Evaluation of MyTeachingPartner-Secondary Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Andrew J. Wayne, Mengli Song, Alex Bishop, Cheryl Graczewski, Sami Kitmitto, & Heleana Lally

American Institutes for Research

September 2023

American Institutes for Research® | AIR.ORG

AIR®

Advancing Evidence.
Improving Lives.

Contents

Exhibits

Acknowledgments

Abstract

Relying on a teacher-level randomized experiment with a sample of 87 middle and high school teachers, this study was designed to examine the implementation and impact of MyTeachingPartner-Secondary delivered by local coaches who were trained and supported by the program provider. Due to disruptions caused by the COVID-19 pandemic, implementation of the 2-year program during the first year (2021–22) was weak, and only four of six study districts continued to participate in the study in the second year. Implementation during the second year was stronger but still did not achieve fidelity based on pre-specified fidelity thresholds. Analyses of teacher survey data revealed that the program had a statistically significant positive impact on teachers' enthusiasm about teaching and a marginally significant positive impact on teachers' self-efficacy at the end of the 2-year program. Nevertheless, the study did not find any significant impact on the quality of classroom interactions or student engagement based on observations at the end of Year 2 or on students' math or English language arts achievement at the end of Year 1. (Year 2 student achievement data are not yet available.) Results from this study need to be interpreted with caution given study limitations resulting largely from the influence of the pandemic.

Introduction

There is overwhelming evidence that teachers are central to student outcomes. Longitudinal studies consistently show that teacher effectiveness varies, both within and across schools, in terms of improving students' academic achievement, social and emotional competencies, and later-life outcomes (Chamberlain, 2013; Chetty et al., 2014; Gershenson, 2016; Kraft, 2019; Rivkin et al., 2005). There is also clear evidence that differences in teacher effectiveness explain some of the inequalities seen in academic outcomes (Isenberg et al., 2016), which have become more severe with the COVID-19 pandemic (e.g., Kuhfeld et al., 2022; West & Lake, 2021).

The National Assessment of Educational Progress (NAEP) documented the extent of inequality in student achievement as of 2022, 2 years after the pandemic began. In mathematics, for example, just under two thirds (62%) of students scored at or above NAEP Basic, and the percentage was lower among students from low-income families, Black students, and Hispanic students. Specifically, among those eligible for the National School Lunch Program, 46% scored at or above NAEP Basic in mathematics, compared to 75% among those not eligible. For Black and Hispanic students, the percentages were 38 and 49, compared to 74% of White students (National Center for Education Statistics [NCES], 2022a). The pattern in reading achievement was similar (NCES, 2022b).

This evidence—about teacher effects, student outcomes, and inequality—has fueled efforts to boost the effectiveness of teachers in schools serving youth from disadvantaged backgrounds, especially through teacher professional learning. Many of these efforts predate the pandemic. School systems have invested billions in teacher professional learning every year, drawing from federal programs that allocate funds based in part on family income (U.S. Department of Education, 2014).

To help school systems use teacher professional learning as a strategy to boost teacher effectiveness and close achievement gaps, researchers have sought to identify professional learning programs that are effective and scalable (Kraft et al., 2018; Wayne & Coggshall, 2022). A recent meta-analysis pooled together studies of the impact of teacher professional learning programs and found a positive average impact on student achievement, suggesting that professional learning is a promising strategy (Garrett et al., 2021). However, the authors also found considerable variation. Examining individual studies, it is clear that some professional learning programs with almost all the features recommended by experts (e.g., that the professional learning be sustained, intensive, and content-focused and that it involve active learning) do not significantly impact achievement (Garet et al., 2016; for additional program

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

reviews see Garrett et al., 2019; Kennedy, 2016; Kraft et al., 2018). In short, we do not yet know enough about how to design teacher professional learning programs.

One professional learning program that stands out in the literature for its effectiveness is an instructional coaching program called MyTeachingPartner (MTP), which is a web-mediated coaching program developed by the Center for Advanced Study of Teaching and Learning at the University of Virginia. It is designed to help teachers improve the quality of classroom interactions as a strategy for boosting student engagement and academic achievement. To do so, MTP targets validated, measurable dimensions of classroom practice defined by the Classroom Assessment Scoring System–Secondary (CLASS), which is a classroom observation instrument designed to measure classroom interactions associated with student engagement and learning (see Appendix A for domains and dimensions measured by CLASS-S). All the MTP processes and resources are organized to support improvement in those aspects of classroom practice that are the focus of CLASS.

The present study focuses on a version of MTP called MyTeachingPartner-Secondary (MTP-S) to indicate that it draws on a version of the CLASS that is tailored to secondary school settings. That version of the CLASS is called Classroom Assessment Scoring System–Secondary (CLASS-S). (See Appendix A for domains and dimensions measured by CLASS-S.) The MTP-S program has been tested in two randomized controlled trials (RCTs), both of which found positive impacts on student achievement (Allen et al., 2011, 2015). There have also been promising results for versions of the MTP program tested in Pre-K (for a list of studies, see What Works Clearinghouse [WWC], 2018) and upper elementary grades (Clark et al., 2022).

Given the positive results, the program developer and provider have sought ways to scale MTP-S while preserving its effectiveness. For example, because it is common for secondary schools to allocate time for teachers to meet in department teams, Stuhlman et al. (2022) adapted the program to take place through a series of team meetings. Their pilot study found some positive associations suggesting that the team meetings that drew on MTP processes may have positive effects, if implemented. However, scaling the approach was difficult, as teams had many other activities already on their agendas.

The present study was designed to assess the effectiveness of a different approach to making MTP-S more scalable. In the two trials demonstrating the program's effectiveness in middle and high schools, the researchers directly hired and supervised a small number of coaches, who worked directly with the teachers. The present study instead enlisted the help of instructional coaches who were already present in the participating districts. Many teachers have access to instructional coaches in their buildings, or from the school district's central office. So, as a way to make MTP-S more scalable, the national provider of the program—Teachstone—sought to

train, monitor, and support instructional coaches who were local to each participating district. These coaches, in turn, delivered individual coaching that was intended to be the same as was provided in the earlier impact studies of MTP-S. Teachstone offers this approach as a service model, and there is some evidence about its use on a large scale in earlier grade levels (see Early et al., 2015).

The research questions (RQs) guiding this evaluation are based on an understanding of the MTP-S program and its theory of action. As detailed in the Program Overview section, which appears after this introduction, an important distinguishing feature of the program is its focus on the quality of classroom interactions—between teachers and students as well as among students. In MTP-S, the coach helps the teacher use the quality of classroom interactions as a lever for improving student engagement and achievement. Thus, the study was designed to answer the following questions about the implementation of the MTP-S program and its impact on teacher and student outcomes, when delivered over 2 years using local instructional coaches.

- RQ 1: To what extent were the key components of the MTP-S program (i.e., annual teacher trainings and coaching cycles) implemented as planned?

- RQ 2: To what extent were the key components of the strategy to scale the MTP-S program (i.e., guiding district selection of coaches, training coaches, and monitoring and supporting coaches) implemented as planned?

- RQ 3: What were the barriers to and facilitators of the implementation of the MTP-S program and the scaling strategy?

- RQ 4: To what extent did the implementation of the MTP-S program and scaling strategy vary across districts, and what contextual factors were perceived as affecting implementation?

- RQ 5: What was the impact of MTP-S on the quality of classroom interactions?

- RQ 6: What was the impact of MTP-S on student engagement and academic achievement?

- RQ 7: To what extent was the impact of MTP-S on student achievement moderated by student and teacher characteristics?

- RQ 8: To what extent was the impact of MTP-S on student achievement mediated by the quality of classroom interactions?

In addition to the above RQs about the implementation and impact of MTP-S as specified in our original evaluation plan, this study allowed us to address the following question about the

impact of MTP-S on teachers' teachers' perceptions of their work and their mental health, which may benefit from a strengths-based professional learning program such as MTP-S:[1]

- RQ 9: What was the impact of MTP-S on teachers' perceptions of their work and their mental health?

Finally, this evaluation also included a cost analysis that addressed the following question:

- RQ 10: What was the cost of the MTP-S program, and what was the cost effectiveness of the program?

The above RQs were addressed with an independent evaluation conducted by the American Institutes for Research® (AIR®) as part of an Education Innovation and Research (EIR) mid-phase project funded by the U.S. Department of Education. As the grantee of this EIR project, AIR organized its project staff into two teams: management and evaluation. The management team focused on overseeing and coordinating the work of all project partners, including the national provider of the MTP-S program. To ensure the independence of the project evaluation, the AIR evaluation team operated separately from the management team and independently carried out all evaluation activities, including creating the evaluation plan, executing random assignment, collecting implementation and outcome data, conducting data analyses, and reporting evaluation findings. The AIR evaluation team played no role in the implementation or scale-up of MTP-S except to provide the program provider with implementation data as feedback.

