

**Group Work Is Not Cooperative Learning:
An Evaluation of PowerTeaching
in Middle Schools**

A Report from the Investing in Innovation (i3) Evaluation

**Shelley Rappaport
Jean Grossman
Ivonne Garcia
Pei Zhu
Osvaldo Avila
Kelly Granito**

with

**Deni Chen
Ashley Kennedy
Joseph Quinn**

June 2017



Funding for this report came from the U.S. Department of Education under its Investing in Innovation (i3) initiative. The i3 grant called for the Success for All Foundation to scale up its middle school PowerTeaching program and for MDRC to conduct an independent evaluation of the implementation and impacts of that expansion.

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 Annie E. Casey Foundation, Charles and Lynn Schusterman Family Foundation, The Edna McConnell Clark Foundation, Ford Foundation, The George Gund Foundation, Daniel and Corinne Goldman, The Harry and Jeanette Weinberg Foundation, Inc., The JPB Foundation, The Joyce Foundation, The Kresge Foundation, Laura and John Arnold 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 website: www.mdrc.org.

Copyright © 2017 by MDRC®. All rights reserved.

Overview

In the next decade, the fastest growing occupations are projected to be in the fields of science, technology, engineering, and mathematics, and will require advanced mathematical and scientific knowledge. Unfortunately, many American students today, especially those in low-income schools, are performing at low levels in math and will have trouble gaining access to these jobs. It is therefore critical that middle school students succeed in math. The PowerTeaching program is a middle school math program that has shown strong evidence of effectiveness. Developed by the Success for All Foundation, it emphasizes cooperative learning to instruct math. In 2011, Old Dominion University received a grant through the U.S. Department of Education's Investing in Innovation program to scale up the PowerTeaching program. In 2012, MDRC began a multiyear evaluation of the scale-up effort, conducting an implementation study and an impact study that included a school-level randomized controlled trial. Over two years, the research team randomly assigned 58 schools, of which 30 (those assigned to the program groups) were part of the scale-up effort. The remaining 28 schools were assigned to the control group and as such were not part of the scale-up group of schools. This report describes the evaluation and presents its findings, key among which are the following:

- Although the Success for All Foundation and the schools in the study provided the requisite time, staff, and materials needed to support teachers in their implementation of the PowerTeaching program, teachers in only a few schools collected and used student assessment data to drive instruction, and most teachers did not receive the kind of training and support needed to create cooperative learning teams in their classrooms.
- Students in both program and control group schools worked in groups often, but students in program group schools spent more time in groups than students in control group schools. Students in program group schools were also more likely to be in longstanding mixed-ability groups. Despite these differences found in group work, many teachers in program group schools did not use the techniques that move group work to true cooperative learning.
- Students in both the program group and control group schools performed equally well on math, as measured by their state math test scores. However, students in schools that enrolled in the study earlier did worse than those students in schools that enrolled later.
- While, overall, implementation of the program was weak in all the schools participating in the evaluation, the non-research scale-up schools — which tended to be smaller and in less urban environments — implemented the PowerTeaching program slightly better than the schools in the study.

Contents

Overview	iii
List of Exhibits	vii
Preface	ix
Acknowledgments	xi
Introduction	1
PowerTeaching	1
The Evaluation	4
The Findings	9
Conclusion	22
References	25

List of Exhibits

Table

1	Background Characteristics for Schools in the Study Sample, Schools in the PowerTeaching Scale-Up Sample, and Similar Schools in the National Population (2012-2013 Academic Year)	6
2	School Achievement Snapshot Scores for Items Related to Schoolwide Structures and Instructional Practices, Study Schools (2014-2015 and 2015-2016 Academic Years)	12
3	Impact of PowerTeaching on Minutes Students Spent Doing Group Activities During the Math Block (Activities Are Not Mutually Exclusive)	15
4	Impact of PowerTeaching on Students' Average Math Achievement by Grade, for Analysis Samples	18
5	Impact of PowerTeaching on Students' Average Math Achievement for Grade 7 Analysis Sample, by Cohort and Student Subgroup	20

Figure

1	Logic Model for the Success for All Math PowerTeaching Program in Middle Schools	3
2	Timeline and Sample Configuration of the Evaluation	8
3	School Achievement Snapshot Scores, by Category and Academic Year, Study Schools Only	10
4	2015-2016 School Achievement Snapshot Scores, by Category, Scale-Up Schools Only	23

Box

1	Cooperative Learning in Action: Becoming a Super Team	5
---	---	---

Preface

To succeed in today's economy, students need both proficiency in the "three Rs" and strong applied skills. Communication skills, team work, and critical thinking have long been at the top of employers' lists of applied skills they seek in employees. States are responding to employers' needs by putting in place new educational standards. These standards include not only higher levels of basic academic knowledge that students are expected to master but also applied skills pertaining to presenting information, explaining one's reasoning, and effectively collaborating in groups. As a result, teachers nationwide are having students work in groups more frequently. This report examines a recent large-scale effort to expand a cooperative learning program in middle schools.

The change in standard instructional practices gives schools a chance to not only teach students applied skills, but improve students' academic learning, if they can help teachers turn "group work" into "cooperative learning teams." PowerTeaching, a structured cooperative learning program, was designed to do just that. Thus, the expansion of PowerTeaching through a federal Investing in Innovation grant offers the education field a unique opportunity to learn what it takes to help teachers create cooperative learning environments in their classrooms.

This report presents the lessons learned from this scale-up effort and findings from a multiyear evaluation of it. It describes how PowerTeaching was implemented over the first few years, how classrooms with the program differed from those without it, and whether students in the program performed better in math. The evaluation found that while teachers who taught with PowerTeaching learned to place their students into longstanding mixed-ability groups, which are thought to be conducive to cooperative learning, teachers did not consistently use the program's instructional techniques that transform group work into cooperative learning. In turn, students' math performance did not differ significantly between schools using the program and schools not using it. A likely cause for the weak implementation was that the ongoing professional development, which is an integral part of the PowerTeaching program, mostly did not occur or focused more on teaching the new material required by recently adopted education standards rather than on cooperative learning techniques. The evaluation thus points to the importance of focused, ongoing training and support when trying to modify teachers' instructional practices.

Gordon L. Berlin
President, MDRC

Acknowledgments

The evaluation of the Investing in Innovation (i3) scale-up initiative of Success for All's middle school PowerTeaching program, and this resulting report, reflects the efforts of many people. Our first debt of gratitude is to the principals, Success For All Foundation (SFAF) coaches, research liaisons, PowerTeaching facilitators, and teachers at the schools that participated in the study; to staff members in the central offices of the five school districts who provided us with critical student records data as well as staffing information; and to district coaches who took part in the interviews. The assistance and cooperation of these individuals were vital for enabling the study to go forward and for providing the rich and detailed information on which this report is based.

John Nunnery and Pamela Arnold-Puchalski at Old Dominion University and Nancy Madden and Paul Miller at SFAF responded promptly and patiently to our many information requests. In particular, Paul Miller spent countless hours giving us a strong understanding of the PowerTeaching program. John Nunnery, Pamela Arnold-Puchalski, Nancy Madden, Robert Slavin, and Paul Miller provided useful critiques at each stage of the report drafting process.

Cristofer Price at Abt Associates provided thorough, thoughtful, and welcomed technical assistance related to the study design and its implementation throughout the project. He went above and beyond in helping us navigate the i3 process.

Thomas J. Smith, Melissa Comerchero, Zach Pinto, Amanda Ferrandino, and Cammie Brown played an active role in the field research and data collection. Melissa Comerchero, Zach Pinto, and Amanda Ferrandino also assisted in coding the data. Helping the research team with the principal and teacher surveys as well as the teacher logs were: Zach Pinto, Seth Muzzy, Bulent Can, Sandhya Bandhi, Usha Krishnan, and Matthew Au. Jillian Verillo helped with the preparation of this report.

Fred Doolittle ensured that the team received both material and moral support at every turn. He, along with Janet Quint, Marie-Andree Somers, Robert Ivry, and Jennie Kaufman carefully reviewed earlier drafts of this report and made comments that improved the final product. Mario Flecha handled a variety of administrative and other tasks. Kelly Granito and Alpesh Shah provided fiscal oversight. Christopher Boland edited the report, and Carolyn Thomas prepared it for publication.

