impart such understanding and skills (termed "SK"). The professional development that the teachers received in both demonstrations was designed to be relevant to the reading and math programs used by their districts. These programs were, in fact, selected for the demonstration because they are in wide use across the country.[7]

The Reading PD study addressed open questions that remained about *how* professional development should be provided. In particular, some practitioners and program developers maintained that the knowledge that teachers obtained in workshops and seminars needed to be reinforced periodically by expert coaching. Coaching is resource-intensive and expensive, however, and there was little rigorous evidence about its effectiveness. As discussed below, the Reading PD study was explicitly intended to fill this gap by comparing two different versions of professional development — one with coaching and one without it.

Testing the Theory

Table 1 shows the key features of the two professional development studies. The Reading PD study began during the summer of 2005 with a teacher workshop and continued over the course of the 2005-2006 school year. Summer workshops that marked the beginning of the Math PD study began two years later, in the summer of 2007. In that demonstration, professional development continued through the 2007-2008 school year and, in half the sites, through the 2008-2009 school year as well.

[5]National Institute of Child Health and Human Development (NICHD) (2000).

[6]Seventh grade was selected as the target grade for the math demonstration because this is typically the last year of formal instruction in rational numbers before students move on to pre-algebra.

[7]The two reading programs were SRA/McGraw-Hill's *Open Court Reading* and Houghton Mifflin's *Legacy of Literacy/The Nation's Choice.* Six districts in the Math PD study used either Glencoe/McGraw-Hill's *Mathematics: Applications and Concepts* or Prentice Hall's *Mathematics,* while the other six used Prentice Hall's *Connected Mathematics.* Study districts had to have been using the specified curriculum for at least two years prior to the inception of the evaluation.

Table 1

Key Features of the Reading Professional Development and Math Professional Development Studies

		Math PD Study	
Key Feature of Study	Reading PD Study	Receipt of PD for 1 Year	Receipt of PD for 2 Years
Impact research design	School-based random assignment	School-based random assignment	School-based random assignment
Implementation year (when PD was delivered)	Summer 2005 School year 2005-2006	Summer 2007 School year 2007-2008	Summer 2007, 2008 School year 2007-2008 School year 2008-2009
Number of districts	6	12	6
Number of schools	90	77	39
Number of teachers	270	195	92
Number of students	5,530		
What was tested	1 year of PD consisting of group workshops and institutes vs. 1 year of PD consisting of group workshops and institutes plus coaching vs. 1 year of whatever PD the control group received	1 year of PD consisting of group workshops and institutes plus coaching vs. 1 year of whatever the PD control group received	2 years of PD consisting of group workshops and institutes plus coaching vs. 1 year of PD consisting of group workshops and institutes plus coaching vs. 2 years of whatever PD the control group received
Length of follow-up	2 years	1 year	2 years
Data sources and when data were collected			
Receipt of PD	Surveys at end of implementation, follow-up years	Survey at end of (single) implementation year	Survey at end of each of 2 implementation years
Teacher knowledge	Pretest before summer institute Posttest at end of implementation, follow-up years	Pretest before summer institute Posttest at end of (single) implementation year	Pretest before summer institute Posttest at end of each of 2 implementation years
Teacher's instructional practices	Classroom observations during implementation, follow-up years	Classroom observation during (single) implementation year	Classroom observation during first implementation year only
Student achievement	Standardized tests at end of implementation, follow-up years	Specially developed test at end of (single) implementation year	Specially developed test at end of each of 2 implementation years

The research design in both demonstrations called for entire schools that served large proportions of low-income students to be randomly assigned to treatment and control conditions. Random assignment of individual teachers within a school was seriously considered — it had the advantage of allowing for a larger research sample — but this idea was rejected. Planners both feared and hoped that, within treatment group schools, teachers would talk with their control group counterparts about the professional development that they were receiving, thereby undermining the distinctiveness of the research conditions; planners also hoped that school-level random assignment would lead to more buy-in and support for the professional development in the treatment group schools.

Both demonstrations ended up examining two variations on the professional development theme. From the outset, as noted above, the Reading PD study was intended to test not only the general theory of action but also, and more specifically, the added value of coaching for teachers over and above the group workshops and seminars that the teachers attended. Thus, each of the 90 schools in six districts that participated in the Reading PD study was randomly assigned to one of three groups:

- **An "institute and seminar" program (or "treatment") group.** Second-grade reading teachers at schools in this group participated in a summer institute and several one-day seminars over the course of the next school year.
- **An "institute, seminar, and coaching" program (a second treatment group).** Teachers at schools in this group received not only the summer institute and seminars but also multiple sessions of individual or group coaching.
- **A control group.** Teachers at schools in this group received only the professional development normally provided by the districts in which they taught.

At the beginning of the Math PD study, 77 schools in 12 districts were randomly assigned to one of two groups:

- **A treatment group.** Seventh-grade math teachers in the treatment group schools in all the districts received the same treatment — a year of professional development, consisting of a summer institute, a series of one-day follow-up seminars during the school year, and coaching.
- **A control group.** Teachers in control group schools received the usual professional development provided by their districts. In this study, the variation concerned the length of the treatment.

Midway through the year (and before the first-year results were in), IES decided to test the effects of a second year of professional development for math teachers in half the districts,

which were selected in large part because they were able and willing to participate for a second year. In the remaining six districts, professional development ended after the first year.

The two studies measured outcomes in the four areas suggested by the theory of change: receipt of professional development, teacher knowledge, instructional practices, and student outcomes. Data on these outcomes came from similar sources in the two studies. Teachers in the program and control groups in both studies completed surveys about the amount and content of the professional development that they had received. Teacher knowledge was measured through tests, starting with a pretest before teachers first participated in the summer institutes. In the Reading PD study, posttests were administered at the end of the implementation and follow-up years; in the Math PD study, they were given at the end of the first year for teachers in all 12 districts and again at the end of the second year for teachers in the districts where two years of professional development were provided. In both studies, classroom observations (although limited in number because of resource constraints) were used to record instructional practices.[8] Finally, in the Reading PD study, standardized assessments administered by the six districts were used to measure student achievement, while in the Math PD study, the test used for this purpose was especially developed for the evaluation.