The remainder of this report starts with an overview of the MTP-S program and a high-level summary of the design of the independent evaluation. Next, three sections present the design details of and findings for the implementation evaluation, impact evaluation, and cost study of the program, respectively. The report concludes with a discussion of the findings and directions for future research.

Program Overview

This study evaluated a 2-year version of the MTP-S program implemented using the *Coach Support* delivery model. In this model, school districts identify local coaches to do the coaching, instead of Teachstone coaches. To implement the model, Teachstone trains, monitors, and supports the local coaches. The Coach Support delivery model may be more appealing to districts interested in MTP-S than receiving coaching directly from Teachstone. It is somewhat

---

[1] MTP-S uses a strengths-based approach to professional learning, whereby teachers identify effective interactions in their own teaching and then increase the depth, duration, and frequency of those interactions. Recent publications highlight the strengths-based approach as a promising alternative to other types of professional learning models and note its potential to increase teacher morale (Foster, 2021; Wells & Foster, 2022).

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

uncommon for districts to pay outside vendors to coach their teachers, perhaps because districts trust local coaches more than outside coaches, and perhaps because districts are eager to find ways to build local coaches' skills.

Exhibit 1 depicts the overall theory of change underlying the program and delivery model that guides the evaluation. As shown in the bottom half of the exhibit, it includes the key components of the MTP-S program and the intended teacher, classroom, and student outcomes. The top half depicts the key components of the strategy to scale the program (i.e., the Coach Support delivery model). Delivered through these supports, the MTP-S program is expected to improve teachers' knowledge of and skill at using effective classroom interactions. Although teacher knowledge and skill are not examined in this evaluation, these and other intermediate teacher outcomes (e.g., teacher self-efficacy and enthusiasm about teaching) are logical precursors to improved quality of classroom interactions—the primary teacher outcome for this evaluation—which in turn may lead to improved student engagement and ultimately improved academic achievement. Below, we provide a high-level description of the key components of the MTP-S program and the strategy to scale the program.

Exhibit 1. Theory of Change for the MTPS Program and the Strategy-to-Scale

STRATEGY TO SCALE

District Selection of MTP-S Coaches
- Recruit coaches
- Interview and select coaches

Coach Specialist Functions

- Guide district selection of coaches
- Train coaches

- Monitor and support coaches
- Ongoing quality checks
- Ongoing, needs-based training
- Individual problem solving

Coaches local to each district

MTP-S PROGRAM

Annual Teacher Training

Coaching Cycles

The coach is expected to
- Select clips and write prompts
- Conduct conference
- Write summary and action plan

The teacher is expected to
- Videotape lesson
- Write responses to clips and prompts
- Participate in videoconference
- Execute plan

Intermediate Teacher Outcomes
- Teacher knowledge of and skill at using effective classroom interactions
- Teacher self-efficacy
- Enthusiasm about teaching

Quality of Classroom Interactions
- Emotional support
- Classroom organization
- Instructional support

Student Outcomes
- Engagement
- Academic learning

The 2-year version of the MTP-S program, as depicted in the bottom half of the theory of change exhibit, includes the following two key components:

- **Annual teacher trainings.** The first year of the 2-year program starts with a 2-hour teacher orientation that introduces the CLASS-S rubric, the MTP process, and the technology involved in filming and uploading classroom videos. A local MTP-S coach leads this training. At the beginning of the second year of implementation, the coach conducts a 1-hour refresher teacher training, primarily about CLASS-S principles and domains.

- **Coaching cycles.** All other aspects of the MTP-S coaching occur in 12 to 20 cycles—six to 10 per year,[2] with each cycle focusing on a lesson video-recorded by the teacher and reviewed by the coach. These coaching cycles are designed to stimulate improvement along the dimensions of teaching measured by CLASS-S. Each cycle follows the five-step process described in Exhibit 2.

Exhibit 2. The Five-Step Process for Each MTP-S Coaching Cycle

- **Step 1.** The teacher video-records his or her classroom, capturing at least 30 minutes of instruction, and submits the video to the coach.
- **Step 2.** The coach views the video, using CLASS-S as a lens. The coach selects three 1-minute clips, each involving examples of classroom interactions and behaviors relevant to a CLASS-S dimension. For each clip, the coach writes a detailed, customized prompt for the teacher, intended to promote awareness of classroom interactions, the role the interactions played in student engagement and learning, and the role of the teacher in fostering these interactions.
- **Step 3.** The teacher views the video and the prompts and posts responses in the online journal on the MTP-S website for the coach to read and prepare before the scheduled conference.
- **Step 4.** During a 30-minute video or in-person conference, the teacher and coach discuss the edited video and the teacher's responses to the prompts. The coach guides the teacher in reflection, helping the teacher see the video through the CLASS-S lens. The teacher and coach also discuss goals for the teacher, focusing on particular dimensions of CLASS-S, and develop an action plan, to include reading more about a CLASS-S dimension, watching video examples that illustrate high-quality interactions on a CLASS-S dimension, and trying a new strategy or behavior while videotaping for the next cycle.
- **Step 5.** The coach sends the teacher a detailed written conference summary and action plan.
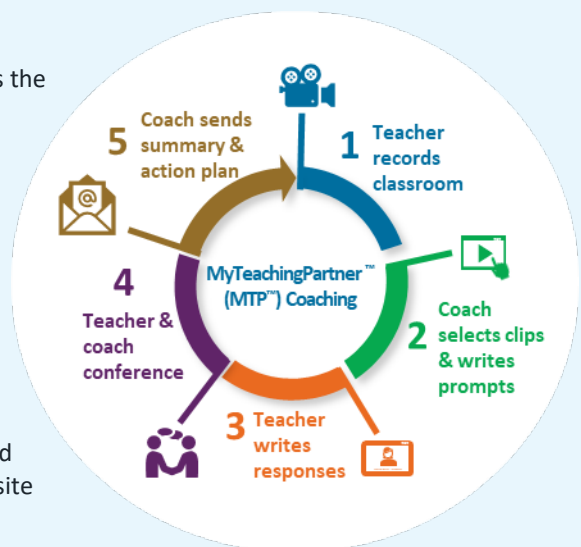
The mechanism to bring about the implementation of the components of the MTP-S program is the strategy-to-scale, shown in the top box in Exhibit 1. The strategy-to-scale is supported by coach specialists employed by Teachstone, who have experience as both MTP-S coaches and coach supervisors. Within Teachstone, the coach specialists assume a variety of responsibilities, reflected in the following three strategy-to-scale components:

1. **Guiding district coach selection.** Teachstone provides guidelines to districts to inform the selection of local MTP-S coaches. The guidelines define ideal coach qualities, such as experience teaching in a middle or high school and experience as a coach or supervisor.

---

[2] Teachstone, the partner organization responsible for the implementation of MTP-S as part of this EIR project, asked participating teachers to complete 10 coaching cycles per year and set the threshold for adequate participation for a teacher as six cycles per year.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

2. **Coach training.** Once the local coaches are selected, Teachstone conducts 5 days of training for the coaches, divided into two trainings. The first training consists of Teachstone's 2-day training for using the CLASS-S rubric to score instruction, after which the coaches demonstrate their scoring skill in an observer certification assessment. After becoming certified observers, coaches participate in the second, 3-day training that covers the MTP-S coaching model and processes.

3. **Coach monitoring and support.** Following the initial coach training, Teachstone's coach specialists prepare local coaches to lead orientations for the teachers receiving MTP-S in their district. The coach specialists then begin ongoing monitoring and support, including:

   – *Monthly group calls among coaches.* In year 1, the coach specialists lead monthly group calls with all coaches to engage them in a supportive group context to reflect on their coaching experiences. By design, no coach specialist-led group calls occur in year 2.

   – *Individual support sessions with each coach.* The coach specialists review artifacts from a recent coaching cycle against a checklist and then meets with the coach to discuss issues and provide tailored 1:1 support.

   – *Ad hoc individual problem-solving with each coach.* When a coach misses a meeting, is struggling, or has questions between meetings, the coach specialists provide individual attention and support to solve problems and provide additional training as needed via text message, email, or phone call.

Evaluation Overview

The independent evaluation of the MTP-S program implemented using the Coach Support delivery model, as described above, included an implementation evaluation, an impact evaluation, and a cost study based on a sample of secondary school mathematics and English language arts (ELA) teachers recruited for an RCT. Those teachers were randomly assigned with equal probabilities to the treatment and control conditions within blocks defined by subjects and schools. Teachers in both conditions continued to participate in "business-as-usual" professional development activities available in their districts; teachers assigned to the treatment condition were expected to participate in the MTP-S program for 2 years. Before random assignment, we asked each study teacher teaching multiple sections in Year 1 to choose a "Year 1 focal class," which would be the focus of the MTP-S coaching during Year 1 for teachers assigned to the treatment condition. At the beginning of Year 2, we asked each treatment teacher teaching multiple sections in Year 2 to choose a "Year 2 focal class," which would be the focus of the MTP-S coaching for those teachers during Year 2.