The Authors

Introduction

Ensuring students are on target in math is critical for at least two reasons. First, the fastest growing occupations in the next decade are projected to be in the science, technology, engineering, and mathematics (STEM) fields and require advanced mathematical and scientific knowledge.¹ Second, math skills act as a filter for better career outcomes since many higher-paying careers (even those that do not require math) require that a student has completed high school or college math prerequisites.² Research shows that achievement in math at the start of high school has a significant effect on students' career aspirations and the courses they choose to take.³ Unfortunately, many American students today are performing at low levels in math — especially those in high-need middle schools where eighth-graders on average consistently test below proficiency — and will have trouble gaining access to these jobs.⁴

To address the underperformance in math of students in high-need middle schools, in 2011, the U.S. Department of Education awarded a five-year Investment in Innovation (i3) grant to Old Dominion University, Johns Hopkins University, and Success for All Foundation (SFAF) to scale up and further test PowerTeaching — a middle-school cooperative learning math program that has shown strong evidence of effectiveness.⁵ This report presents findings from a multiyear evaluation of this i3 scale-up effort.

PowerTeaching

SFAF developed the PowerTeaching model used in the i3 scale-up (PTi3 for short) based on over 25 years of extensive research and refinement of the model. Its components are intended to provide teachers with the necessary tools to incorporate cooperative learning strategies into their instructional practices.⁶ It can be implemented within any school's or district's existing curriculum since it focuses on instructional practices rather than specific math material.

¹Hanushek, Peterson, and Woessmann (2010).

²Sherman (1982).

³Shapka et al. (2006).

⁴The Nation's Report Card (2017).

⁵PowerTeaching was formerly known as Student Teams-Achievement Division. There have been 14 evaluations of this strategy in either primary or secondary schools (Nunnery and Chappell, 2011). The average impact on math test scores was a positive shift of 0.60 of a standard deviation for secondary school students and a 0.13 standard deviation shift for primary school students. The average impact of the studies that met the evidence standards of the What Works Clearinghouse was 0.42.

⁶Cooperative learning, as will be discussed in detail in this report, is different than group learning. In cooperative learning, students work as a team, holding each other accountable for the learning of the group as opposed to group learning where students work together or in close proximity but ultimately are only responsible for their own individual learning.

Figure 1 presents the logic model for PowerTeaching. SFAF recruits school districts that are interested in adopting this approach to math instruction. The leadership in a middle school that will receive PTi3 must commit to supporting the program for three years and the school must provide a part-time math facilitator for each school. SFAF provides training to the principal, facilitator, and math teachers before the school year starts and ongoing training to the math facilitators who in turn train the math teachers in the PTi3 schools. Because mastery of cooperative learning takes time, teachers are expected to participate in continuous improvement meetings, specifically biweekly PTi3 professional development sessions (“component team meetings”) led by the facilitator. The meetings are intended to help teachers set PTi3-specific instructional goals, monitor teachers’ implementation of the program, discuss classroom challenges, and review student progress. Data showing teacher and student progress are shared and discussed.

If the training and ongoing support are adequately delivered, teachers will be able to incorporate PTi3’s instructional strategies into their math classes. In particular, they would place students in longstanding heterogeneous skill groups and provide them with structured opportunities to practice the three key elements of cooperative learning teams:⁷

- Team recognition — students embrace team identity and care about the team’s performance
- Equal opportunities for all students to help the team — all team members, no matter their ability, can contribute to the team goals by improving on their past performance
- Team interdependence — team success depends on each individual’s learning, while an individual’s grade depends only on his or her own performance

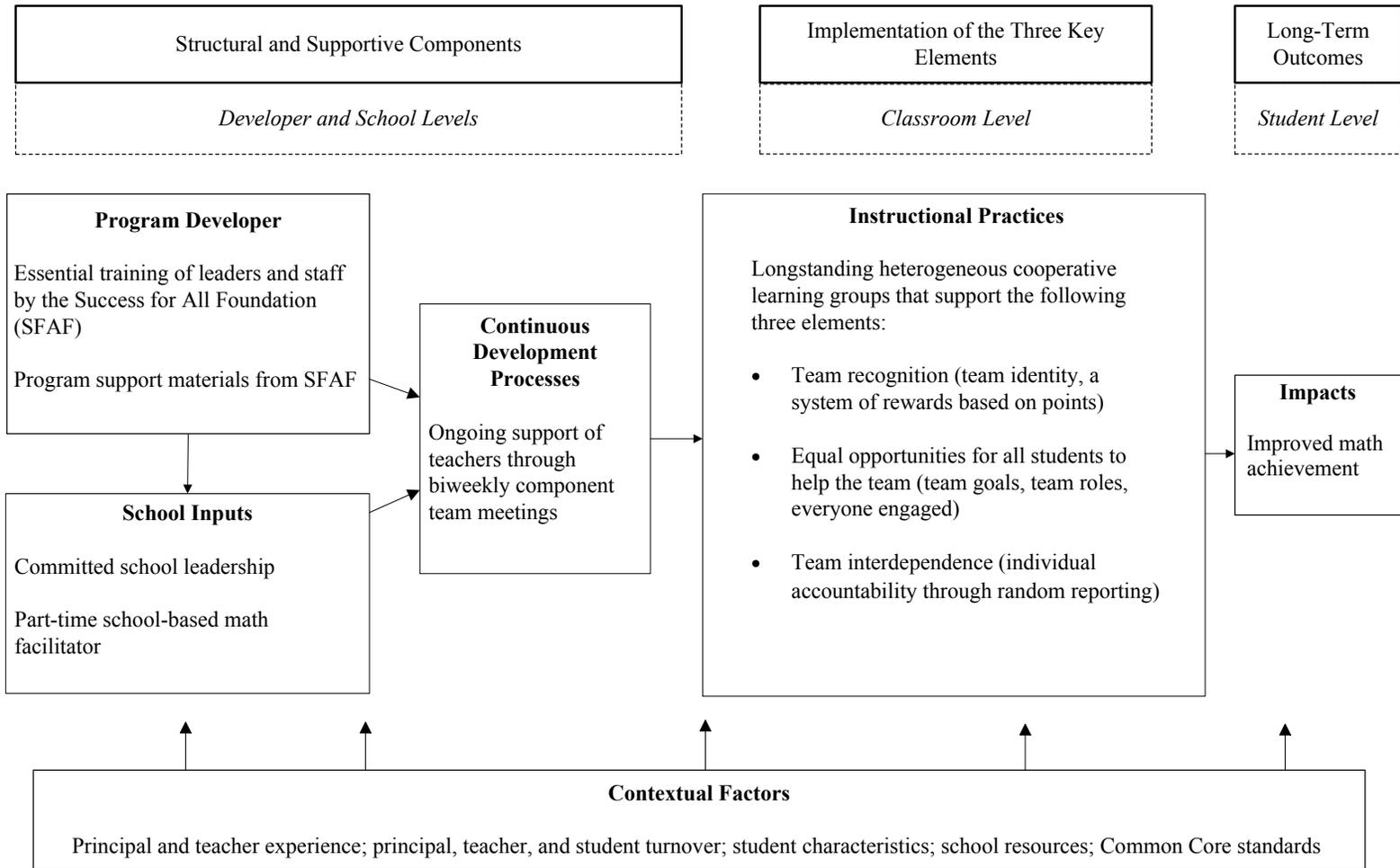
The PTi3 model posits that this team structure coupled with the use of specific cooperative learning strategies creates an environment in which students help each other learn the material and hold each other accountable for both their learning and behavior.

In order for students to feel accountable to themselves and to their team, the three cooperative learning elements must be simultaneously in place. First, students need to care about the recognition earned by their team. The PTi3 model suggests teachers use strategies such as having students name their teams name and decorating a box with team-related pictures.

⁷SFAF calls these features “the three central concepts” (team recognition, equal opportunities for success, and individual accountability). The names of the last two concepts were changed in this report for the sake of clarity.

Figure 1

Logic Model for the Success for All Math PowerTeaching Program in Middle Schools



Second, to prevent less skilled students from disengaging from learning tasks and relying on the more skilled students to do the work, they must also be able to win points for the team. One specific strategy used in PTi3 is developing team goals that require individual team members to do better academically or behaviorally than they have done previously by, for example, improving their academic or behavioral performance, bringing in their homework more often, or increasing their level of team collaboration. Another strategy is to assign “team roles” (such as recorder or leader) to individual students.

The third element, team interdependence, is the heart of cooperative learning. With the first two elements in place, the PTi3 model creates interdependence by using “random reporter” and other similar strategies, in which the team’s performance is assessed based on the performance of one randomly selected team member. In other words, a team earns points based on the quality of a randomly selected team member’s homework, exam, or explanation of solutions to math problems when called on during class. The randomness gives students an incentive to help each other understand the math to ensure that *all* team members can represent the team well when the teacher selects their work. Prior research shows that when these three essential elements of PTi3 are simultaneously in place, PTi3 increases students’ academic performance. Box 1 shows what the combination of the key elements might look like in the classroom and provides an example of simple group work that is now common in middle school math classes.