The basic analytic strategy in each study was to compare outcomes for schools randomly assigned within each district to program and control conditions. Two-level models (teachers nested within schools) were used to estimate impacts on teacher knowledge and practices, while three-level models (students nested within teachers' classrooms and classrooms nested within schools) were used to estimate impacts on student achievement.

The Professional Development That Was Delivered

In order to foster a sense of collective participation, in both demonstrations, invitees to the professional development from the treatment schools included the schools' principals, teachers in the appropriate grades, special education and/or teachers of English Language Learners working in these grades, and subject-area specialists (the lead reading teacher in the Reading PD study and the math department chair in the Math PD study). Teachers and specialists generally attended the professional development sessions; the principals' participation was more sporadic.

The professional development that was offered went well beyond what teachers would otherwise have received in their districts. In the Reading PD evaluation, the professional development for both treatment groups involved eight full days of content-focused institutes and seminars, which were offered during the summer of 2005 and the 2005-2006 school year. The

[8]In the Reading PD study, all second-grade classrooms were observed three times: during the fall and spring of the implementation year and during the fall of the follow-up year. The Math PD study involved one observation per classroom, conducted during the first year of the study.

topics covered were relevant to second-grade reading instruction and included phonemic awareness, phonics, fluency, analyzing student work, vocabulary, reading comprehension, and differentiated instruction. In addition, in schools assigned to the second treatment group, teachers were provided with a coach who worked with the school on a half-time basis. Coaches received training for their roles, and it was expected that teachers would receive, on average, 60 hours of group and individual coaching during the school year.[9]

During the first year of the Math PD evaluation, the study-provided professional development for seventh-grade math teachers at the treatment group schools included a three-day summer institute, a series of five one-day follow-up seminars held during the school year, and ten days of within-school coaching conducted in association with the seminar days and delivered by the seminar trainers. During the second year, the professional development was scaled back to two days of summer institutes, three seminar days, and eight days of in-school individual and group coaching, along with a special two-day "make-up" for teachers who joined the study after the first-year summer institute. The institutes and seminars included several opportunities for teachers to solve mathematics problems individually and in groups, explain how they solved them, and receive feedback on the solutions and their explanations. Teachers also discussed student misconceptions associated with rational numbers and planned lessons that they would teach during the coaching visits. The coaching visits were designed to help teachers apply what they had learned in the seminars to their classroom instruction.[10]

Summary of the Impacts

Tables 2 and 3 show key impact findings for the Reading PD and Math PD studies, respectively, along measures of each of the variables in the theory of change. Measures for which there is a treatment-control difference favoring the treatment group that is statistically significant (that is, with a probability of 1 in 20 or less of having arisen by chance) are marked by an "X," and outcomes for which there is no statistically significant difference are indicated by double hyphens ("--"). "NA" signifies that the outcome was not measured for a particular research group or in a given year.[11]

---

[9]The teacher institute and seminar series for the Reading PD study were based on the *Language Essentials for Teachers of Reading and Spelling (LETRS®)* professional development curriculum developed by Louisa Moats of Sopris West Educational Services (Moats, 2005) and were delivered by LETRS facilitators from Sopris West. The coaching was delivered by facilitators at the Consortium on Reading Excellence.

[10]Two organizations provided professional development in the Math PD study: America's Choice and Pearson Achievement Solutions.

[11]Information about the precise magnitude of treatment-control differences, effect sizes, significance levels, and other statistics is available in the reports on which this synthesis is based (Garet et al., 2008; Garet et al., 2010; Garet et al., 2011). This synthesis report follows the practice of the earlier report in its treatment of findings; it considers these findings to be nonetheless potentially meaningful when the earlier reports did so.

Table 2

The Reading Professional Development Study: Summary of Impacts

Impact Area	Impacts After Implementation Year		Impacts After Follow-Up Year	
	Institutes and Seminars	Institutes, Seminars, and Coaching	Institutes and Seminars	Institutes, Seminars, and Coaching
Receipt of professional development				
Hours of reading seminars and institutes	X	X	NA	NA
Hours of coaching	--	X	NA	NA
Teacher knowledge				
Total score	X	X	---	---
Word-level knowledge[a]	X	X	---	---
Meaning-level knowledge[b]	---	---	---	---
Instructional practices				
Teacher uses explicit instruction	X	X	---	---
Teacher encourages independent student activity	---	---	---	---
Teacher uses differentiated instruction	---	---	---	---
Student test scores	---	---	---	---

NOTES: All impacts compare outcomes for the specified treatment with those for the control group. "X" indicates that there was an impact that is statistically significant at the level of 5 percent or less; "- -" indicates that the impact is not statistically significant at the level of 5 percent or less. "NA" indicates that the variable was not measured.

[a]Word-level knowledge includes the areas of phonemic awareness, phonics, and fluency.

[b]Meaning-level knowledge includes the areas of vocabulary and comprehension.

In both studies, the two-year outcomes were measured for all teachers present at the end of the second year in the study schools, regardless of the length of time that they had been at those schools or how much — if any — professional development the teachers at the treatment group schools had received. In evaluation parlance, the analyses register the effects of the "intent to treat," rather than the effects of the "treatment on the treated."