To facilitate iterative refinement of the MTP-S program, we had planned to conduct the evaluation based on three successive cohorts of teachers that would start to participate in the study in the 2018–19, 2019–20, and 2020–21 school years, respectively. However, starting in spring 2020, the COVID-19 pandemic forced the project to suspend implementation and evaluation activities that had begun for the first two cohorts, as indicated in the revised timeline in Exhibit 3. In the end, we had to exclude the first two cohorts of teachers from the evaluation and rely on Cohort 3 as the only cohort for the evaluation.[3]

To minimize the adverse effects of the pandemic on Cohort 3 recruitment, program delivery, and data collection, we decided to delay Cohort 3 by 1 year: from 2020–21 to 2021–22. Thus, the evaluation presented in this report focuses on Cohort 3 teachers, who were expected to participate in the evaluation during 2021–22 (Year 1) and 2022–23 (Year 2), as shown in the bottom row of Exhibit 3.

Exhibit 3. Revised Timeline for Implementation and Evaluation Activities, by Cohort

Cohort	2018–19				2019–20				2020–21				2021–22				2022–23			
	Summer	Fall	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter	Spring
C1		■	■	■	■	■	■	O												
C2						■	■	O	O	O	O	O								
C3														■	■	■	■	■	■	■

Note. ■ = Ongoing implementation and evaluation
O = Suspension of implementation and evaluation due to the COVID-19 pandemic

Given the emphasis of the EIR program on supporting high-need students, recruitment for this project targeted schools where at least one of the following criteria was met: (1) At least 66.7% of students were eligible for free or reduced-price lunch, or (2) at least 66.7% of students were students of color, *and* at least 50% were eligible for free or reduced-price lunch. From the schools recruited for this project, we sought teachers who did not participate in an induction program involving regular mentoring, and who taught a yearlong mathematics or ELA class during the first program year. For each teacher, at least one such class needed to have an end-

---

[3] The most intractable challenges were the suspension of state testing and in-person schooling. The spring 2020 state test data are not available for the second program year for Cohort 1 teachers (N = 27) and not available for the first program year for Cohort 2 teachers (N = 58). In addition, because MTP-S was not designed for use in online schooling, program delivery was impossible for some time periods. Cohort 2 treatment teachers participated in MTP-S for only one year (2019–20) and were unable to continue to participate in the program for a second year.

of-course examination or be composed mostly of students at grade levels that are tested by the district in mathematics and ELA in the spring.

In total, we recruited 87 Cohort 3 teachers from 16 public schools (13 middle schools and three high schools) located in six districts in five states. Of the six districts, three are urban districts, two are suburban, and one is a rural district. Of those teachers, 44 were randomly assigned to the treatment group and 43 to the control group. These teachers and their students were the basis of the implementation evaluation, impact evaluation, and cost study reported in the subsequent sections.

In addition, this EIR project served about 100 teachers not included in the implementation or impact evaluation. Some of those teachers participated in the two cohorts that began before the COVID-19 pandemic. The others began partway through the 2022–23 school year and limited their participation to five cycles of coaching over a few months. Their participation was part of an effort to expand the number of districts, schools, and teachers who had experience with the MTP-S program and elicit additional feedback for Teachstone on program design and scaling strategies.

Conditions in schools during the 2-year study time period for Cohort 3 (2021–22 and 2022–23) made it difficult to carry out RCTs focused on teacher professional learning. Schools across the United States continued to respond to the pandemic, including the Delta and Omicron variants during the 2021–22 school year. In addition, districts had severe difficulty hiring and retaining essential staff. As a result, many school-based staff were given extra responsibilities and thus had less availability for activities outside of their school responsibilities (Lieberman, 2022; Will, 2022). In this RCT and others, AIR found district and school staff less responsive than usual to efforts to keep program implementation and evaluation activities on track.

In these conditions, the level of sample attrition was greater than expected. Most notably, two of the six participating districts with six study schools dropped out at the end of Year 1. In one district, a new leader decided to drop out to give local coaches as much time as possible to fulfill unmet staffing needs that had become especially challenging. In the other, a new leader decided to discontinue participation to facilitate the launch of a district-wide coaching initiative using a different model. We discuss sample attrition in detail where it is relevant in each section.

Implementation Evaluation

In this section, we describe the implementation evaluation, which addresses the first four RQs for the overall evaluation (see Introduction). We first describe the measures used to determine

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

fidelity of implementation and what data sources were used to capture this information. Then, after defining the analytic samples, we present the evaluation findings, including whether components were implemented with fidelity in Years 1 and 2 (RQs 1 and 2) and what barriers and facilitators coaches, teachers, and coach specialists perceived as affecting implementation (RQ 3). Lastly, we discuss implementation differences across districts and the contextual factors that affected implementation (RQ 4). AIR provided summaries of findings on these topics to Teachstone at approximately 6-month intervals during implementation to inform efforts to refine program components and ensure implementation quality. In this report, the findings about implementation provide the context for interpreting the impact findings.

Measures and Data Sources

The MTP-S program has two key components—the annual teacher trainings and coaching cycles. These components were carried out by trained MTP-S coaches with teachers in each district. There are three strategy-to-scale components—guiding district coach selection, coach training, and coaching monitoring and support. These components were carried out by coach specialists employed by Teachstone, each of whom had experience as an MTP-S coach and as a coach supervisor (see the Program Overview section for additional detail). The project team used a variety of methods to collect implementation information on the program and strategy-to-scale components, including interviews, surveys, observations, attendance and system records, checklists, and monitoring forms.

For each component, we defined measurable indicators of implementation as summarized in Exhibit 4 and detailed in Appendix B. AIR worked with Teachstone to define these measures through working group sessions. The exhibit also indicates the specific data sources used to collect information for each component's indicators. For example, the first component, annual teacher trainings, has two indicators (the teacher attendance rate at training events and the extent to which planned topics were covered in the trainings). A coach survey was administered after the training to collect data on these indicators.

Exhibit 4. Implementation Evaluation Measures and Data Sources, by Component

Key Component	Indicators	Data Sources
Program Components		
1. Annual Teacher Trainings	a. Attendance rate of teachers at teacher orientation and refresher teacher training b. Extent to which topics were covered as intended in teacher orientation and refresher teacher training	a. and b. Coach survey

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Key Component	Indicators	Data Sources
2. Coaching Cycles	a. Number of MTP-S coaching cycles completed b. Fidelity of coach's products c. Teacher responsiveness to coaching d. Quality of coach–teacher relationship	a. Online system records b. Fidelity checklist applied to online system records for a randomly selected sample of coaching cycles for each coach c and d. Coach survey
Strategy-to-Scale Components		
1. Guiding District Coach Selection	a. Program provides support to districts to select high-quality coaches. (Coach has 5 or more years of teaching experience, has worked in the district for 2 or more years, and has some coaching or supervisory experience.)	a. Coach interviews
2. Coach Training	a. CLASS-S certification obtained and maintained b. MTP-S coach training attended c. MTP-S coach training topics covered as intended d. Coach self-confidence in skills	a. Online completion records b. Attendance records c. Coach Training Fidelity and Continuous Improvement Form d. Coach survey
3. Coach Monitoring and Support	a. Individual support sessions attended b. Monthly group support sessions attended c. Individual problem solving provided as needed	a. Attendance records b. Support monitoring tool c. Coach and coach specialist interviews

Analytic Samples for the Implementation Analyses

As noted in the evaluation overview, the level of attrition across the 2-year implementation period was higher than expected (e.g., two of the six districts stopped participating at the end of Year 1). For the implementation findings, the attrition in the treatment group makes it important to define the implementation analysis samples precisely.

The purpose of providing implementation findings is to inform interpretations of the impact findings. For example, if the dosage or quality is limited, one would expect the impact to be limited as well. For this reason, we define three analytic samples for presenting the implementation findings, as shown in the three columns of Exhibit 5.

- The first analytic sample focuses on Year 1 in all six districts. The reader can use findings based on this sample to aid in interpretation of the Year 1 impact findings.

- The remaining two analytic samples focus on Years 1 and 2, respectively, in the subset of four districts that participated for 2 years. Results based on these samples are necessary to interpret the Year 2 impact findings, which represent the cumulative effect of implementing MTP-S in both years.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Exhibit 5 shows the composition of the analytic samples for these three groups, accounting for districts, teachers, coaches, and coach specialists.

Exhibit 5. The Three Analytic Samples for Findings About Implementation

Participants	Year 1 (2021–22) in All Six Districts	Year 1 (2021–22) in the Four 2-Year Districts	Year 2 (2022–23) in the Four 2-Year Districts
Districts	6	4	4
Teachers	44	30	22
Coaches	16	11	5
Coach Specialists (Teachstone)	3	3	1

As shown in the bottom row, the sample for Year 1 in all six districts included 44 treatment teachers. In the four 2-year districts, there were 30 teachers present in Year 1 and 22 present in Year 2.