The Evaluation

In 2012, MDRC began a multiyear evaluation of PTi3 that included an implementation study to document how PTi3 operated, a school-level randomized controlled trial to determine PTi3’s impact on standardized math test scores, and a scale-up study to examine if the goals of the scale-up had been reached. The research team recruited schools in five districts that volunteered to participate in the study and these schools entered the study over the course of two years — 24 schools (all in one state) began in the 2013-2014 academic year (Cohort 1) and 34 schools began in the 2014-2015 academic year (Cohort 2). In each of the five districts in the study, the research team assigned schools to either a program group that received the PTi3 intervention or to a control group that did not. In program group schools, all sixth-, seventh-, and eighth-grade math teachers received PTi3-specific training and support. Teachers in the control group schools received whatever training and support they would usually receive in the absence of the study. The research team continued to recruit schools into 2016 in order to meet the i3 grant’s scale-up goals. These schools were not part of the randomized controlled trial and are referred to as “scale-up schools” in this report.

Box 1

Cooperative Learning in Action: Becoming a Super Team

Students are working in teams of four in Ms. Martin's seventh-grade math class. When it comes time to collect homework, it turns out that three members of the purple team have brought in their homework but one, Rudy, has not. The three team members that did their homework are disappointed; not only will they not receive a team "celebration point" for Rudy's homework, but they will not receive the extra point that teams earn when all members of the team bring in their homework. What is more, the team's goal for the week was to improve on completing homework, so now they are behind in their progress toward reaching their goal and accumulating enough points by the end of the math unit to be rated as a Super Team. Rudy promises to make a greater effort to bring in his homework. Ms. Martin then gives the teams a math problem to solve. Rudy and another team member, Malia, have some ideas about how to approach the problem, but the other two are stumped. Malia and Rudy share their ideas with the others to help them understand the problem and how to solve it, because they know that Ms. Martin might randomly call on any one of them to represent the team and explain its solution to the math problem. After giving the teams enough time to work on the problem, Ms. Martin randomly calls on a member of the purple team, Rosario, to share the solution. Rosario's explanation is clear and correct and he receives a high score that counts towards his grade. His team also receives celebration points. By the end of the unit, Rudy has gotten much better at bringing in his homework and the team continues to work collaboratively on math problems. As a result, the purple team receives enough celebration points to become a Super Team and the class celebrates their achievement.

More Typical Group Work That Is Not Cooperative Learning

Ms. Martin collects everyone's homework at the start of class. Most, but not all, have finished it. She then gives the class a math problem to solve and asks the students to form groups of three. Malia, Rosario, and Marie — who are good friends — get together. Malia has some ideas about how to solve the problem, but the other two are stumped. Malia tells the others her solution and assures them that the answer is right, so Rosario and Marie relax. Ms. Martin calls on Rosario to share the solution to the math problem. Rosario tries his best to repeat Malia's answer but knows he is getting it wrong. "Malia can explain it better," he says. "Ok, Malia, what's the answer?" asks Ms. Martin. Malia's answer is correct and very clearly explained. Ms. Martin is pleased because the class has gotten to hear the correct answer explained well.

The majority of the 58 schools in the study are located in urban areas, such as cities, large towns, and on the outskirts of urban areas, or "urban fringe." Table 1 shows how the study sample compared with middle schools nationwide and with the schools in the PTi3 scale-up effort. The first column of Table 1 shows that more than half of the schools in study are located in the West. In order for the research team to conduct a random assignment study of middle schools within a district, the district had to have more than one middle school. This requirement

Table 1
Background Characteristics for Schools in the Study Sample, Schools in the
PowerTeaching Scale-Up Sample, and Similar Schools in the National Population
(2012-2013 Academic Year)

Characteristics	Study Sample	Scale-Up Sample	National Population ^a
Geographic region (% of schools)			†
Northeast	6.9	12.7	16.4 *
South	25.9	29.6	29.3
Midwest	15.5	22.5	27.4 *
West	51.7	35.2 *	26.7 *
Urbanicity (% of schools)			†
Large or mid-sized city	44.8	25.4 *	22.3 *
Urban fringe and large town	51.7	40.8	29.4 *
Small town and rural area	3.4	33.8 *	48.3 *
Title I status (% of schools)	91.4	91.5	100.0 *
Eligible for free or reduced-price lunch (average % of students)	72.0	68.7	61.1 *
Race/Ethnicity (average % of students)			
White non-Hispanic	11.5	38.0 *	51.5 *
Black non-Hispanic	31.3	23.3 *	17.2 *
Hispanic	50.5	28.4 *	23.0 *
Asian	4.9	3.6	2.8 *
Other	1.7	6.7 *	5.5 *
Male (average % of students)	48.8	50.7 *	52.1 *
Enrollment (average number of students in Grades 6-8)	961.34	686.76 *	251.66 *
Full-time teachers (average % of teachers of Grades 6-8)	51.6	43.2 *	29.7 *
Sample size	58	71	31,102

(continued)

made it very difficult to include small middle schools in rural areas in the study sample, which explains why more schools in the sample were located in urban areas, compared with PTi3 schools generally or schools nationally.

On average, 72 percent of students enrolled in schools in the study were eligible for free or reduced-price lunch, and 91 percent of the schools were designated Title I schools in the 2012-2013 academic year. On average, schools in the study enrolled more non-white students, compared with PTi3 schools generally or schools nationally. Across all schools in the study, 12 percent of students were non-Hispanic White, 31 percent were non-Hispanic Black, and 51 percent were Hispanic. On average, the middle schools in the study enrolled about 1,000 sixth-, seventh-, eighth-graders and employed 52 teachers, making these middle schools much larger than PTi3 schools generally or middle schools nationally.

Table 1 (continued)

SOURCE: 2012-2013 Common Core of Data.

NOTES: Due to missing values for some variables, the number of schools included varies by characteristic.

"*" indicates a statistically significant difference ($p\text{-value} \leq 0.05$) between the study sample and either the scale-up sample or the national population of schools for given characteristics. A two-tailed t-test was applied to each comparison.

"†" indicates a statistically significant difference ($p\text{-value} \leq 0.05$) between the study sample and either the scale-up sample or the national population of schools for categorical characteristics. A chi-square test was applied to each of such comparisons.

To examine whether there is any systematic difference between the study sample and the scale-up sample, an F-test was conducted in a regression model controlling all school characteristics reported in this table ($p = 0.892$). A similar test was conducted for systematic difference between the study sample and the national population ($p < 0.001$).

^aThe national population includes Title I schools with Grades 6 through 8 only.