The Reading PD Study

Receipt of professional development in reading. As expected, there were marked differences between teachers in both of the treatment groups and their control group counterparts in the amount of professional development in reading that they received through institutes and seminars during the implementation year and the summer that preceded it. There was no difference between the two treatment groups in this regard. However, teachers in the second

Table 3

The Math Professional Development Study: Summary of Impacts

Impact Area	Impacts After 1 Year for All Districts	Impacts After 2 Years for 2-Year Districts
Receipt of professional development		
Hours of math-related PD	X	X
Hours of math seminars and institutes	X	X
Hours of coaching	---	X
Teacher knowledge		
Total score	---	---
Common knowledge of math	---	---
Specialized knowledge of math for teaching	---	---
Instructional practices		
Teacher elicits student thinking	X	NA
Teacher uses representations	---	NA
Teacher focuses on mathematical reasoning	---	NA
Student test scores	---	---

NOTES: All impacts compare outcomes for the specified treatment with those for the control group. "X" indicates that there was an impact that is statistically significant at the level of 5 percent or less; "- -" indicates that the impact is not statistically significant at the level of 5 percent or less. "NA" indicates that the variable was not measured.

treatment group, as planned, received many more hours of coaching than either teachers in the first treatment group or control group teachers.

Receipt of professional development was not measured during the follow-up year, when it was assumed that all teachers would get the professional development that their districts arranged for them.

Teacher knowledge. Both at the outset and at the end of the implementation and follow-up years of the Reading PD study, teacher knowledge was assessed through the Reading Content and Practices Survey (RCPS), which was developed specifically for the study to assess teachers' knowledge of reading instruction and which, accordingly, emphasized topics relevant to second-grade reading. The test yielded an overall score and scores on two subscales measuring word-level knowledge (phonemic awareness, phonics, and fluency) and meaning-level knowledge (vocabulary and comprehension).

In the spring of the implementation year, teachers in both treatment groups had significantly higher overall scores than teachers in the control group: 57 percent of the teachers in the two treatment groups gave a correct answer to a typical item on the assessment, compared with 51 percent of their control group counterparts. Teachers in the treatment groups also had significantly higher scores on the subscale measuring word-level knowledge but not on the subscale measuring meaning-level knowledge. No statistically significant differences were registered between teachers in the two treatment groups.

At the end of the follow-up year, statistically significant treatment-control differences were no longer evident on any of the knowledge measures. Another way of stating this is that the follow-up year impacts, though favoring the treatment group, are too small to be deemed statistically significant, given the available sample size. One cannot conclude that treatment group teachers forgot what they had learned during the implementation year, however, because implementation-year and follow-up-year treatment-control impacts are not significantly different from one another.

Instructional practices. The Reading PD study measured the extent to which teachers in the study used three teaching practices that had been emphasized in the professional development that the teachers in the two treatment groups received: teacher-led explicit instruction, independent student activity, and instruction that was differentiated to meet individual students' needs. During the spring of the implementation year, teachers in both treatment groups were more likely to use explicit instruction than teachers in the control group. There were no statistically significant differences in the extent to which teachers across the three groups employed the other two recommended practices. In the fall of the follow-up year, when teachers were observed again, there were no significant differences among the groups on any of the instructional practices measured.

Student achievement. Student achievement in reading was measured in two ways: (1) average scores on the standardized test used to assess reading achievement in each of the study districts and (2) the percentage of students scoring at or above the average score for their district's last cohort before the professional development intervention began (school year 2004-2005). The two professional development interventions did not register impacts on either of these achievement measures in either the implementation or the follow-up year. The impacts of the two different treatments are statistically indistinguishable in magnitude.

The Math PD Study

Receipt of professional development in math. On average, in both demonstration years, teachers in the treatment group schools received significantly more professional development in math than did their control group counterparts. During the first year, the difference was driven by the fact that teachers in the treatment group spent significantly more hours in

institutes and seminars than their counterparts in the control group; treatment group teachers also received more coaching than control group teachers, but this difference is not statistically significant. During the second year, treatment group teachers got significantly more workshops and seminars *and* more coaching than did teachers in the control group.

Teacher knowledge. In both years of the Math PD study, teachers in the treatment group schools did not exhibit significantly higher levels of overall mathematical knowledge than their control group counterparts. Nor did they score higher on the subscale measuring common knowledge of mathematics. There is one statistically significant subgroup difference that favored the treatment group: At the end of the first year, treatment group teachers in the districts that received only one year of professional development scored higher on the measure of specialized knowledge of mathematics for teaching than did control group teachers. At the end of the second year, 76 percent of teachers in the treatment group answered test items of average difficulty correctly, compared with 75 percent of teachers in the control group.

The researchers conducted additional exploratory analyses to take advantage of the added power provided by a "pooled" sample of teachers who were in the first-year analysis only (from all 12 study districts), who were in the second-year analysis only (from the 6 two-year districts), or who were in both analyses. The analysis using this sample indicated that one year of professional development did not produce a statistically significant impact on the overall score or on the subscale of common knowledge; however, teachers in the treatment group scored significantly higher on the subscale of specialized knowledge of mathematics for teaching.

Instructional practices. The impacts of the professional development on math teachers' instructional practices were measured through classroom observations during the first implementation year only. Statistically significant differences were found for one of the three measures of practice examined: On average, on an hourly basis, treatment group teachers engaged in 3.5 activities that elicited students' thinking (for example, asking students whether they agreed or disagreed with a student's response, asking them to provide additional strategies for solving a problem), compared with 2.4 such activities per hour for control group teachers. The professional development had a positive impact that just missed statistical significance on the treatment group teachers' use of visual representations (for example, number lines) and had no effect on the frequency with which the teachers engaged in activities focused on mathematical reasoning (asking, for example, why an answer did or did not make sense).

Student achievement. In neither year of the demonstration did the Math PD study have a statistically significant impact on students' knowledge of rational numbers. On a test specifically devised to measure this knowledge, students in treatment group schools did not have higher average overall scores than students in control group schools, nor did they have higher

scores on either of two subscales measuring their knowledge of fractions and decimals and of ratios and proportions.