The implementation samples included a few teachers who were noncompliers. Their inclusion lowers the average dosage reported. These are teachers who remained in their schools as reading/ELA or mathematics teachers but chose not to participate in any coaching cycles, despite being assigned to the treatment group and told to complete 6 to 10 cycles per year. Across all six districts in Year 1, three teachers were noncompliers. In the four 2-year districts, there was one noncomplier in Year 1, and there were two noncompliers in Year 2.

There was also some teacher mobility during the study. Because these teachers became ineligible to participate, they are not counted in the implementation analyses. Across all six districts, no teachers left their school or teaching assignment area during Year 1. In the four 2-year districts, eight of the 30 teachers left between Year 1 and Year 2, either individually (two teachers) or because their school decided to discontinue participation (six teachers).

Finally, there was some mobility among the coaches. The implementation analyses for each year use data from coaches who were present and delivering coaching in that year. Coaches were replaced as needed. Thus, coach mobility does not translate into reduced coaching dosage. However, in theory, having one's coach replaced could lead to lower quality of coaching or coach–teacher relationships. In the four 2-year districts, three teachers were coached by a Teachstone coach after their local coach discontinued their participation.

Findings

Implementation Fidelity for the Program Components (RQ 1)

Both program components were carried out each year, reaching all teachers except the noncompliers. First, local coaches conducted the annual teacher trainings, including the teacher orientation in Year 1 and refresher teacher training in Year 2. Average attendance was consistently above 90%, as shown in Exhibit 6. The coaches reported fully covering about two thirds of the intended topics in Year 1 and all of them in Year 2.

For the coaching cycles, the minimum required dosage was six cycles per year. The mean number of cycles was always above six, as shown in Exhibit 6. However, the ranges show that dosage varied across teachers, from zero for the noncompliers up to the maximum of 10. In Year 1, only about two thirds of teachers completed the minimum of six cycles (59% in all six districts and 67% in the four 2-year districts). In Year 2, 91% of teachers met the minimum.

To further understand the coaching cycles, we examined three additional indicators. The first is the fidelity of coaches' products for teachers, which is based on a 23-item checklist used to review the video clips, reflection prompts, and summaries prepared by the coach. The second and third are the coaches' perceptions of teacher responsiveness and of the quality of the coach–teacher relationship. These indicators suggest a largely positive story, as shown in the bottom three rows of Exhibit 6.

Exhibit 6. Participation in the Annual Teacher Training and Coaching Cycles

Key Program Indicators	Implementation Sample		
	Year 1 (2021–22) in All Six Districts	Year 1 (2021–22) in the Four 2-Year Districts	Year 2 (2022–23) in the Four 2-Year Districts
Key Component: Annual Teacher Trainings			
Percentage of teachers attending	93%	97%	91%
Percentage of teacher training topics fully covered as planned	69%	64%	100%
Key Component: Coaching Cycles			
Mean number of cycles teachers completed	6.16 (0–10)	6.93 (0–10)	7.64 (0–10)
Mean number of items coaches achieved on 23-item fidelity checklist	21.63 (17–23)	21.40 (17–23)	21.75 (17–23)

Key Program Indicators	Implementation Sample		
	Year 1 (2021–22) in All Six Districts	Year 1 (2021–22) in the Four 2-Year Districts	Year 2 (2022–23) in the Four 2-Year Districts
Percentage of items about teacher responsiveness to coaching rated *agree* or *strongly agree*	86%	78%	100%
Percentage of items about coach–teacher relationship rated *agree* or *strongly agree*	93%	89%	100%

Despite reaching all but the noncompliers, the implementation of the two program components did not meet the component-level thresholds for adequate implementation fidelity that AIR and Teachstone had defined for the evaluation. First, to meet the threshold for the annual teacher trainings, all districts would have had to meet the threshold for attendance (100% of teachers) and the threshold for adequate topic coverage (100% of coaches reporting that all topics were covered "to a sufficient extent" and "overall orientation goals were met"). Not all districts met those thresholds, as shown in the first row of Exhibit 7. Teachers from two districts in Years 1 and 2 did not attend at all or did not attend a sufficient amount of time to meet the threshold. Additionally, not all Year 1 training in one district covered each intended topic/activity in sufficient detail. Some coaches reported that "practice with technology" was not sufficiently covered, and some disagreed that "overall orientation goals were met." Only about half of coaches met the overall threshold in Year 1, as shown in Exhibit 7, and all but two did in Year 2.

In addition, the coaching cycles component did not meet the threshold for adequate fidelity. To meet it, each coach would have had to complete the minimum number of coaching cycles (six per teacher per year) for each teacher served *and* meet thresholds for the quality of coach–teacher relationships, teacher responsiveness to coaching, and fidelity of coach's products. Only about half of coaches met the overall threshold in Year 1, as shown in Exhibit 7, and all but one did in Year 2. The criterion most often failed was the number of coaching cycles (eight of the 16 coaches failed it in Year 1, and one did in Year 2).

Exhibit 7. Scoring of Adequate Implementation of Each Key Program Component, by Year

Key Program Components	Implementation Sample		
	Year 1 (2021–22) in All Six Districts	Year 1 (2021–22) in the Four 2-Year Districts	Year 2 (2022–23) in the Four 2-Year Districts
Annual Teacher Trainings	**Fidelity Not Met** 50% (3 of 6) of districts met threshold	**Fidelity Not Met** 75% (3 of 4) of districts met threshold	**Fidelity Not Met** 50% (2 of 4) of districts met threshold
Coaching Cycles	**Fidelity Not Met** 50% (8 of 16) of coaches met threshold	**Fidelity Not Met** 55% (6 of 11) of coaches met threshold	**Fidelity Not Met** 80% (4 of 5) of coaches met threshold

Implementation Barriers for the Program Components (RQ 3)

Qualitative data were used to identify perceived barriers to program implementation as well as facilitators of successful implementation. The qualitative collections included interviews with coaches and coach specialists as well as open-ended items on the teacher survey completed by treatment teachers. Although these data do not allow us to quantify perceptions, they illustrate the types of barriers and facilitators that participants encountered.

In Year 1, a few teachers reported having had trouble using the recording devices to capture their classroom videos. Wi-Fi in schools was not always reliable, and teachers sometimes needed additional time for their classroom videos to upload so the coach could watch, select clips, and write prompts for the teacher. Teachstone provided technical support to troubleshoot issues. In Year 2, only one teacher reported that they had experienced similar issues.

Apart from issues with technology posing a barrier, factors resulting from the pandemic— including staff turnover, burnout, and limited availability—negatively affected implementation. The Delta variant was particularly disruptive at the beginning of Year 1, and the Omicron variant struck in the winter of Year 1. Some teachers and coaches described not implementing as many coaching cycles as they would have liked or feeling less engaged in the coaching process. Among them, the most commonly cited barrier was time constraints (e.g., scheduling, other professional or personal commitments).

In Year 2, in the districts that continued implementation, some teachers and coaches reported that time constraints continued to pose a challenge, but the challenges were less severe. It is possible that reduced time constraints may explain the improvement in implementation seen from Year 1 to Year 2 in the four 2-year districts, evident on measures shown in Exhibit 6 (e.g., training content coverage, the number of coaching cycles).

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Implementation Facilitators for the Program Components (RQ 3)

In addition to barriers, qualitative data from teachers and coaches illustrated some implementation facilitators. Years 1 and 2 teacher survey responses and coach interview data indicated that an important facilitator of implementation was the extent to which teachers were receptive to aspects of the coaching model.

Based on teacher survey responses, two program features stood out as making teachers particularly receptive to the program. First, teachers indicated that the MTP-S program helped foster self-awareness through intentional reflection on teacher strengths, through watching the short video clips of their own instruction. Second, teachers reported that action plans developed by their coach helped them apply the effective strategies they collaboratively identified with their coach during conferences.

In interviews, MTP-S coaches described the following factors as facilitating teacher receptiveness to MTP-S coaching and their ability to successfully coach:

1. framing MTP-S as a continuous, non-evaluative reflective process;

2. demonstrating commitment to developing and maintaining trusting relationships with teachers (see Appendix C for additional findings);

3. proactively scheduling frequent conferences with teachers and opportunities for feedback; and

4. having clear expectations and shared accountability for the coach and teacher to follow the coaching system steps as designed.