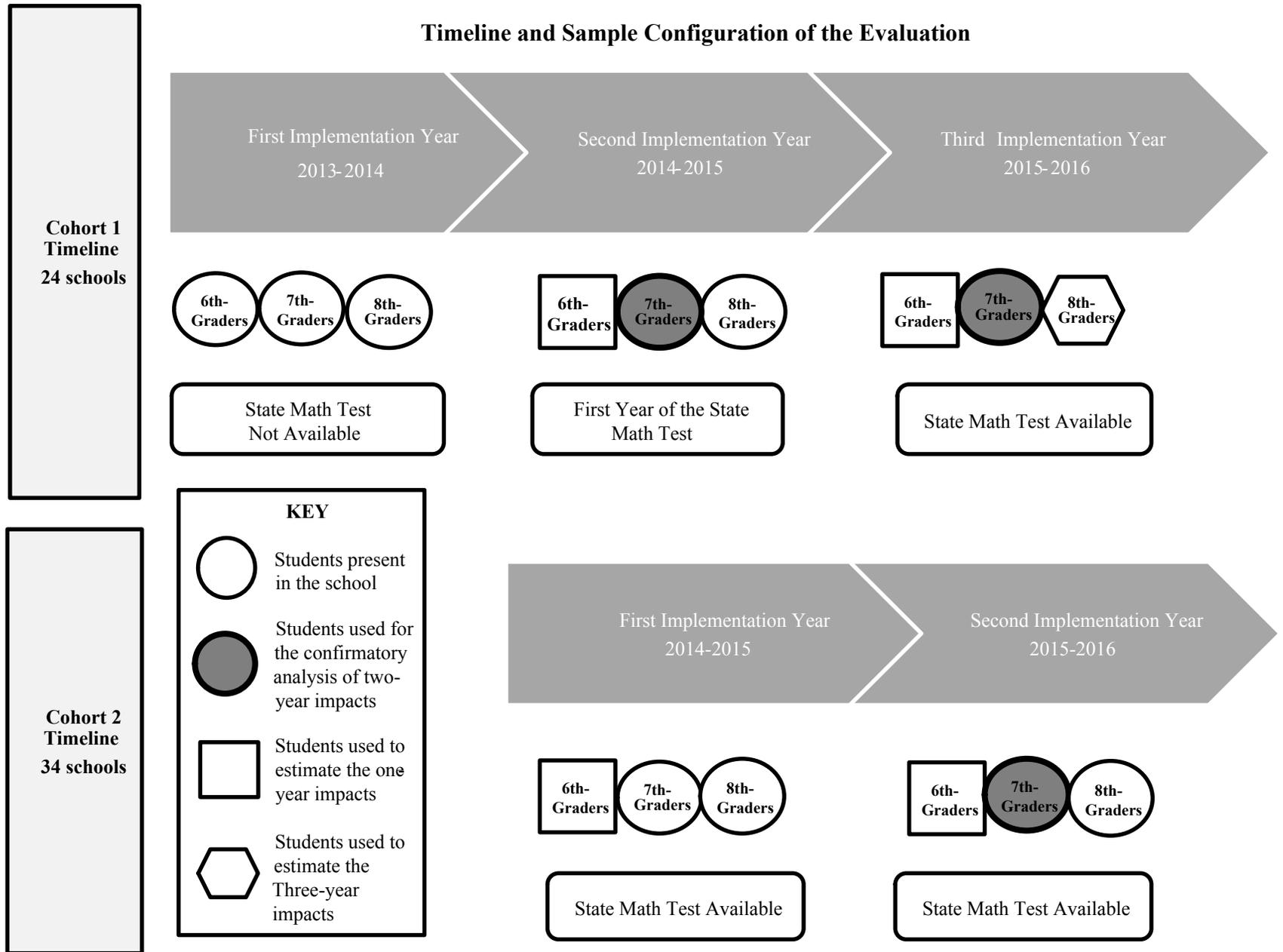
The research team collected data from many sources for the evaluation. The implementation findings use information gathered from both program and control group schools from in-person interviews of school staff, teacher surveys, and instructional logs completed by teachers on a set of randomly chosen students. The impact findings are based on school records collected for students attending both program and control group schools through the 2015-2016 academic year.⁸

Given the time frame of the evaluation, the impact analysis focuses on the one- and two-year (steady state) impacts of PTi3 on students who could have been exposed to the program from the beginning of middle school. Figure 2 indicates which sets of students the research team used to estimate these impacts, showing Cohorts 1 and 2 separately. Because all schools in Cohort 1 were in the same state and the state was refining a standardized test that aligned with its new state standards in the 2013-2014 academic year, test scores were not available that year, making it impossible to observe the one-year impacts in Cohort 1 schools. The one-year impacts are estimated for the sixth-graders (denoted by the square in Figure 2), and the two-year impacts (the confirmatory test) are for seventh-graders who could have experienced PTi3 in both the sixth and seventh grades (denoted by the shaded circle). There is only one set of students who could have been exposed to PTi3 for three years, namely those in Cohort 1 who were eighth-graders in the 2015-2016 academic year (denoted by the hexagon). While the sample of eighth-graders is too small to rigorously test the impact on this group, the research team examined the effect, not expecting statistical significance.

⁸A working paper that is based on data from the 2014-2015 academic year provides a detailed description of these data and is available on the MDRC website in the "Supplemental Materials" for this report (Rappaport et al., 2017).

Figure 2

Timeline and Sample Configuration of the Evaluation



For interested readers, a longer working paper based on the 2014-2015 academic year is available on the MDRC website in the “Supplemental Materials” for this report.⁹

Most multiyear evaluations encounter obstacles that impinge on the researchers’ original plans. A major challenge this evaluation faced was that one of the states in which the study was conducted was adopting new educational standards during the study period. Teachers in program group schools were not only being asked to adopt PTi3 instructional practices, the material they were expected to cover was dramatically changing. Similarly, students were being assessed with new standardized tests. Thus, the educational environment in which teachers in program group schools tried to implement PTi3 may not have been representative of the usual environment in a less stressful time.

The Findings

This section follows the logic model in Figure 1. It begins by determining whether the PTi3 components external to the classroom were in place, then examines the instructional practices teachers used and the dynamics within the cooperative learning teams, and finally assesses the impacts the program generated from this level of implementation. The section concludes with a description of the PTi3 scale-up experience.

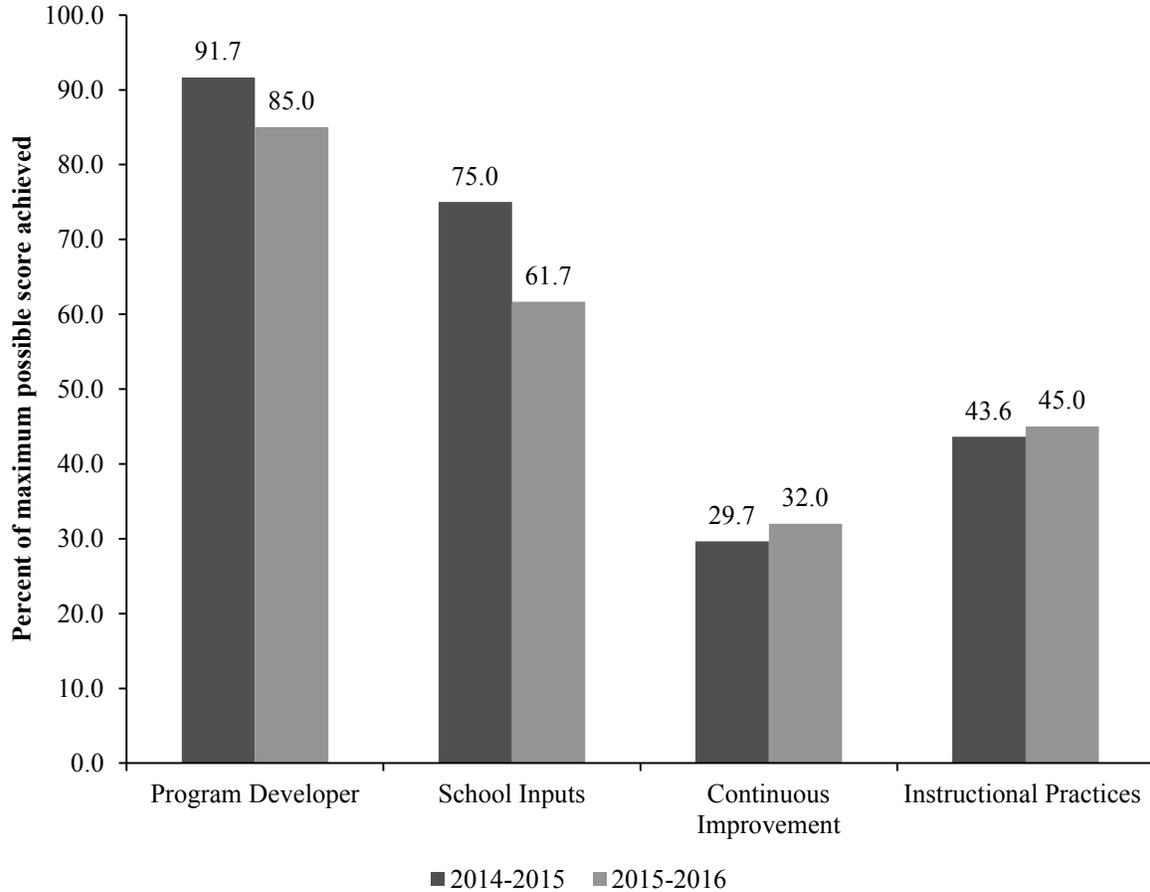
- **SFAF and the PTi3 schools provided the requisite time, staff, and materials needed to support teachers in their implementation of the program.**

To gauge the fidelity with which schools implemented PTi3, the research team assessed select scores that each school earned on the School Achievement Snapshot (Snapshot), an instrument created by SFAF to guide schools in a continuous improvement process. The facilitators and SFAF coaches jointly decided on which scores from the Snapshots would be used to gauge fidelity. At the beginning of the study, SFAF determined that, for the purpose of this study, if a school achieves a total score of 50 percent or more of the maximum possible score on an implementation measure, then the school should be deemed as having implemented that dimension with adequate fidelity, even if it may be operating PTi3 with fewer than optimal PTi3 teachers or program components in place.