Explaining the Findings

What accounts for these disappointing findings? This section explores a number of possible explanations for the results:

- Teachers' backgrounds and attitudes
- The content, quantity, and quality of the professional development that was delivered
- Methodological issues associated with the evaluation
- Teacher and student turnover
- The underlying theory of change

These factors are theoretically separable, but they are potentially inextricably intertwined in practice. For example, problematic measures potentially make it hard to establish relationships among the variables in the theory of change. High teacher turnover may mean that too few teachers got enough of the professional development for the theory to receive a fair test.

Table 4 summarizes the analysis of the findings, posing specific questions within each category of explanations and presenting answers to these questions, where these are known. (A question mark indicates that systematic data to answer the question are unavailable.) It is important to state at the outset that large parts of this discussion are speculative. It is possible to determine that some potential explanations are *wrong*, but it is not equally possible to be confident that other potential explanations are *right.* And the evidence needed to answer some questions is ambiguous or unavailable, as the text makes clear. While the analysis cannot provide definitive answers on a number of points, it can supply the grist for further conversation about the studies and their findings.

Teachers' Backgrounds and Attitudes

Treatment effects could potentially be influenced by the characteristics of teachers and students in the study sample. It is worth asking, for example, whether the professional development was aimed at teachers who could benefit from it. Did teachers have the basic experience required to benefit from professional development that was focused on content and pedagogy, rather than on basic procedures like maintaining classroom discipline? The answer is almost certainly yes. Overall, 85 percent of the teachers in the Reading PD study had been teaching for at least four years at baseline. And at the start of the Math PD study, about 70 percent of the

The Professional Development Synthesis Report

Table 4

Possible Explanations for the Findings

Area of Inquiry and Questions Suggesting Explanations	Reading Professional Development	Math Professional Development
Teachers' backgrounds and attitudes		
Were teachers' levels of experience at baseline high enough that teachers could benefit from content- and pedagogy-focused professional development?	Yes	Yes
Were teachers' levels of knowledge low enough at baseline that teachers could benefit from the professional development that was delivered?	Yes	Yes
Did teachers like the professional development and think that it was providing them with new information?	Probably yes	Probably yes
Content, quantity, and quality of professional development		
Should the professional development have placed greater emphasis on some topics?	?	?
Did the professional development cover the intended topics?	Yes	Yes
Was the intended quantity of professional development delivered?	Yes	Yes
Was the information that was conveyed accurate?	Probably yes	Probably yes
Methodological issues		
The measures used		
Do the measures capture what the professional development emphasized?		
Teacher knowledge	Yes	Yes
Instructional practices	Yes	Yes
Student achievement	Only in part	Yes
Are the measures reliable?		
Teacher knowledge	Yes	Yes
Instructional practices	Uncertain	No
Student achievement	Yes	Yes
Are there important unmeasured constructs?	Yes	Yes

(continued)

Table 4 (continued)

Area of Inquiry and Questions Suggesting Explanations	Reading Professional Development	Math Professional Development
Sample size		
Is the overall sample size adequate?	Yes, but…	Yes, but…
Did the random assignment produce treatment and control groups that were fully equivalent at baseline?	Yes	?
Teacher and student turnover		
Did teacher turnover mean that many teachers for whom impacts were measured did not receive a full dosage of the treatment?	Yes	Yes
Were impacts attenuated for this reason?	No	Yes
Did student turnover weaken impacts?	No	?
Theory of change		
Are measures of teacher knowledge related to measures of student achievement?	Yes	Yes
Are measures of instructional practices related to measures of student achievement?	Yes, to some extent	?

NOTE: A question mark indicates that systematic data to answer the question are unavailable.

teachers across the treatment and control groups had four or more years of teaching experience (much of that time was spent teaching middle school mathematics).

If teachers already knew the material that they were being taught or if they were already using at high levels the teaching strategies that the professional development advocated, there would be little room for improvement. The data indicate that this was definitely not the case. In the Reading PD study, on average, teachers had about a 53 percent chance of answering a typical item on the baseline knowledge test correctly (compared with 81 percent for a group of experienced professional development providers who also took the test). At the beginning of the Math PD study, 46 percent of teachers in the treatment group and 51 percent of teachers in the control group answered teacher-knowledge test items of average difficulty correctly (compared with 93 percent of the first-year professional development providers).[12] In fact, some have argued that, as a group, the math teachers knew too little — rather than too much — to benefit from the professional development. Analyses show, however, that the effects of the professional development did not differ for teachers with different levels of background knowledge.

[12]It is worth noting that less than 30 percent of the middle school math teachers had majored in math or a related subject in college. As previously noted, the scale that was used to measure teacher knowledge has two subscales, one measuring general knowledge of rational numbers and the other measuring specialized knowledge useful for teaching. On the first subscale, teachers in the treatment group scored a good deal lower than teachers in the control group; the difference is significant at the 10 percent level but not at the 5 percent level adopted in the report.

Teachers who valued the professional development that they received and who felt that they were learning important new things might be expected to strive to put its precepts into practice. The same cannot be said of teachers who responded to the professional development with indifference. Unfortunately, systematic information was not collected through survey questions or other methods about teachers' attitudes toward the professional development, although there is reason to think that teachers in both studies enjoyed it and felt that they were benefiting from it.[13]

It is important to note, though, that teachers were not told their scores on the teacher-knowledge tests used in the evaluations. If they had been, teachers at the treatment group schools might have appreciated more fully the need to improve their knowledge and skills.

The Content, Quantity, and Quality of the Professional Development That Was Delivered

A second category of potential explanations concerns the professional development that was delivered. Another experiment would be needed to determine whether professional development that included other topics or gave additional emphasis to some of the topics that were already included would produce larger impacts. In retrospect, however, some observers speculated that the professional development would have been more beneficial if it had been more closely linked to the curricula that teachers were using and to teachers' classroom activities and if had included more sessions during which teachers developed lesson plans for specific topics. They also suggested that the math professional development might have been more effective if teachers had been required to work more problems in the course of the training.