> "I valued the positive approach to my teaching that the MTP-S program focused on. It allowed me to be intentional about reflecting on my strengths as a teacher, so that I could teach with more confidence and less anxiety."   – Teacher
>
> "It was helpful for someone else to acknowledge the good that is happening within our classrooms and to be reminded that we are doing good; the simplest forms of appreciation and recognition go a long way!"   – Teacher

> **Receptiveness and engagement led to perceived positive changes in the classroom:**
>
> "I've noticed, in her [teacher] teaching practice, little things I don't see in my other teachers, like when she's answering kids' questions, even at their desk and helping them, she really makes a conscious effort to respond with questions and keep the kids doing the answering. The majority of my teachers that I don't do MTP-S with list out the steps and tell them [students] what to do."
>   – MTP-S Coach

Implementation Fidelity for the Strategy-to-Scale Components (RQ 2)

To implement the first two strategy-to-scale components, Teachstone provided support to all districts to guide the selection of local coaches. Teachstone then trained and certified the coaches. These two components reached all the districts and coaches, as shown in the first two panels of Exhibit 8. Specifically, everyone who provided coaching in Year 1 or Year 2 held the initial qualifications recommended by Teachstone (e.g., 5 or more years in teaching, previous experience coaching or supervising others) and completed the necessary training, refresher, certification, and recertification.

The implementation of the remaining strategy-to-scale component—Coach Monitoring and Support—did not consistently reach all coaches in Year 1. Some coaches participated in the support sessions at the intended rate—about monthly for individual support sessions and for group support sessions. But four coaches participated in less than half of the individual support sessions, and four participated in less than half of the group support sessions. Among them, one coach did not participate in support sessions of any kind. The mixed participation in Year 1 is reflected in the percentages in the bottom two rows of Exhibit 7, which are lower than intended.

In Year 2, as planned, Teachstone reduced the frequency of individual support sessions to every other month and did not provide group support sessions. All coaches participated in four individual support sessions. Several coaches had been replaced between Years 1 and 2, as discussed earlier.

Exhibit 8. Coaches' Participation in the Strategy-to-Scale Components

Strategy-to-Scale Indicators	Implementation Sample		
	Year 1 (2021–22) in All Six Districts	Year 1 (2021–22) in the Four 2-Year Districts	Year 2 (2022–23) in the Four 2-Year Districts
Key Component: Guiding District Coach Selection			
Percentage of coaches meeting coach qualifications criteria	100%	100%	100%
Key Component: Coach Training			
Percentage of coaches who attended CLASS-S training and achieved certification/recertification	100%	100%	100%
Percentage of coaches who attended MTP-S training/refresher	100%	100%	100%
Percentage of intended topics fully covered during coach MTP-S training/refresher	91%	100%	100%
Percentage of coaches with an average of at least *4.0–somewhat confident* on the self-confidence survey items	86%	88%	100%

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Strategy-to-Scale Indicators	Implementation Sample		
	Year 1 (2021–22) in All Six Districts	Year 1 (2021–22) in the Four 2-Year Districts	Year 2 (2022–23) in the Four 2-Year Districts
Key Component: Coach Monitoring and Support			
Percentage of coaches reporting that coach specialists provided individual problem solving as needed	100%	100%	100%
Percentage of coaches who attended at least eight individual support sessions in Year 1 and four individual support sessions in Year 2	63%	55%	100%
Percentage of coaches who attended at least six monthly group support sessions	50%	45%	n/a

The implementation of the strategy-to-scale components did not fully meet the prespecified implementation thresholds for fidelity in Year 1 but did in Year 2. Exhibit 9 provides an at-a-glance summary indicating whether each strategy-to-scale component (detailed in Exhibit 4 and Appendix B) was adequately implemented with fidelity each year. All districts across years met the threshold for the component Guiding District Coach Selection. Districts assigned qualified coaches (e.g., previous coaching/supervisory experience, 5 or more years in teaching). The other two components did not meet thresholds in Year 1: Coach Training and Coach Monitoring and Support.

The overall threshold for Coach Training was not met in Year 1 because of challenges on two of the four indicators for the component. All coaches successfully earned their CLASS-S certification and attended the MTP-S training, as described earlier. However, thresholds were not always met for the indicators: (1) MTP-S training topics covered as intended and (2) coach self-confidence. Observation data from two trainings in Year 1 indicated that a few activities were skipped or only partially covered. And coach survey results indicated that coaches from two districts felt less confident than required about their coaching skills (e.g., prompt writing, recommending appropriate exemplar videos) and their knowledge of CLASS domains (e.g., Content Understanding, Analysis and Inquiry).

As was true for Coach Training in Year 1, the overall threshold for Coach Monitoring and Support was not met in Year 1. For this component, coach specialists were asked to report on attendance by the coaches whom they were assigned to support. Coach specialists struggled with getting some coaches to attend the individual and monthly group sessions, as discussed earlier. As a result, only one of the three coach specialists active in Year 1 met the threshold for this component, which required that the coach specialist engage each coach through eight individual support sessions and six group support sessions.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Exhibit 9. Scoring of Adequate Implementation of Each Key Strategy-to-Scale Component

Key Scaling Components	Implementation Sample		
	Year 1 (2021–22) in All Six Districts	Year 1 (2021–22) in the Four 2-Year Districts	Year 2 (2022–23) in the Four 2-Year Districts
Guiding District Coach Selection	**Fidelity Met** 100% (6 of 6) of districts met threshold	**Fidelity Met** 100% (4 of 4) of districts met threshold	**Fidelity Met** 100% (4 of 4) of districts met threshold
Coach Training	**Fidelity Not Met** 88% (14 of 16) of coaches met threshold	**Fidelity Not Met** 82% (9 of 11) of coaches met threshold	**Fidelity Met** 100% (5 of 5) of coaches met threshold
Coach Monitoring and Support	**Fidelity Not Met** 33% (1 of 3) of coach specialists met threshold	**Fidelity Not Met** 33% (1 of 3) of coach specialists met threshold	**Fidelity Met** 100% (1 of 1) of coach specialists met threshold

Implementation Barriers for the Strategy-to-Scale Components (RQ 3)

Interview data collected in all six study districts in Year 1 indicated that the main barrier to implementation of the strategy-to-scale components was limited time for the coaches to participate in the supports offered by Teachstone. The time required for the coach role had been communicated early on. However, as noted earlier, principals and central office leaders sometimes assigned coaches unexpected responsibilities during the school year (e.g., they were pulled to serve as a substitute or were asked to take on additional classes). Their limited availability resulted in frequent absences for coaches in two districts—in particular, from individual support sessions with their coach specialist and monthly group sessions with other coaches participating in the study.

In Year 2, coaches and coach specialists reported no barriers or challenges related to the implementation of the strategy-to-scale components.

Implementation Facilitators for the Strategy-to-Scale Components (RQ 3)

During spring interviews in Years 1 and 2, some MTP-S coaches identified elements of the scaling supports that they valued and that motivated them to continue participating in the supports. In interviews, all MTP-S coaches who participated in individual support sessions with their Teachstone coach specialist said the sessions were positive and very helpful. The coaches found the following topics discussed in the support sessions to be the most helpful:

- strategies for developing action plans for teachers with clear next steps,

- instructional approaches aligned with CLASS-S dimensions that coaches could suggest to their teachers, and

- issues or challenges that the coach proposed for discussion.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Some coaches additionally mentioned specific features of group support sessions, such as having the coach specialist model coaching. And some coaches additionally mentioned the value of the coach specialists' availability for informal communications (e.g., text, email, quick calls outside of scheduled formal meetings) for additional support.

Implementation Variation Across Districts and Contextual Factors Affecting Implementation (RQ 4)

The implementation results presented above identify some indicators on which implementation varied. Much of this variation occurred across districts. For example, for the program components, the teacher trainings covered all topics as intended in some districts and not others. The indicators that varied the most by district are the mean number of cycles teachers completed and the mean number of individual support sessions coaches attended in each year. Exhibits 10 and 11 highlight these differences.

Exhibit 10 shows that of the districts participating in Year 1, three achieved average dosages in the range of seven to 10 cycles per teacher, and all those districts continued into Year 2, achieving similar averages that year. By contrast, the three other districts fell in the range of four to six cycles in Year 1, and only one of those districts continued to participate in Year 2.

Exhibit 10. Variation in the Number of Cycles Implemented by Each District, by Implementation Sample

District	Year 1 in All Six Districts: Average # of cycles completed per teacher	Year 1 in the Four 2-Year Districts: Average # of cycles completed per teacher	Year 2 in the Four 2-Year Districts: Average # of cycles completed per teacher
District A	10.0	10.0	10.0
District B	10.0	10.0	10.0
District C	7.0	7.0	9.0
District D	5.0	5.0	4.5
District E	4.0	n/a	n/a
District F	5.7	n/a	n/a

Note. n/a means the district did not participate in the second year of implementation.