Figure 3 shows that schools received high scores in the Program Developer category, or SFAF’s provision of initial training and materials to the schools. The schools also earned

⁹Rappaport et al. (2017). The paper provides many more details on the study’s methodology, as well as more detailed evidence supporting the findings.

Figure 3
School Achievement Snapshot Scores, by Category and Academic Year, Study Schools Only



SOURCE: 2014-2015 and 2015-2016 School Achievement Snapshots.

NOTE: The sample includes 12 schools in Cohort 1 and 18 schools in Cohort 2. For Cohort 1 schools, the 2014-2015 and 2015-2016 academic years were Years 2 and 3 of implementation. For Cohort 2 schools, the 2014-2015 and 2015-2016 academic years were Years 1 and 2 of implementation.

adequate scores in the School Inputs category, or schools provision of the required resources such as a part-time facilitator.¹⁰ However, while implementation was over 50 percent on both of these measures in both years, scores declined in the 2015-2016 academic year.

¹⁰The scores for Program Developer and School Inputs categories were each based on two Snapshot items.

- **Few schools carried out the ongoing continuous improvement activities specified by the PTi3 model, as the low scores (below the adequacy threshold of 50 percent) in the Continuous Improvement category indicate. Few schools collected and used student assessment data to drive instruction, and most teachers did not receive the kind of training and support from the school-level facilitator that the PTi3 model specified in order to help teachers create cooperative learning teams in their classrooms.**¹¹

Figure 3 shows that the study schools scored about 30 percent of its maximum value in the Continuous Improvement category both years. The score was slightly higher in the 2015-2016 academic year, but still below the adequacy threshold.¹² As shown in Table 2, the schools held component team meetings less often than twice a month as prescribed by the PTi3 model. When these meeting did occur, they often did not focus on setting goals, monitoring program implementation, and going over student data — the core purpose of these meetings — since very few schools collected the necessary data on student assessment and teacher implementation. Finally, very few schools used the coaching method prescribed by SFAF to help math teachers master cooperative learning in their classrooms. Thus, the average score in this category is low across the sample. Despite receiving less support than prescribed by the PTi3 model, a significantly greater proportion of teachers in program group schools than in control group schools reported in surveys that they had received coaching.¹³

- **Students in both program and control group schools worked in groups often, but math teachers in program group schools were more likely than teachers in control group schools to put students in longstanding mixed-ability groups (per the PTi3 model), which are more conducive to building strong cooperative learning teams. Students in program group schools spent more time each day in these teams.**

The educational standards recently adopted by the states in which the evaluation was conducted support frequent student collaboration in math classes, and the study found that most classrooms in the study sample included group work. However, simply putting students in a

¹¹One of the main responsibilities of the school-level facilitator is to support and train teachers in the PTi3 model. SFAF coaches train the facilitators, who are then expected to train the teachers in their schools.

¹²The scores for the Continuous Improvement were based on four Snapshot items.

¹³Math teachers in both program and control schools were surveyed at the end of the 2014-2015 academic year.

Table 2
School Achievement Snapshot Scores for Items
Related to Schoolwide Structures and Instructional Practices,
Study Schools (2014-2015 and 2015-2016 Academic Years)

Item	2014-2015 Percent of maximum possible score	2015-2016 Percent of maximum possible score
<u>Program developer</u>		
All leaders and staff have received essential training	83.3	76.7
Materials for program implementation are complete	100.0	93.3
<u>School inputs</u>		
School-based math facilitator is a part-time position	80.0	56.7
The principal is fully involved with PowerTeaching	70.0	66.7
<u>Continuous improvement processes</u>		
Component teams meet at least twice a month	43.3	45.0
Each teacher submits a quarterly classroom assessment summary	18.3	5.0
Instructional component teams set targets, chart progress, and work to meet targets	25.0	31.7
The school-based math facilitator uses PowerTeaching coaching process	18.3	34.5
<u>Instructional practices</u>		
<i>Teachers...</i>		
Use basic lesson structure, objectives, and available media regularly and effectively	57.7	54.0
Use think-pair-share, whole-group response, or random reporter frequently and effectively	45.3	46.7
Provide time for partner and team talk to allow mastery of learning objectives by all students	55.7	62.0
Facilitate partner and team discussion	38.3	46.3
Randomly select students to report for their teams during class discussion, use rubrics to evaluate responses, and award teams with points	30.7	37.3
Effectively summarize and address misconceptions or inaccuracies during class discussion	20.0	14.7

SOURCES: 2014-2015 and 2015-2016 School Achievement Snapshots.

NOTES: The sample includes 12 schools in Cohort 1 and 18 schools in Cohort 2. For Cohort 1 schools, the 2014-2015 and 2015-2016 academic years were Years 2 and 3 of implementation. For Cohort 2 schools, the 2014-2015 and 2015-2016 academic years were Years 1 and 2 of implementation.

group does not necessary mean they will collaborate as a cooperative learning team to solve problems. Cooperative learning teams require structure and guidance and take time to gel.¹⁴ The teacher survey results show that, on average, student groups in PTi3 schools stayed together for longer periods of time, allowing students to bond as a team. Teachers in PTi3 schools were also less likely to separate individual students from their groups when academic or behavioral issues arose. Finally, groups in PTi3 schools were more likely to have four students,¹⁵ compared with groups of three students in control group schools. Thus, longstanding, mixed-ability teams, as prescribed by the PTi3 model, were more prevalent in program group schools. Students in PTi3 schools also spent significantly more time working in groups (59 percent of class time) than students in control group schools (40 percent of class time). When teachers logged the activities of a randomly chosen set of students, the data showed that students in an average math class in program group schools spent significantly more time doing team work than students in control group schools by an average of 10 minutes (31 minutes versus 21 minutes).

- **Despite differences in how math teachers in program group and control group schools structured student groups and used teamwork, fewer than half of the teachers in program group schools incorporated the PTi3 instructional strategies in their classrooms. Many teachers in program group schools put cooperative learning team structures in place but did not use the strategies that move group work to true cooperative learning teams.**

Students in cooperative learning teams collaborate on problems, not just to find the right answer, but to ensure all team members know how to solve similar problems, or have learned the skill. The scores in the Instructional Practices category in Figure 3 and the scores on the corresponding items in Table 2 indicate that teachers in program group schools were not implementing, with adequate fidelity, the PTi3 strategies intended to foster cooperative learning teams. Table 2 shows that the teachers' scores on using basic PTi3 lesson plans and objectives and providing time for partner and team talk were just over the adequacy threshold of 50 percent. However, teachers scored below the adequacy threshold on all the other items in the Instructional Practices category. Importantly, teachers overall scored below the adequate mark on using the random reporter, a critical strategy that gives students an incentive to ensure that all team members can answer questions correctly.

¹⁴Tuckman (1965).

¹⁵With four students in a group, there is more opportunity for partnership as it is possible to have students work in pairs within the group.

- **The dynamics between students inside the groups did not appear to differ greatly between program and control group schools, although a few different behaviors were observed.**

The instructional logs provide a detailed picture of what students did in their groups.¹⁶ On average, students in PTi3 schools spent 31.3 minutes working in groups, compared with 21.3 minutes for students in control schools. The top row of Table 3 shows that, on average, students in PTi3 schools spent 8.4 minutes jointly solving math problems using an algorithm, compared with 5.2 minutes for students in control group schools. The average student in PTi3 schools also spent a little more time than students in control group schools engaging in activities not related to the group assignment (3.8 minutes versus 1.9 minutes), with students in PTi3 schools spending 1.5 of those minutes bothering other students (versus 0.6 minutes for students in control group schools). No other significant differences were observed.

Subgroups of students defined by skill level drove these differences. The students in PTi3 schools that the teacher rated as in the top third of their math class had similar in-group experiences in the program and control group schools. Students in the middle or bottom third of their class demonstrated changed behaviors. These students were more likely than their counterparts in control group schools to jointly solve math problems. (Students in the middle third spent 8.5 minutes jointly solving math problems, compared with 4.8 minutes for their counterparts in control group schools; students in the bottom third spent 7.1 minutes, compared with 3.7 minutes for their counterparts in control group schools). Differences in negative behaviors were entirely concentrated among students ranked in the bottom third. (These students in the lower third spent 6.4 minutes engaging in negative behaviors, versus 3.0 minutes for their counterparts in control group schools.)