The evaluation plan called for researchers to address the *fidelity* of the professional development that was delivered to what was planned: whether it covered the topics that it was supposed to cover and whether the right amounts of it were delivered. As Table 4 shows, observations of the training institutes and seminars indicate that the professional development was implemented faithfully with respect to coverage and allotted time; the same was true of coaching. If the impact findings are disappointing, it is not because the delivery of the profes-

[13]For example, researchers who observed the Reading PD sessions and the conversation during training breaks noted that teachers were often heard to say that what they were learning was new and useful and that they wished they had known this material when they began to teach.

The decision not to collect attitudinal data from the teachers was driven both by resource constraints and by a preference for objective measures over ones that tapped teachers' subjective experiences. The protocol that researchers used when they observed the professional development asked the researchers to rate teacher engagement during the training sessions. While the results appear to indicate that teachers were engaged, the measure of engagement does not set a high bar: Teachers were counted as actively engaged if they were observed to be "working problems" or "contributing to the discussion" but also if they were judged to be "watching the facilitator" or "listening."

sional development fell short of what was intended in terms of coverage and quantity. What about the quality of the professional development? Did the trainers provide information that was accurate? Because a solid mathematics background was not a criterion for selecting the researchers who observed the training, the observations could not address this question. However, all institute and seminar materials were reviewed by content experts for quality and accuracy, and, as previously noted, training facilitators scored high on measures of their content knowledge. So the evidence suggests that fidelity of implementation to what was planned was tantamount to high-quality implementation. There is more reason to be concerned about the quality of the coaching. In the Reading PD study, coaches scored considerably higher than did teachers on the test of content knowledge, but they also scored considerably lower than did the professional development institute facilitators. In the Math PD study, the coaches were the institute facilitators, but while they had a great deal of content-matter expertise as well as experience in leading professional development workshops, they may have had less expertise in coaching.

Methodological Issues Associated with the Evaluation

Two aspects of each study's research design might help to explain the impact findings: the *measures* used and the *sample size* (that is, the number of schools in the study).

Measures employed. Table 4 raises a number of questions about the measures of teacher knowledge, teacher practice, and student achievement that constituted the key outcomes in each study.

First, were the outcome measures related to the contents of the professional development? If the professional development stressed some concepts and behaviors and the measures were of quite different concepts and behaviors, there would be no reason to expect that exposure to the professional development would affect the outcomes being measured. Fortunately, that does not appear to be the case in these studies, with one exception: Students' reading skills were measured using the standardized tests normally employed in the study districts, since these were believed to be of the greatest importance to policymakers. These tests tended to emphasize students' passage-comprehension skills rather than the word-level knowledge stressed in the professional development that their teachers received. The other outcome measures did not exhibit this alignment problem: In both studies, the measures of teacher knowledge and practice were integrally tied to what was covered in the professional development, and this was true as well for the measure of students' understanding and manipulation of rational numbers, which was especially developed for the math study.

Second, were the measures reliable? That is, were they an accurate reflection of the teachers' and students' true performance? Here, as Table 4 shows, the measures of teachers' instructional practice are the major cause for concern. These were gleaned through classroom

observations — a form of data collection that is resource-intensive and expensive. Given the priority placed on having enough schools in the research sample to yield robust conclusions about student impacts, the budget could accommodate only a very limited number of observations per teacher: three for the Reading PD study (two during the implementation year and one during the follow-up year) and only one for the Math PD study (during the first year only, with no observations conducted during the second year for the two-year sites). With such a small number of data points, it seems likely that, especially in the Math PD study, the observations captured practices that were not representative of teachers' typical classroom behavior across the school year.

Finally, did important constructs go unmeasured? At the end of most investigations, the researchers can point to data that they wish they had collected. As noted above, the evaluations would have been richer had they collected systematic data on teachers' responses to the professional development. Furthermore, the observational data that were collected on teachers' practices center on how teachers conveyed content but provide no information on the accuracy of what was taught, since observers were not well enough grounded in the subject matter that they were observing to know when mistakes were being made. Misinformation that was conveyed — even if using the teaching strategies recommended in the professional development — would have detracted from student achievement.

Sample size. The smaller the impact that evaluators want to establish as being statistically significant, the larger the research sample must be to detect it. The number of schools in the Reading PD and Math PD studies reflected input from the U.S. Department of Education about the magnitude of effects that department officials believed to be policy-relevant. Department officials were not particularly interested in finding small but statistically significant effects on adults; they reasoned that the impacts on teachers' knowledge and instructional practices would need to be substantial in order to affect student achievement to a policy-relevant degree. Consequently, the studies were designed to detect relatively large effects and to involve relatively small numbers of teachers.[14] As it turned out, however, the interventions' actual effects on teacher knowledge and instruction are, in many cases, smaller than the effects that the studies had been designed to detect as statistically significant. Had the Math PD study involved a larger sample, a couple of the impacts on teachers might be deemed statistically significant.[15] But the professional development would still not have made a difference for student outcomes.

[14]"Minimum detectable effect size" is the smallest true effect that a study has a good chance of detecting. The Reading PD study was designed to detect a minimum detectable effect size of 0.40 for teacher outcomes and 0.20 for student outcomes. The second year of the Math PD study was designed to detect a minimum detectable effect size of 0.59 for teacher knowledge and 0.20 for student achievement.