For coaches' participation in support sessions, Exhibit 11 shows a split across districts in the percentage of coaches who attended eight or more sessions in Year 1. In three districts (A, B, F), 100% of coaches met this threshold. In the other three (C, D, E), only one third to two thirds did so.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Exhibit 11. Variation in the Percentage of Coaches Attending Individual Support Sessions, by District and Implementation Sample

District	Year 1: All Six Districts % of coaches who attended at least eight support sessions	Year 1: Four 2-Year Districts % of coaches who attended at least eight support sessions	Year 2: Four 2-Year Districts % of coaches who attended at least three support sessions
District A	100%	100%	100%
District B	100%	100%	100%
District C	67%	67%	100%
District D	38%	38%	100%
District E	0%	n/a	n/a
District F	100%	n/a	n/a

Drawing on all the implementation data collected, we hypothesize that implementation varied across study districts due primarily to three factors:

1. the extent to which the COVID-19 pandemic interfered with school operations,

2. where the coach was based (inside or outside of school), and

3. buy-in to the program and engagement of school and district personnel.

First, respondents in three districts (D, E, and F) said that regularly implementing MTP-S in Year 1 was not feasible due to interference related to the COVID-19 pandemic. As a strategy to manage the problem, they reported that they decided to attempt to implement fewer cycles in fall and more in spring. The result was that in Year 1 the coaches in these districts completed fewer cycles per teacher than the other districts, as noted above and shown in Exhibit 10. Among these three districts, only District D continued to participate in the project in Year 2.

The second factor we hypothesize led to variation in implementation was the situational demands associated with a coach being school-based, district-based, or Teachstone-based. Unlike the other districts, districts D, E, and F opted to use some school-based coaches, meaning coaches worked at the same school as their teachers. School-based coaches in these districts were sometimes teachers who taught classes during the school day and were, by definition, on hand when principals needed to fill in for absences of critical staff. By contrast, district-based and Teachstone-based coaches may have had more time to dedicate to MTP-S compared to school-based coaches, because of more predictable responsibilities and dedicated time allotted for coaching.

Finally, we suspect an important factor that led to variation in implementation by district was buy-in and engagement by leaders. This factor varied across the six districts that participated in Year 1. According to coach interview data and responses from some teachers to open-ended

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

questions on the teacher survey, leaders in Districts D, E, and F did not communicate MTP-S as a priority to teachers and coaches amid other priorities, preventing them from having the time needed to adequately implement the program.

Impact Evaluation

In this section, we provide details about the design of the MTP-S impact evaluation and report findings about program impacts on teacher and student outcomes. These findings are based on analyses addressing RQs 5–7 and RQ 9 of the overall evaluation (see Introduction for the list of RQs).

Design of the Impact Evaluation

The impact evaluation of the MTP-S program was based on a blocked cluster-RCT, in which teachers were randomly assigned to the treatment and control conditions within blocks defined by subjects and schools. For this evaluation, teachers are the appropriate unit of assignment because the "unit of treatment" of MTP-S is individual teachers rather than groups of teachers or whole schools. One commonly voiced concern about within-school, teacher-level random assignment is the risk of contamination, which would occur if some control teachers received some of the treatment. Although contamination does not affect the internal validity of intent-to-treat impact estimates, it may weaken the service contrast and lead to underestimated treatment effects. For the MTP-S evaluation, contamination was likely limited for two reasons. First, most MTP-S program materials were stored on a secure, password-protected website that could not be accessed by control teachers. Second, treatment teachers were asked not to share what they learned from the MTP-S program with other teachers in their school before the study ended.[4] Next, we describe in detail the sample, measures and data sources, and analytic methods used for the impact evaluation.

Sample

Teacher sample. The teacher sample for the impact evaluation included a total of 87 math and ELA teachers from 16 study schools in six districts. The majority (69) of those teachers taught middle school grades, and 18 teachers taught high school grades. Those teachers were randomly assigned within subjects and schools to the two study conditions at the beginning of the first program year (fall 2021) and were the basis for both Year 1 and Year 2 impact analyses of teacher outcomes. No teacher joiners were included in any analyses of teacher outcomes. Six schools had separate random assignment blocks for math and ELA teachers, and the other

---

[4] No evidence of contamination was found in prior studies of MTP-S (Allen et al., 2011; Allen et al., 2015) or in an RCT of a teacher professional development program conducted by AIR using a similar design (Garet et al., 2016).

10 schools each had a single block of math or ELA teachers. Thus, there were 22 blocks in total across the 16 study schools. Exhibit 12 presents the sample size information for the teacher sample, overall and by subject.

Exhibit 12. Teacher Sample Size for the MTP-S Impact Evaluation, Overall and by Subject and Study Condition

Teacher Sample	N of Random Assignment blocks	N of Teachers Randomized, by Study Condition		
		Treatment Group	Control Group	Total
Overall teacher sample	22	44	43	87
Math teacher sample	11	21	22	43
ELA teacher sample	11	23	21	44

Student samples. To assess the impact of the 2-year MTP-S program on student achievement, we defined the student sample for assessing Year 1 impact and the sample for assessing Year 2 impact as follows:

- *Year 1 student sample:* all students enrolled in sections of math or ELA taught by study teachers prior to random assignment in the fall of Year 1.

- *Year 2 student sample:* all students enrolled in sections of math or ELA taught by study teachers in the spring of Year 2.

While the Year 1 student sample identified prior to random assignment does not include any joiners (i.e., students who entered study teachers' classes after random assignment), the Year 2 student sample does include joiners by design. Those joiners may pose a risk of bias if their placement into study teachers' classrooms was affected school staff's knowledge about teachers' treatment status. To minimize the potential risk of bias due to joiners for the Year 2 student impact analyses, during recruitment, we explicitly required that study schools not manipulate students' classroom placement in Year 2 according to teachers' treatment status.

Measures and Data Sources

Measures of Teacher Outcome: Quality of Classroom Interactions and Survey-Based Teacher Outcomes

Our primary teacher outcome is the quality of classroom interactions as measured through classroom observations coded using CLASS-S. For each study teacher, we video-recorded one lesson in the early fall of Year 1 (as baseline) and up to three lessons in the spring of Year 2. These video-recoded lessons were coded by certified CLASS-S observers at AIR, using CLASS-S, which covers 12 dimensions of classroom practice grouped into four domains: *Emotional Support*, *Classroom Organization*, *Instructional Support*, and *Student Engagement*. For each

lesson, CLASS-S coding generated 12 dimension scores, domain scores for the first three domains,[5] and an overall score. We analyzed the Year 2 CLASS-S overall score as the primary teacher outcome measure and analyzed the three CLASS-S domain scores as supplemental teacher outcome measures.

Given concerns about potential overalignment between teacher outcomes measured with CLASS-S and the MTP-S program, certified observers also coded the video-recorded lessons using an alternative instrument—Charlotte Danielson's Framework for Teaching (FFT). FFT is a non-subject-specific classroom observation instrument that focuses on dimensions of teacher practice similar to those measured by CLASS-S and has demonstrated evidence of validity in prior research (Bill & Melinda Gates Foundation, 2012; Goe & Little, 2008). (See Appendix A for domains and dimensions measured by CLASS-S and FFT.) As a robustness check, we used the FFT overall score and scores for the two FFT domains that are amenable to classroom observation (i.e., the Classroom Environment and Instruction domains) as measures of the quality of classroom interactions. The variance decomposition of the CLASS-S and FFT measures based on the Year 2 observations as well as the inter-rater reliability estimates based on four video-recorded observations coded by all 11 coders (six CLASS-S coders and five FFT coders) are presented in Exhibit 13.[6]

Exhibit 13. Variance Decomposition and Inter-rater Reliabilities of CLASS-S and FFT Measures

Measure	Decomposition of Total Variance Within Districts		Inter-rater Reliability
	Lesson level	Teacher level	
CLASS-S Overall Score	**71%**	**29%**	**0.86**
CLASS-S Domain Scores:			
Emotional support	83%	17%	0.76
Classroom organization	62%	38%	0.69
Instructional support	79%	21%	0.85
Student engagement	89%	11%	0.52
FFT Overall Score	**84%**	**16%**	**0.82**
FFT Domain Scores:			
Classroom environment	85%	15%	0.73
Instruction	91%	9%	0.83

---

[5] The fourth CLASS-S domain—Student Engagement—includes only one dimension; therefore, a separate domain score is not generated for this domain, as it would be the same as the dimension score.

[6] For each CLASS-S or FFT measure, we estimated the percentage of the total within-district variance at the lesson and teacher levels with a multilevel model controlling only for study districts and estimated the interrater reliability using Stata's "icc" command. We did not distinguish teacher-level variance from school-level variance in the results presented in Exhibit 13, as the number of schools (seven schools in four districts) included in Year 2 observations is too small to allow reliable estimates of school-level variance.

In addition to the quality of classroom interactions measured by observations, we collected data on teachers' perceptions of their work and their mental health as secondary teacher outcomes based on a teacher survey administered in the spring of Year 2. These teacher outcomes include the following four scales: Self-Efficacy, Improvement in Teaching Practices, Enthusiasm about Teaching, and Symptoms of Depression. The study team created the Improvement in Teaching Practices scale, and the other three scales are based on existing survey instruments.[7] The specific items comprising each scale and the reliability of each scale are presented in Appendix D.