- **The three elements that must be simultaneously in place to create interdependent cooperative learning teams — team recognition, equal opportunity for all students to contribute to the team’s success, and team interdependence — were not consistently in place. Strategies to promote interdependence, in particular, were not well understood or implemented by teachers.**

Teacher surveys and focus groups indicate that some teachers in PTi3 schools used strategies supporting team recognition or team identity, but primarily at the beginning of the

¹⁶In the 2014-2015 academic year, math teachers in program and control group schools were asked to fill out a daily log for a two-week period. Each entry would focus on one student, and teachers would complete daily entries for up to eight students that researchers randomly selected from their classes. Logs record what the student did during class.

Table 3

**Impact of PowerTeaching on Minutes Students Spent Doing Group Activities During the Math Block
(Activities Are Not Mutually Exclusive)**

Activity	Mean Program Minutes	Mean Control Minutes	Estimated Impact ^a			90% Confidence Interval
				-4	0	
Solving mathematical problems by using an algorithm	8.42	5.22	3.19 *			(0.74 5.64)
Discussing and working on a problem with multiple solution methods	8.06	6.98	1.08			(-2.64 4.80)
Applying mathematical concepts to "real world" problems	7.82	6.84	0.98			(-2.71 4.67)
Representing and analyzing relationships using tables or models	6.71	6.35	0.36			(-3.38 4.10)
Analyzing data to make inferences or draw conclusions	5.54	5.80	-0.26			(-3.76 3.24)
Explaining a solution to a problem to other students	5.47	4.89	0.57			(-1.52 2.67)
Helping other students solve math problems	5.20	4.68	0.52			(-1.42 2.47)
Asking other students clarifying questions	5.06	4.29	0.77			(-1.20 2.73)
Asking other students for help in solving a math problem	4.85	4.17	0.67			(-1.29 2.64)
Exchanging work with other students for review and checking	4.82	4.99	-0.17			(-2.00 1.65)
Suggesting a strategy to a partner or group members	4.61	4.23	0.38			(-1.52 2.28)
Building on or challenging the ideas of other students	3.93	3.99	-0.06			(-1.82 1.70)
Engaging in discussion or activities not related to assigned activity	3.77	1.94	1.83 *			(0.51 3.16)
Jointly reading a textbook or supplementary materials	2.48	3.44	-0.97			(-3.28 1.35)
Discussing and jointly working on multiple choice exercises	2.23	2.33	-0.11			(-1.39 1.18)
Making fun of, belittling, or bothering a partner or group members	1.49	0.64	0.85 *			(0.27 1.43)

(continued)

Table 3 (continued)

SOURCE: Teacher logs administered in spring 2015.

NOTES: Sample consists of 2,941 logs (1,567 in the program group and 1,374 in the control group). There was a low response rate from teachers in one school district. As a result, all instructional logs (n = 84) from this district were dropped.

All estimations are based on a three-level hierarchical model with individual logs nested within teachers and teachers nested within schools. A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by an asterisk (*) when the p-value is less than or equal to 5 percent.

In the instructional logs, teachers were asked how much time randomly selected students spent working in three different configurations during the mathematics period: in groups, in pairs, and individually. If teachers indicated that selected students worked in a group during the mathematics period, they were asked approximately what proportion of group time the students spent engaging in specific activities using a five-category scale: 0 percent, less than 10 percent, 10 to 25 percent, 26 to 50 percent, and more than 50 percent. In the analysis of the logs, the research team converted the proportion of group time into minutes by calculating the midpoint of the ranges in this scale and then multiplying the midpoint of the selected range by the total amount of time spent working in a group. If selected students did not spend time working in a group, the time spent on each activity was set to zero.

^aThe horizontal lines on each side of the impact estimate represent the "confidence interval" — that is, the range of estimated values of the impact, within which there is a 90 percent probability that the true value falls. The impact estimate is statistically significant when the range of the confidence interval (defined by the upper and lower bounds) crosses the vertical line.

year. Teachers seemed to inconsistently use strategies promoting equal opportunities for all students to contribute to the team's success. Data from focus groups indicate that some teachers in program group schools used strategies such as rewarding team points, assigning team roles, and developing team goals, but they did not do so consistently. Given that the logs showed misbehavior was higher in program group schools than in control group schools — and concentrated in the students in the bottom third of their class — it appears that more needed to be done to fully engage this group. Finally, teachers did not seem to understand random reporter and other strategies that promote group interdependence very well.

While many teachers in both program and control group schools reported that they sometimes used a random reporter strategy to call on students to answer math questions, some of them reported that they allowed randomly selected students to pass the questions to a teammate or to confer with their team before answering, and some described only randomly picking students who had not yet had a chance to share a response.¹⁷ As a result, these teachers did not create team interdependence since the most advanced team member often answered questions when a less advanced team member could not, without repercussions such as withholding team points or rewards. Teachers in PTi3 schools demonstrated a similar misunderstanding of interdependence strategies when they reported letting students work on their tests together. This

¹⁷Teachers also reported using other ways of calling on students, including calling on volunteers, calling on students who rarely volunteered or who were not paying attention, or calling on lower-performing students when it was evident that they had a strong response to share.

approach, once again, allowed for the most advanced team member to correct the errors made by the less advanced members, rather than ensure that all team members could correctly complete the test themselves. Teachers would have created group interdependence if they had encouraged team members to prepare for the test together but required them to take their tests independently.

- **Based on the analysis of the full sample, the PTi3 program did not produce statistically significant impacts on math performance of the sixth-, seventh-, or eighth-graders, as measured by their scores on the state’s standardized math test.**

As Table 4 shows, the estimated one-year impact on sixth-graders’ standardized math score is approximately zero and insignificant. The estimated two-year impact on math scores for seventh-graders, our confirmatory hypothesis, is also close to zero (with a p-value of 0.931). Finally, the estimated impacts on eighth-graders in Cohort 1 — who were the only students in the sample that could have experienced PTi3 during all three years of middle school — is also not statistically different from zero (with a p-value of 0.222). However, this estimate is based on a sample of less than half the schools in the study, which is not large enough to determine statistical significant at the desired 80 percent level of power. (In other words, the estimation is underpowered).

It is important to note that schools in this sample were experiencing the PTi3 program with differing maturity, and some schools were administering a new state standardized math test for the first time. The research team conducted several sensitivity analyses to see if the estimated impacts differed by program maturity or if they were obscured because some schools were using the new standardized test for the first time. None of these analyses yielded statistically significant findings.¹⁸

- **Impact analyses on subgroups of the confirmatory sample, however, found that the impacts differed significantly by cohort. The estimated impact for seventh-graders in Cohort 2 who could have received the**

¹⁸An analysis of only students in seventh grade during their schools’ second year implementing PTi3 (seventh-graders in Cohort 1’s 2014-2015 academic year and seventh-graders in Cohort 2’s 2015-2016 academic year) did not reveal any significant impacts. In addition, if the seventh-graders in Cohort 1’s 2014-2015 academic year were dropped from the analysis because they were assessed using a new state test for the first time that year and if the estimated impact is based on only seventh-graders in the 2015-2016 academic year, then the estimated impact is 0.01 and not significant (p-value of 0.91). However, the study is not powered for this kind of subgroup analysis, therefore all results presented here are considered exploratory and should be interpreted as such.

Table 4

**Impact of PowerTeaching on Students' Average Math Achievement
by Grade, for Analysis Samples**

Sample	Program Group	Control Group	Estimated Impact (in Effect Size or Percentage Point)	P-Value	95% confidence interval of the impact ^a (in standard deviation units)		
					-0.25	0	0.25
<u>Grade 6 full sample (exploratory)</u>							
Standardized state math test score	-0.07	-0.05	-0.01	0.747			
Percentage at or above proficiency level	30.1	31.8	-1.8	0.304			
Number of schools	58						
<u>Grade 7 full sample (confirmatory)</u>							
Standardized state math test score	-0.05	-0.05	0.00	0.931			
Percentage at or above proficiency level	32.1	32.3	-0.2	0.896			
Number of schools	58						
<u>Grade 8 full sample (exploratory)</u>							
Standardized state math test score	-0.16	-0.06	-0.09	0.222			
Percentage at or above proficiency level	21.2	27.0	-5.8	0.049 *			
Number of schools	24						

(continued)

Table 4 (continued)

SOURCES: School district student records data from the 2015-2016, 2014-2015, 2012-2013 academic years (for Cohort 1), and the 2013-2014 academic year (for Cohort 2).

NOTES: The analysis sample consists of students from 58 schools (30 program group schools and 28 control group schools) and includes any student who had a valid spring test score in the spring of 2015 or spring of 2016.