[15]After the first year of implementation, the professional development had a positive impact on teachers' use of representations that just missed being statistically significant at the 5 percent level. (The p-value is (continued)

An additional issue about sample size concerns its relationship to the comparability of the treatment and control groups. It is an axiom of evaluation research that random assignment will produce fully equivalent treatment and control groups (or, more accurately, that any treatment-control differences will themselves be randomly distributed), and this is true — provided that the sample is sufficiently large. Otherwise, nonrandom differences may enter into play. In the Math PD study, teachers in the control group schools scored higher than their treatment group counterparts at the beginning of the study. While this difference is not statistically significant, it is possible that the difference was real, and that statistical adjustments may not have fully corrected for it. If so, the professional development might have had to fight an uphill battle in order to produce impacts on the knowledge and behavior of teachers in the treatment group schools and on the achievement of their students.

Teacher and Student Turnover

Teacher turnover is a common occurrence in low-performing schools like those participating in these studies. Teacher mobility could potentially affect both teacher and student outcomes. Teachers departing the treatment schools during the study were replaced by teachers who did not receive the full amount of professional development that was intended. Since impacts were measured for all teachers in the schools, regardless of the amount of professional development that they had received, turnover could reduce impacts on teacher knowledge and instructional practice. And since impacts were measured for all students, some of whom were taught by teachers whose exposure to the professional development fell short of what was planned, turnover could reduce student achievement as well.

Teacher mobility may have affected impacts in one of the two studies. Teacher turnover does not appear to be the explanation for the lack of impacts on either teacher knowledge or instructional practices during the follow-up year of the Reading PD study, despite the fact that one-third of the teachers in the treatment schools left their schools between the start of the implementation year and the end of the follow-up year. An exploratory analysis examined these outcomes for a stable group of teachers who remained in the study schools throughout both years, and no treatment-control impacts on these outcomes were found. (Because this analysis is based on a nonrandom subset of all treatment and control group teachers, the findings are necessarily less definitive than those based on the full sample.) In the Math PD study, teacher

---

0.054.) After the second year, there was a positive impact on the measure of specialized knowledge of math for teaching that is statistically significant at the 10 percent level but not at the 5 percent level.

Nonexperimental analyses examined the effects of the professional development on teacher knowledge, combining the effect of the first year of the intervention for teachers present at the end of the first year with the additional effect of the second year of the intervention for teachers in the schools in both years or in the second year only, thereby increasing the sample size. Again, with this larger sample, there is a statistically significant effect on specialized knowledge of math for teaching.

turnover may have made more of a difference. Nearly half of the teachers present in treatment schools in the two-year districts (22 of the 45 teachers) did not receive the two years of professional development that the evaluation was intended to test.

Mobility obviously can affect students as well as teachers, and student mobility is a commonplace phenomenon in urban districts like those participating in the evaluation. If students moved into treatment schools at some point after the beginning of the school year, they would not have been exposed for the full year to teachers whose instruction was informed by the professional development that the teachers had received. This could attenuate the ability of the professional development to affect student performance.

In the Reading PD study, in treatment and control group schools, analyses that compared outcomes for "stable students of stable teachers" were conducted to address the issue of student mobility and its effects. addressed this possibility. (The stable-student analysis excluded students who were enrolled in the study school six weeks or less of the implementation year.) No impacts on student achievement were found, suggesting that student mobility does not explain the absence of impacts. Analyses conducted as part of the Math PD study yielded similar conclusions.

The Underlying Theory of Change

Yet another possible reason for the generally weak impacts is that the theory of change is inadequately specified with respect to its intermediate steps and that, in reality, teacher knowledge and instructional practices do not affect student outcomes. One way to test the theory is to see whether teacher knowledge and practice predict student test scores once other variables (such as students' background characteristics and prior achievement) have been controlled for. Positive associations would have to be viewed as suggestive rather than causal; nonetheless, such associations, if found, would alleviate concerns about the validity of the theory.[16]

Table 4 presents the results. In the Reading PD study, correlational analyses indicate that greater teacher knowledge is, in fact, associated with higher student test scores but that sizable gains in teacher knowledge make for much smaller test score gains — a point that is reprised below. The evidence with respect to the relationship between instructional practices and student test scores is more mixed, but it suggests that differentiated instruction, in particular, is associated with higher student achievement.

[16]Failure to find such associations does not necessarily mean that the theory is wrong, however. Poor measures of the constructs could result in weak associations even if the theory is correct.

The first-year results of the Math PD study show no significant relationships between teacher knowledge or practice and student achievement. In the second year, the teacher knowledge findings resemble those for the Reading PD study: Higher levels of teacher knowledge are associated with higher student test scores, but, again, large increments in teacher knowledge are associated with much smaller student gains. (The math study did not examine the association between instructional practices and student test scores, since the former were not measured during the second year of the study.)

All in all, then, the evidence suggests that increasing teacher knowledge can increase student test scores, although not by as much as policymakers might hope. The relationship between changed practice and test scores is less certain, but this may be partly a result of inadequate measures of that practice, as described above.

* * *

In both demonstrations, a number of factors reduced the effectiveness of the professional development and the researchers' ability to evaluate that effectiveness. In both studies, the professional development was delivered as planned and was delivered to teachers who could have benefited from it. But in the Math PD study, the baseline difference in teacher knowledge between the treatment and control groups, which would probably have been smaller had the sample been larger, may have decreased the likelihood of detecting statistically significant effects. (This was particularly the case when it came to testing the effects of two years of professional development, since only half the original number of schools were involved.) In the Math PD study, too, teachers may not have understood the gaps in their knowledge of rational numbers and may not, therefore, have taken full advantage of the learning opportunities that the professional development offered. Finally, teacher turnover in that study meant that nearly half the teachers in the two-year sites did not receive two years' worth of professional development. In the Reading PD study, the test of student reading ability focused on comprehension, whereas the professional development had centered on word-level skills.