Measures of Student Outcomes: Engagement and Achievement

Student engagement. We measured student engagement at the classroom level using the CLASS-S *Student Engagement* dimension score from the classroom observations described above. This dimension captures the degree to which all students in the class were focused and participating in the learning activity presented or facilitated by the teacher.

Student achievement. For the secondary school students in our study sample, achievement data came from both state end-of-grade (EOG) tests and end-of-course (EOC) tests in math and ELA. Based on our original evaluation plan, we were to recruit Grades 6–12 teachers who taught a yearlong math or ELA class with an EOC test and use their students' EOC test scores from the spring of Year 2 as the primary measures of student achievement. However, given challenges with recruitment during the pandemic, we relaxed the teacher eligibility criteria by allowing teachers to join the study as long as they teach at least one class with either an EOC or EOG test at the end of the first program year, provided that they also meet the other eligibility criteria. Largely as a result of the relaxed teacher eligibility criteria and the fact that the large majority (79%) of the teachers in our sample were middle school teachers rather than high school teachers, very few study teachers had students with EOC test data, and most teachers had students with EOG test data based on the Year 1 administrative records obtained from the study districts.[8] Therefore, for each subject, we created a new measure based on all available test data—(EOG+EOC) score—as the primary achievement measure, which would be the EOG test scores for students with such data and would be the EOC test scores for students with EOC test scores but not EOG test scores in the given subject. Before creating the new measure, we

---

[7] The measure for Self-Efficacy is based on items from the Teacher Sense of Efficacy scale (Tschannen-Moran & Woolfolk-Hoy 2001). The items for the measure Enthusiasm about Teaching come from the State of the American Teacher survey (Doan et al., 2022), which drew the items from the teacher questionnaire in the National Teacher and Principal Survey (NCES, 2017) as well as the Seidman-Zager Teacher Burnout scale (Seidman & Zager, 1987). Doan et al. (2022) and Steiner et al. (2022) use the item responses to form a dichotomous indicator of teacher burnout. We totaled the item responses to form the scale Enthusiasm About Teaching. Finally, the Symptoms of Depression scale is the Patient Health Questionnaire-2 (PHQ-2) (Kroenke et al., 2003) and was also used in the State of the American Teacher survey.

[8] Year 2 student achievement data from spring 2023 are not yet available.

standardized the EOC and EOG test scores within each study district based on the district mean and standard deviation for the specific test.

We used the EOG test scores in math and ELA as the supplemental measures of student achievement and did not analyze EOC test scores separately due to the limited data available. In addition to subject-specific analyses of (EOG+EOC) test scores and EOG test scores, we also conducted analyses of these achievement measures with data pooled across both math and ELA study teachers. For students who were taught by both a math study teacher and an ELA study teacher and had both a math score and an ELA score in the pooled data set, we randomly selected one score for inclusion in the pooled achievement analysis so that each student contributed only one unique record to the pooled analysis of the "overall" (EOG+EOC) test scores or the "overall" EOG test scores.

For the Year 1 student achievement analyses presented in this report, the baseline measure corresponding to each achievement outcome measure is students' test scores in the same subject(s) from the prior spring (spring 2021). For students in Grades 6–9, we used their EOG test scores in math and ELA as the baseline measures for the achievement impact analyses. Since no high school students in our Year 1 achievement analysis sample had EOG test scores from the prior year, we used their EOC test scores from the most comparable course (standardized within course and district) from spring 2021 as the baseline achievement measure.

Covariates/Potential Moderators

To adjust for chance imbalance in certain sample characteristics and to improve the precision of impact estimates, we included teacher and student background characteristics as covariates in the impact analyses as appropriate. Impact analyses of teacher outcomes include the following teacher characteristics as covariates: years of teaching experience (4–10, 11–20, and 21+ years, with 1–3 years as the omitted reference), whether the teacher has a graduate degree, and baseline classroom observation score (for observation-based outcomes). Impact analyses of student outcomes include the teacher characteristics listed above as teacher-level covariates, as well as the following student characteristics as student-level covariates: prior achievement scores, gender, race/ethnicity (White vs. non-White), eligibility for free or reduced-price lunch, English learner status, and special education status. These covariates are also potential moderators in the differential impact analyses addressing RQ 7.

Data on teacher characteristics came from a survey of teachers' coaching experiences administered in the spring of each program year. Data on student background characteristics were requested together with achievement data from study districts.

Sample Sizes and Attrition

As explained in the Evaluation Overview section, substantial attrition occurred before the start of the second year of the MTP-S program. Two of the six study districts with six study schools dropped out of the project before Year 2 and did not participate in the Year 2 classroom observations or the teacher survey. Substantial teacher attrition also occurred in the other four districts. Data on Year 2 teacher outcomes are available only for teachers in the remaining four study districts (i.e., the "2-year districts") that participated in the study-conducted data collection in the spring of Year 2. Exhibit 14 presents the number of teachers with the data needed for analyses of outcomes based on the Year 2 classroom observations and teacher survey, as well as the number of teachers in the full RCT sample by study group. The last column of the exhibit indicates whether each teacher outcome had high or low attrition according to the WWC attrition standard. Unfortunately, all observation-based teacher outcomes and one of the four survey-based teacher outcomes (self-efficacy) had high attrition according to optimistic attrition boundaries specified by the WWC, due largely to the high overall attrition rates (55.2% to 58.6%) rather than the differential attrition rates (1.0% to 3.3%).

Exhibit 14. Sample Sizes at Randomization and in Analytic Samples for Year 2 Teacher Outcomes, by Study Group

Teacher Outcomes	Treatment Teachers		Control Teachers		
	# Randomized	# Analytic Sample	# Randomized	# Analytic Sample	Attrition
Year 2 Observation Measures					
CLASS-S overall score and domain scores	44	19	43	20	High
FFT overall score and domain scores	44	19	43	20	High
Year 2 Survey Measures					
Self-efficacy	44	18	43	19	High
Enthusiasm about teaching	44	19	43	19	Low
Improvement in teaching practices	44	18	43	18	Low
Symptoms of depression	44	19	43	18	Low

Note. Seven of the 16 study schools were included in the analyses of Year 2 observation measures and teacher survey measures. One of the CLASS-S domain scores (i.e., Student Engagement score) is considered a student outcome in this study even though it was based on teacher observation data.

Although we obtained Year 1 student achievement data from all six study districts, we had to exclude one district from the Year 1 student achievement analyses because the data provided by that district do not allow us to link students to their teachers.[9] Exhibit 15 presents the

---

[9] We are continuing to communicate with that district in an attempt to obtain the data needed for including the district in our Year 1 student achievement analyses.

sample size information for Year 1 student achievement outcomes within the five study districts included in Year 1 achievement analyses. For completeness, we include in the exhibit sample size information for three EOC-based achievement outcomes, even though we did not conduct separate analyses of these measures given the limited data availability.

For each achievement outcome, the exhibit presents for each study group the number of students and the number of teachers included in the impact analysis, as well as the number of teachers in the original RCT sample in the five study districts whose students were expected to have the relevant achievement data and the number of students taught by those teachers based on the fall of Year 1 class rosters. The exhibit also indicates whether each achievement outcome had high or low attrition according to WWC's attrition standards. There was no teacher-level attrition for any of the achievement measures. However, there was high student-level attrition for the ELA EOC score and math EOG score, due largely to the high differential attrition rates for these two measures.[10]

Exhibit 15. Sample Sizes at Randomization and in Analytic Samples for Year 1 Student Achievement Outcomes in Five Study Districts, by Study Group

| Year 1 Student Achievement Outcomes | Treatment Group | | | | Control Group | | | | |
| | Teachers | | Students | | Teachers | | Students | | |
	# Random-ized	# Analytic Sample	# Random-ized	# Analytic Sample	# Random-ized	# Analytic Sample	# Random-ized	# Analytic Sample	Attrition
ELA EOC score	5	5	721	627	4	4	221	162	High
ELA EOG score	16	16	1,028	967	15	15	836	793	Low
ELA (EOG+EOC) score	21	21	1,749	1,594	19	19	1,057	955	Low
Math EOC score	2	2	202	167	1	1	141	108	Low
Math EOG score	16	16	1,449	1,146	19	19	1,704	1,560	High
Math (EOG+EOC) score	18	18	1,651	1,547	20	20	1,845	1,707	Low
Overall EOC score	7	7	981	970	5	5	280	272	Low
Overall EOG score	32	30	1,669	1,669	34	34	1,965	1,965	Low
Overall (EOG+EOC) score	39	37	2,650	2,639	39	39	2,245	2,237	Low

Note. Of the 16 study schools, nine were included in the analysis of Year 1 ELA (EOG+EOC) scores, eight were included in the analysis of Year 1 math (EOG+EOC) scores, and 13 were included in the analysis of Year 1 overall (EOG+EOC) scores.