The student sample size for Grade 6 (Cohort 1 and Cohort 2 schools in both 2014-2015 and 2015-2016 academic years) is 32,288 students (17,354 in the program group schools and 14,934 in the control group schools). The sample size for Grade 7 (Cohort 1 schools in the 2014-2015 and 2015-2016 academic years, and Cohort 2 schools in the 2015-2016 academic year) is 26,808 students (14,489 in the program group schools and 12,319 in the control group schools). The student sample size for Grade 8 (Cohort 1 schools in the 2015-2016 academic year) is 9,139 students (5,117 in the program group schools and 4,022 in the control group schools).

The estimated impacts are based on a two-level model with students nested within schools, controlling for random assignment block and school- and student-level covariates. The program group and control group columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the program group schools as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

"*" indicates a statistically significant difference (p-value ≤ 0.05) between comparison groups for given characteristics. A two-tailed t-test was applied to each comparison.

^aThe confidence intervals are for the estimated impacts on the standardized test scores.

PTi3 program for two years was positive, but not statistically different from zero, while the estimated impact for similar seventh-graders in Cohort 1 was negative and significant.

The top panel of Table 5 shows that the impact for Cohort 1 is -0.12 and statistically significant (with a p-value of 0.047), while the impact for Cohort 2 is 0.08 and not statistically significant (with a p-value of 0.285). These two estimated impacts are statistically different from each other (p-value of 0.044). An impact analysis by cohort for sixth-graders (not shown in Table 5) yielded no statistically significant findings. The impact for sixth-graders in Cohort 1 was -0.10, while it was 0.04 for sixth-graders in Cohort 2.

One possible explanation for Cohort 1's smaller impact findings could be that the study period overlapped with the year when one of the states in which the study took place adopted the state's new educational standards. This state — where all Cohort 1 schools are located — introduced a standardized test to hold schools accountable for the new standards the same year schools in Cohort 1 entered the study. Teachers in Cohort 1 schools were being asked to adopt PTi3 instructional strategies *and* to cover new material. Interviews with teachers indicate that they struggled with how best to teach the material in accordance with the new standards. SFAF responded by developing teaching resources for math that aligned with the new standards during the first two years of the study. However, it took SFAF a few years to fully develop and refine these curricular materials. Thus, it is not surprising that students who experienced the

Table 5
Impact of PowerTeaching on Students' Average Math Achievement
for Grade 7 Analysis Sample, by Cohort and Student Subgroup

Subgroup	Number of Observations	Program Group	Control Group	Estimated Impact (in Effect Size or Percentage Point)	P-Value
<u>By cohort (exploratory)</u>					†
Cohort 1					
Standardized state math test score	18,539	-0.11	0.01	-0.12	0.047 *
Percentage at or above proficiency level	18,539	22.3	26.6	-4.3	0.056
Cohort 2					
Standardized state math test score	8,269	-0.01	-0.09	0.08	0.285
Percentage at or above proficiency level	8,269	38.5	35.9	2.6	0.368
<u>By performance rank at baseline</u>					
Top third	7,408	0.68	0.72	-0.04	0.589
Middle third	7,927	0.04	0.04	0.00	0.920
Bottom third	7,680	-0.74	-0.77	0.03	0.521
<u>By proficiency level at baseline</u>					
At or above proficiency	14,604	0.43	0.45	-0.02	0.687
Below proficiency	8,378	-0.60	-0.61	0.01	0.807
<u>By gender</u>					
Boys	13,696	-0.09	-0.08	-0.01	0.835
Girls	13,023	0.00	0.00	0.00	0.931
<u>By race/ethnicity</u>					
Hispanic	18,118	-0.19	-0.14	-0.04	0.377
White, Non-Hispanic	2,917	0.39	0.46	-0.07	0.397
Black, Non-Hispanic	3,451	-0.25	-0.28	0.02	0.623
<u>By family income</u>					
Eligible for free and reduced-price lunch	19,458	-0.08	-0.09	0.01	0.753
Not eligible for free and reduced-price lunch	4,880	0.20	0.30	-0.10	0.149
<u>By English-language learner (ELL) status</u>					
ELL	4,036	-0.70	-0.63	-0.08	0.125
Non-ELL	18,773	0.08	0.07	0.01	0.838
<u>By special education (SPED) status</u>					
SPED	2,985	-0.82	-0.84	0.02	0.744
Non-SPED	23,499	0.07	0.07	0.00	0.975

(continued)

Table 5 (continued)

SOURCES: District student records data from the 2015-2016, 2014-2015, and 2012-2013 academic years (for Cohort 1), and the 2013-2014 academic year (for Cohort 2).

NOTES: The Grade 7 analysis sample consists of students from 58 schools (30 program group schools and 28 control group schools) and includes any student who had a valid spring test score in the spring of 2015 (Cohort 1) or spring of 2016 (both Cohorts 1 and 2). The sample size for Grade 7 is 26,808 students (14,489 in the program group schools and 12,319 in the control group schools).

The estimated impacts are based on a two-level model with students nested within schools, controlling for random assignment block and school- and student-level covariates. The program group and control group columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the program group as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

The difference between the impact estimates for Cohorts 1 and 2 is significant at the 5 percent level ($p = .044$).

"*" indicates a statistically significant difference ($p\text{-value} \leq 0.05$) between comparison groups for given characteristics. A two-tailed t-test was applied to each comparison.

"†" indicates a statistically significant difference ($p\text{-value} \leq 0.05$) in impacts among subgroups.

early materials (students in Cohort 1 schools) fared less well on the standardized test than those who experienced the more refined materials (students in Cohort 2 schools).

- **A subgroup analysis of students based on socioeconomic and academic characteristics did not reveal that impacts varied across these groups, nor did it find any statistically significant impacts.**

The research team also explored potentially heterogeneous impacts across different student subgroups defined by baseline characteristics. These subgroups included those defined by students' performance levels in math, gender, race or ethnicity, English language learner (ELL) status, poverty status, and special education status. Table 5 presents results of this exploratory analysis for the sample of seventh-graders. Overall, the findings indicate that the PTi3 program did not produce any statistically significant impacts across a range of student subgroups, and these findings did not seem to vary across such subgroups.

- **As of October 2016, Old Dominion University calculated that the PTi3 program served 132,166 students in 106 high-need schools over the course of the i3 scale-up effort, which put them at 98 percent of their target of 135,000 students. Most of these scale-up schools were smaller and implemented PTi3 slightly better than the schools in the study, according to Snapshot scores.**

The second column of Table 1 shows that the scale-up schools, similar to the schools in the study, were high-need schools (92 percent were Title 1 schools), but they were more

geographically dispersed, less urban, and smaller in size than schools in the study sample.¹⁹ The total of 132,166 students served includes the students in 8 pilot schools,²⁰ the 30 schools in the study's program group, and 41 additionally recruited scale-up schools.²¹ After recruitment for the study ended, districts and schools could join the project as scale-up sites and were eager to do so. Of the schools offered to participate in the scale-up, about 90 percent joined the project. Scale-up schools did not have to go through the random assignment process and the i3 grant paid for implementation costs to varying degrees, depending on the year that a school was recruited and its level of participation in helping to improve the program.

The research team also gauged the fidelity with which these additional scale-up schools implemented PTi3 using their Snapshot scores in the 2015-2016 academic year, shown in Figure 4.²² Similar to the program group schools, the scale-up schools scored above the adequacy threshold in the Program Developer and School Inputs categories. Unlike the schools in the study, the scale-up schools also scored above the adequacy threshold in the Continuous Improvement category. However, the scale-up schools similarly fell short of the adequacy threshold in the Instructional Practices category. Thus, while implementation of PTi3 remained weak, the scale-up schools implemented the program with more fidelity than schools in the study.

Conclusion

The evaluation's findings show that training middle school math teachers to create an effective cooperative learning environment is hard. When a school-based intervention requires both structural and instructional changes, as PTi3 did, it is not uncommon to observe that it is easier for the schools to put the new structures in place than for the teachers to change their instructional practices. For example, a recent evaluation of Diplomas Now — a comprehensive, schoolwide model that includes structural changes, the introduction of new instructional

¹⁹In order to conduct a random assignment of middle schools within a district, the district had to have more than one middle school. This requirement made it almost impossible to include small rural middle schools in the study sample.