These problems with the studies notwithstanding, the evidence suggests that the theory of change makes sense: professional development can increase teacher knowledge and that teacher knowledge and improved instruction can change student test scores. But changes in teacher knowledge and practice would have had to be considerably greater than they were to move the needle on student achievement significantly. This finding is not unique to the two evaluations discussed in this report. Correlational evidence from other studies has also established that fairly sizable changes in teacher-related variables are associated with much smaller changes in student learning outcomes.[17] In other words, teachers who are considerably above

[17]See, for example, Hill, Rowan, and Ball (2005) and Rockoff, Jacob, Kane, and Staiger (2008).

average in knowledge tend to have students whose gains in reading or math are only somewhat above average.

Reflections and Future Directions

The foregoing raises two questions: How might professional development produce larger, more lasting impacts for teachers, and larger impacts for students as well, than were found in the Reading PD and Math PD studies? And how might evaluations yield more and more useful implementation findings while remaining affordable?

Improving Professional Development

First, the professional development that was delivered in these studies might have made more of a difference if teachers had been aware that they needed it. While reading teachers were well aware that the professional development that they received was covering new and previously unexplored territory, math teachers did not seem to realize that their knowledge of rational numbers was often shaky, and their scores on the pretest were not shared with them. Planners may have wished to maintain the confidentiality of teachers' responses — and to avoid embarrassing them. But the consequence may have been that the math teachers, not knowing how much they did not know, were less motivated to take the professional development seriously.

There are ways around such a dilemma. For example, the overall results of the pretest — the average score on the test as a whole, for example, or the percentage of teachers answering particular items correctly — could be made known. Individual scores could then be shared privately with teachers who request this information or with teachers who assert that they do not need the professional development.

What about the professional development itself? In these studies, the professional development that was tested represents a marked difference from — and improvement over — the "one-shot" professional development that has been widely criticized. Through periodic institutes and seminars during the school year and especially through coaching, it sought to refresh and reinforce what teachers had learned in the intensive summer training. It also focused on both content knowledge and pedagogical skills, rather than privileging one at the expense of the other.

At the same time, the professional development that was evaluated had limitations. It may not have been sufficiently tied to the curricula that teachers used in their classrooms or may not have placed adequate emphasis on how desirable instructional practices could be incorporated into lesson plans. Perhaps most fundamentally, it was not part of a schoolwide instructional improvement effort — it was narrowly targeted toward teachers in particular grades, rather than an element of a broader plan to change how reading and mathematics were taught in the

target schools. The engagement of a school's principal, many experts believe, is critical to the sustained success of initiatives to change instruction; although principals were invited to the professional development, their presence was not expected or required. In the interventions that involved coaching, the coaches held grade-level meetings at which participating teachers were encouraged to examine student work and discuss instruction, but there is no evidence that these meetings occurred when the coaches were not there to lead them. In short, the professional development did not have built-in mechanisms to transform teachers' practices in a long-term way.

New thinking emphasizes not just formal "professional development" activities but a broader conception of "professional learning" that includes not only externally provided activities but also ones that arise in the school context, where teachers are part of a community of learners.[18] Structures and initiatives that can promote professional learning within the community include opportunities for teachers to observe colleagues within their own schools and in other schools, occasions for teachers to examine student work and test scores together, and common planning time for teachers to talk with one another about what instructional practices they have tried, what worked and what did not, and how the practices might have worked better. In this conception, the principal has a critical role in helping to define the school as an institution committed to ongoing learning and improvement, in setting the priorities (such as providing for regular meeting times for teachers) that foster change and in monitoring both the content of teacher meetings (to make sure that teachers remain focused on instruction) and student outcomes. In short, professional development is transformed from something that a specific group of teachers does for a limited period of time in isolation from their peers to something that is an integral part of what all teachers do all the time and that involves a continuous and collegial cycle of learning, practice, reflection, and improvement.

There have been no randomized trials that have tested whether professional development that is reinforced within a professional learning community has value over and above professional development that is more narrowly targeted. But it is worth testing the hypothesis that teachers will change their practice more when they are part of a peer group whose members are expected and encouraged to try out new, research-driven instructional methods, to discuss their experiences, and then to try again.

It may also be worth considering how content-focused professional learning can be diffused throughout an entire district. Given high rates of student and staff mobility, especially in urban districts, efforts to ensure that teachers throughout the district are adopting the same

[18]This discussion owes much to the work of Linda Darling-Hammond and her colleagues at the School Redesign Network at Stanford University. See Darling-Hammond et al. (2008) and Wei et al. (2009).

approach to student learning — and to their own instructional practices — could mitigate the disruptions that occur when students change schools.

Notwithstanding, it may be unrealistic to expect to see a major transformation of either teacher practices or student achievement within the relatively short time period of a year or two. It seems likely that unless teachers are mandated to adopt certain instructional approaches and their adherence to these techniques is closely monitored, they will stick with practices that are familiar to them, incorporating new ones gradually as they learn how they have worked for others and see how they work for themselves. Given the finding that very substantial changes in practice are necessary to bring about student improvement, a fundamental design lesson of the two professional development interventions that were examined may be that more than one year of professional development is needed to produce large and lasting change.

Improving Evaluations

Another direction for change that emerges from these two studies concerns evaluation methodology rather than the substance of what was evaluated. The implementation analysis in the studies largely centered on assessing the fidelity of the professional development that was offered to what was initially planned — whether the same topics were covered and for how long — since, without such fidelity, the professional development could not have been said to receive a fair test.

Other interesting but unanswered implementation questions, however, turn out not to have concerned fidelity all. Instead, as suggested above, it would have been useful to understand in a more systematic way how teachers were responding to the professional development, including whether it prompted them to think about how it might be integrated into their classes. It would have been even more helpful to know more about the quality of classroom instruction, through classroom observations that focused not just on the number of times that teachers demonstrated the instructional practices that had been emphasized in the professional development but also on whether that instruction was engaging to students — and whether it was accurate. Whether or not teachers' practices changed in the desired directions, if teachers were passing along the wrong information, it seems likely that the achievement of their students would have suffered.