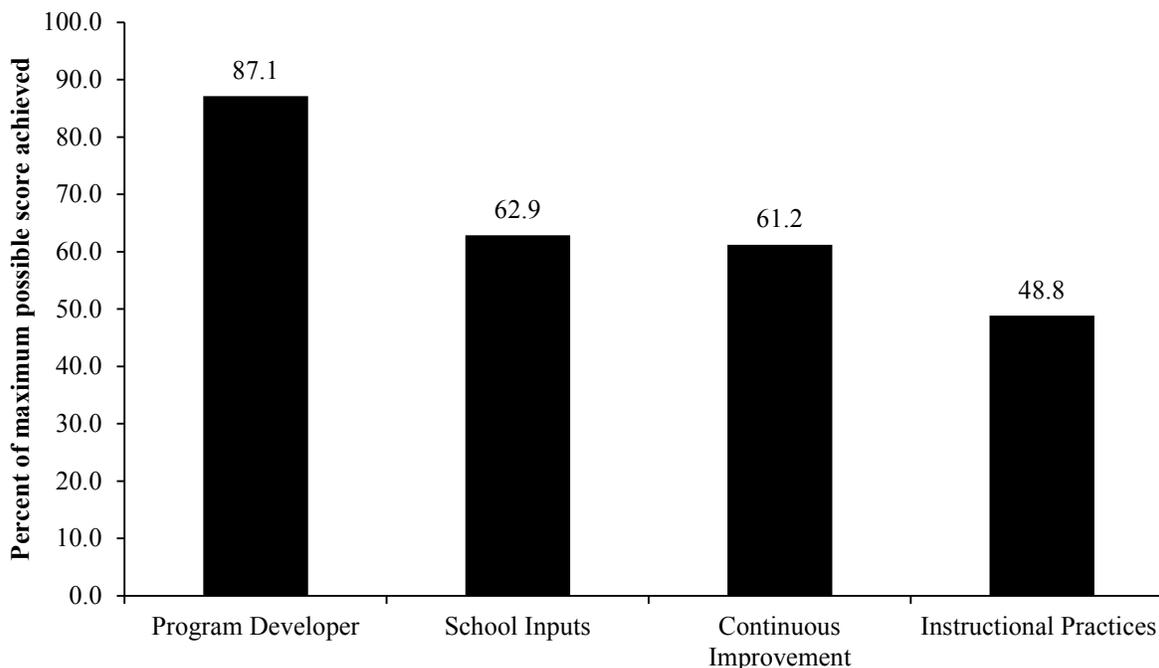
²⁰In the first year of the PTi3 project, before the study schools and the scale-up schools were recruited, the program was piloted in eight schools.

²¹By the 2016-2017 academic year, a total of 157,183 students had been served in 134 schools, which put the number of students being served over the 135,000 student target. The 28 additional schools were the study's control group schools, which received access to the program after the final follow-up data was collected.

²²There were a total of 41 non-study scale-up schools by the 2015-2016 academic year, but Snapshot scores were only available for 39 schools. The Snapshot scores for schools in the study were from the 2015-2016 academic year. The scale-up schools also varied in the number of years they had been implementing the program.

Figure 4

**2015-2016 School Achievement Snapshot Scores, by Category,
Scale-Up Schools Only**



SOURCE: 2015-2016 School Achievement Snapshots.

NOTE: In total, there were 41 non-study scale-up schools by 2015. Snapshot scores were available for 39 of these schools.

materials and curricula, teacher and administrator coaching and support, and a student early warning system — found that, while schools were reasonably successful at implementing many parts of the model, they were least successful at making the instructional and curricular changes.²³ These findings are quite similar to those found in the PTi3 evaluation. Changing instruction and teaching practices may just take more time and concerted attention. Similar to earlier studies, the present evaluation’s findings also suggest that placing students in mixed-ability groups has a positive impact on student performance only if true interdependence among group members occurs. However, since math teachers today incorporate student group work in their classrooms more frequently, the potential to improve student performance in math through effective cooperative learning strategies is great. Research shows that enhancing the instructional practices of math teachers has the largest marginal effect on improving performance among secondary students, compared with other commonly used practices such as changing the math curriculum or supplementing teacher instruction with computer-assisted instruction.²⁴

²³Sepanik et al. (2015).

²⁴Slavin, Lake, and Groff (2009).

Given that overall implementation of PTi3 among schools in the study was weak, the findings are not a fair measure of the program's true effect when properly implemented. Implementation of PTi3 during the study period was hampered by an unusual event — states where the evaluation was conducted were adopting new educational standards. As a result, school districts were introducing new and much more difficult standardized tests that aligned with these standards. Principals and teachers were struggling with how best to instruct the material in accordance with the new standards and standardized tests. While teachers attempted to implement PTi3 and adopt its instructional strategies, they also struggled to teach new and different material and prepare their students for new tests. Thus, the continuous improvement component of the PTi3 model, which is critical to helping teachers master effective cooperative learning strategies, did not occur at the level the model specifies.

This finding points to an important yet underemphasized need to help teachers across the country understand how to use group work in a way that creates an effective cooperative learning environment. Unlike the 1980s and 1990s, when middle school teachers mostly relied on the traditional teaching method of demonstration followed by individual student practice,²⁵ middle school teachers today almost universally incorporate group work into their classroom instruction. This study showed that 96 percent of teachers in the control group schools were using peer-learning strategies, namely partner or group work, in their classes. However, the qualitative data also shows that, for the most part and similar to the teachers in the program group schools, they were not creating environments of positive interdependency among group members. Thus, while it is not difficult to convince math teachers that group learning activities are a useful instructional strategy, there still remains a crucial need to help teachers turn this group work into effective cooperative team learning.

One solution may be to scale up a more refined version of the PTi3 model to more schools. Indeed, the scale-up schools appeared to implement PTi3 with greater fidelity. However, this study's findings show that no matter what instructional program schools adopt to improve cooperative learning, they must provide sufficient training and support to ensure teachers understand and master the strategies that engender a truly interdependent cooperative learning environment. Simply having students work in groups, even in longstanding heterogeneous teams, is not enough. Teachers must manage their classroom so that all students are invested in their teams' achievement, have opportunities to help the team, and receive both individual and team recognition from their individual performance and not from the performance of other team members.

²⁵McKinney and Frazier (2008).

References

- Chappell, Shanan, John Nunnery, Shana Pribesh, and Jane Hager. 2011. "A Meta-Analysis of Supplemental Educational Services (SES) Provider Effects on Student Achievement." *Journal of Education for Students Placed at Risk* 16, 1: 1-23.
- Hanushek, Eric A., Paul E. Peterson, and Ludger Woessmann. 2010. "U.S. Math Performance in Global Perspective: How Well Does Each State Do at Producing High-Achieving Students?" PEPG Report No. 10-19. Harvard University, Program on Education Policy and Governance: Cambridge, MA.
- McKinney, Sueanne, and Wendy Frazier. 2008. "Embracing the Principles and Standards for School Mathematics: An Inquiry into the Pedagogical and Instructional Practices of Mathematics Teachers in High-Poverty Middle Schools." *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* 81, 5: 201-210.
- The Nation's Report Card. 2017. "2015 Mathematics and Reading Assessments." Website: https://www.nationsreportcard.gov/reading_math_2015/#mathematics/groups?grade=8.
- Rappaport, Shelley, Jean Grossman, Ivonne Garcia, Pei Zhu, Osvaldo Avila, and Kelly Granito. 2017. "Group Work Is Not Cooperative Learning: An In-Depth Look at the 2014-2015 Academic Year." Working Paper. MDRC: New York City.
- Sepanik, Susan, William Corrin, David Roy, Aracelis Gray, Felix Fernandez, Ashley Briggs, and Kathleen K. Wang. 2015. *Moving Down the Track: Changing School Practices During the Second Year of "Diplomas Now."* MDRC: New York City.
- Shapka, Jennifer D., José F. Domene, and Daniel P. Keating. 2006. "Trajectories of Career Aspirations Through Adolescence and Young Adulthood: Early Math Achievement As a Critical Filter." *Educational Research and Evaluation* 12, 4: 347-358.
- Sherman, J.A. 1982. "Mathematics As a Critical Filter: A Look at Some Residues." *Psychology of Women Quarterly* 6, 4: 428-444.
- Slavin, Robert E, Cynthia Lake, and Cynthia Groff. 2009. "Effective Programs in Middle and High School Mathematics: A Best-Evidence Synthesis." *Review of Educational Research* 79, 2: 839-911
- Tuckman, Bruce W. 1965. "Developmental Sequence in Small Groups." *Psychological Bulletin* 63, 6: 384-399.

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; Oakland, California; Washington, DC; and Los Angeles, 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 members 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.