Conducting classroom observations is expensive to begin with; hiring observers who can assess the accuracy of what is being taught would make these observations even more expensive — but also potentially much more useful. It is worth asking whether there are ways to reduce the cost of learning what teachers do in the classroom. Possibilities include:

- Using outstationed researchers with flexible schedules, who may be better able to maximize classroom visits than researchers who are deployed from a central office
- Asking teachers to maintain logs of their activities in the classroom over a given time period (and compensating them for their efforts)
- Analyzing assignments that teachers give to their students, along with student work done in response to those assignments
- Making use of implementation data maintained by program sponsors about teachers' practices

In addition, it would be worth exploring the feasibility of public-private funding partnerships in which public sources could support evaluations of program impacts while private sources could address questions of why and how programs succeed or fail.

* * *

Perhaps the chief implication of the studies, and one that applies to the design of both professional development interventions and evaluations of them, is that more investment — of time and personnel, not just money — may lead to bigger payoffs. At the same time, in-service training for teachers who are already in place should not be the only vehicle for increasing student achievement. Strengthening undergraduate or graduate school course requirements for prospective teachers might help to ensure that the teachers have adequate mastery of the subjects that they will be teaching. Programs like Math for America, which provides fellowships to mathematically able students to teach in secondary schools while earning a master's degree, should also be encouraged. Altering certification requirements might also make for a better-qualified teaching force. But even the best teachers will have only a limited ability to improve learning if students arrive in the classroom sleepless, hungry, or unable to see the blackboard. Better professional development is just one of the many tools needed to increase the educational achievement of America's most disadvantaged young people.

References

Darling-Hammond, Linda, Ruth Chung Wei, Alethea Andree, Nikole Richardson, and Stelios Orphanos (School Redesign Network at Stanford University). 2008. *Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad.* Dallas: National Staff Development Council.

Garet, Michael S., Stephanie Cronen, Marian Eaton, Anja Kurki, Meredith Ludwig, Wehmah Jones, Kazuaki Uekawa, Audrey Falk (American Institutes for Research), Howard S. Bloom, Fred Doolittle, Pei Zhu, and Laura Sztejnberg. 2008. *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement.* NCEE2008-4030. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Garet, Michael S., Andrew J. Wayne, Fran Stancavage, James Taylor, Marian Eaton, Kirk Walters, Mengli Song, Seth Brown, Steven Hurlburt, Pei Zhu, Susan Sepanik, and Fred Doolittle. 2011. *Middle School Mathematics Professional Development Impact Study: Findings After the Second Year of Implementation.* NCEE 2011-4024. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Garet, Michael S., Andrew J. Wayne, Fran Stancavage, James Taylor, Kirk Walters, Mengli Song, Seth Brown, Steven Hurlburt, Pei Zhu, Susan Sepanik, and Fred Doolittle. 2010. *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.

Hill, Heather C., Brian Rowan, and Deborah Loewenberg Ball. 2005. "Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement." *American Educational Research Journal* 42, 2: 371-406 (Summer).

Moats, Luisa C. 2005. *Language Essentials for Teachers of Reading and Spelling (LETRS®).* Longmont, CO: Sopris West Educational Services.

National Institute of Child Health and Human Development. 2000. *Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction: Reports of the Subgroups.* NIH Publication No. 00-4754. Washington, DC: U.S. Government Printing Office.

Rockoff, Jonah E., Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger. 2008. "Can You Recognize an Effective Teacher When You Recruit One?" Working Paper No. 14485. Cambridge, MA: National Bureau of Economic Research (NBER).

Wei, Ruth Chung, Linda Darling-Hammond, Alethea Andree, Nikole Richardson, and Stelios Orphanos (School Redesign Network at Stanford University). 2009. *Professional Learning*

in the Learning Profession: A Status Report on Teacher Development in the U.S. and Abroad. Technical Report. Dallas: National Staff Development Council.

Yoon, Kwang Suk, Teresa Duncan, Silvia Wen-Yu Lee, Beth Scarloss, and Kathy L. Shapley. 2007. *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement.* Issues & Answers Report, REL 2007-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. Web site: http://ies.ed.gov/ncee/edlabs.

About MDRC

MDRC is a nonprofit, nonpartisan social and education policy research organization dedicated to learning what works to improve the well-being of low-income people. Through its research and the active communication of its findings, MDRC seeks to enhance the effectiveness of social and education policies and programs.

Founded in 1974 and located in New York City and Oakland, California, MDRC is best known for mounting rigorous, large-scale, real-world tests of new and existing policies and programs. Its projects are a mix of demonstrations (field tests of promising new program approaches) and evaluations of ongoing government and community initiatives. MDRC's staff bring an unusual combination of research and organizational experience to their work, providing expertise on the latest in qualitative and quantitative methods and on program design, development, implementation, and management. MDRC seeks to learn not just whether a program is effective but also how and why the program's effects occur. In addition, it tries to place each project's findings in the broader context of related research — in order to build knowledge about what works across the social and education policy fields. MDRC's findings, lessons, and best practices are proactively shared with a broad audience in the policy and practitioner community as well as with the general public and the media.

Over the years, MDRC has brought its unique approach to an ever-growing range of policy areas and target populations. Once known primarily for evaluations of state welfare-to-work programs, today MDRC is also studying public school reforms, employment programs for ex-offenders and people with disabilities, and programs to help low-income students succeed in college. MDRC's projects are organized into five areas:

- Promoting Family Well-Being and Children's Development
- Improving Public Education
- Raising Academic Achievement and Persistence in College
- Supporting Low-Wage Workers and Communities
- Overcoming Barriers to Employment

Working in almost every state, all of the nation's largest cities, and Canada and the United Kingdom, MDRC conducts its projects in partnership with national, state, and local governments, public school systems, community organizations, and numerous private philanthropies.