---

[10] For those two measures, the overall attrition rates were 16.2% and 14.2%, respectively, and the differential attrition rates were -13.7% and 12.5%, respectively.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Analytic Methods

In this section, we describe the analytic methods used to assess the impacts of the MTP-S program on both teacher outcomes (RQs 5 and 9) and student outcomes (RQ 6) as well as the differential impacts on student achievement by student and teacher characteristics (RQ 7). Given that Year 2 student achievement data from spring 2023 assessments are not yet available, we are not able to conduct the mediation analyses to assess the extent to which the impact of MTP-S on student achievement at the end of the 2-year program is mediated by the quality of classroom interactions (RQ 8). Therefore, in this section, we focus on the analytic methods used to estimate the main impacts and differential impacts of the MTP-S program, which are also the focus of the Findings section to follow.

Main Impact Analyses

Our main impact analyses are intent-to-treat (ITT) analyses, which estimate the impacts of being randomly assigned to receive MTP-S coaching on the quality of teacher and student outcomes. Given the limited number of random assignment blocks in the study sample, we treated blocks as fixed rather than random effects in our main impact analyses, as described below. Compared with a random-effects model, the fixed-effects model generally has greater statistical power and does not require the assumption that the blocks included in this study are representative of a larger population of blocks.

Impact on the Quality of Classroom Interactions (RQ 5). To assess the impact of MTP-S on the quality of classroom interactions (RQ 5) as measured with classroom observations conducted in spring of Year 2, we estimated the following two-level model, where observed lessons are nested within teachers. Since the CLASS-S and FFT measures of the quality of classroom interactions are not subject-specific, we conducted this set of analyses with data pooled across both math and ELA teachers.

<u>Level 1 (lesson level):</u> $\qquad Y_{ij} = \beta_{0j} + \varepsilon_{ij}$ $\qquad\qquad\qquad\qquad$ (1)

where

- Y_{ij} is an observation score (e.g., CLASS-S or FFT overall score or domain score) for lesson i taught by teacher j; and

- ε_{ij} is a random error associated with lesson i taught by teacher j.

<u>Level 2 (teacher level):</u> $\qquad \beta_{0j} = \sum_{k=1}^{8} \gamma_{00k} \text{BLOCK}_{kj} + \gamma_{01} \text{MTPS}_j + \gamma_{02} W_j + r_{0j}$ $\qquad$ (2)

where

- BLOCK_{kj}, k = 1–8, is a set of dummy indicators for the eight random assignment blocks, including in the Year 2 observation sample;

- $MTPS_j$ is a dummy indicator for treatment status (coded 1 if teacher j is in the treatment group and 0 otherwise);

- W_j is a vector of teacher characteristics (i.e., baseline observation score, years of teaching experience, and whether the teacher has a graduate degree), grand-mean centered; and

- r_{0j} is a random error associated with teacher j.

The estimate of primary interest from the Level 2 model is γ_{01}, which represents a precision-weighted overall ITT effect on teachers' Year 2 observation score across all random assignment blocks.

Impact on Survey-Based Teacher Outcomes (RQ 9). We assessed the impacts of MTP-S on survey-based measures of teachers' perceptions of their work and their mental health using a teacher-level regression specified similarly to Equation (2). Those impact analyses were conducted separately for teacher outcomes based on Year 1 and Year 2 survey data.

Impact on Student Engagement and Achievement (RQ 6). Given that we relied on the CLASS-S Student Engagement dimension score as a classroom-level measure of student engagement, we assessed the impact of MTP-S on student engagement using the same model described above for assessing impact on the quality of classroom interactions. To assess the impact of MTP-S on student achievement in the spring of Year 1, we estimated the following three-level model (students/courses/teachers) separately for math and ELA teachers as well as with data pooled across teachers of both subjects:

Level 1 (student level): $\qquad Y_{snj} = \pi_{0nj} + \pi_{1nj}X_{snj} + \varepsilon_{snj}$ $\qquad\qquad$ (3)

where

- Y_{snj} is the test score in a given subject for student s in course n taught by teacher j, standardized based on district mean and standard deviation;

- X_{snj} is a vector of student background characteristics (i.e., prior achievement, gender, race, eligibility for free or reduced-price lunch, English learner status, and special education status), grand-mean centered; and

- ε_{snj} is a random error associated with student s in course n taught by teacher j.

Level 2 (course level): $\qquad \pi_{0nj} = \beta_{00j} + r_{0nj}$ $\qquad\qquad$ (4)

$\qquad\qquad\qquad\qquad\qquad \pi_{1nj} = \beta_{10j}$ $\qquad\qquad$ (5)

Level 3 (teacher level): $\qquad \beta_{00j} = \sum_{k=1}^{K} \gamma_{000k}BLOCK_{kj} + \gamma_{001}MTPS_j + \gamma_{002}W_j + u_{00j}$ $\quad$ (6)

$\qquad\qquad\qquad\qquad\qquad \beta_{10j} = \gamma_{100}$ $\qquad\qquad$ (7)

The independent variables in Equation (6) are similar to those in Equation (2), with the exception that the vector of teacher characteristics (W_j) does not include baseline observation score. The estimate of primary interest from the above model is γ_{001}, which represents a precision-weighted overall ITT effect on student achievement, adjusted for student and teacher background characteristics.

Differential Impact Analyses (RQ 7)

In addition to the main impact analyses, we conducted differential impact analyses to examine whether the impact of MTP-S on student achievement in the spring of Year 1 was moderated by student and teacher characteristics. To estimate the differential impact by a student characteristic, we incorporated a cross-level interaction between the treatment status and the given student characteristic into the main student achievement impact model (Equations 3–7) presented above. To estimate the differential impact by a teacher characteristic, we incorporated an interaction between the treatment status and the teacher characteristic into the teacher-level equation of the main student achievement impact model.

Handling of Missing Data

We handled missing data according to the acceptable methods as specified in the current WWC (2022) standards. For analyses of both teacher outcomes and student outcomes, we imputed missing values on covariates—including baseline measures of the outcomes—using the dummy variable adjustment approach and excluded participants with missing values on the outcome measures.

Findings

In this section, we present the findings from the impact evaluation of the MTP-S program based on the data that we have collected as of summer 2023. We first present findings on the program's impacts on key teacher outcomes (RQs 5 and 9) and student outcomes (RQ 6), and then findings about differential impact on student achievement (RQ 7).

Impacts on Teacher Outcomes

Exhibit 16 presents findings about the impact of the MTP-S program on the quality of classroom interactions as measured in the spring of Year 2 based on video-recorded classroom observations coded using both CLASS-S and FFT.[11] None of those impact estimates were statistically significant. Contrary to our expectation, however, the impact estimates for all but one of the observation-based measures were in the negative direction, with effect sizes (Hedges' g) ranging from -0.38 to 0.01 for CLASS-S measures and from -0.18 to -0.12 for FFT

---

[11] Given that the student engagement outcome for this study was measured with classroom observations, we include the impact result for the CLASS-S Student Engagement dimension score in Exhibit 16 together with observation-based teacher outcomes.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

measures. These results need to be interpreted with caution for at least two reasons. As shown in Exhibit 14, less than half of the teachers in our study sample provided Year 2 observation data, and all observation-based measures had high attrition according to WWC's attrition standard. Moreover, there were substantial and statistically significant group differences in most of the baseline observation measures among teachers in the Year 2 observation impact sample, with all the baseline differences favoring the treatment group (see Appendix Exhibit E1 for detailed baseline equivalence results).[12] The high attrition rates, combined with the lack of baseline equivalence, cast doubt on the internal validity of the findings about the impact of MTP-S on the quality of classroom interactions based on observations conducted at the end of the 2-year program.

Exhibit 16. Year 2 Impact of the MTP-S Program on the Quality of Classroom Interactions

Measure of the Quality of Classroom Interactions	Treatment Group		Control Group		Treatment –Control Difference	Standard Error	Effect Size	p-value
	Mean	SD	Mean	SD				
CLASS-S Overall Score	4.24	0.80	4.44	0.90	-0.20	0.20	-0.23	0.320
CLASS-S Domain Scores:								
Emotional support	4.06	1.02	4.12	1.06	-0.06	0.24	-0.06	0.805
Classroom organization	6.12	0.88	6.11	1.05	0.01	0.23	0.01	0.954
Instructional support	3.14	0.98	3.51	0.96	-0.37	0.20	-0.38	0.078
Student engagement	4.58	1.00	4.79	1.30	-0.21	0.26	-0.18	0.424
FFT Overall Score	2.47	0.48	2.56	0.49	-0.09	0.10	-0.18	0.412
FFT Domain Scores:								
Classroom environment	2.68	0.56	2.75	0.56	-0.07	0.12	-0.12	0.600
Instruction	2.26	0.51	2.35	0.50	-0.09	0.10	-0.18	0.350

Note. Sample size = 19 teachers and 57 observations for the treatment group, and 20 teacher and 60 observations for the control group. The treatment group means are unadjusted means; the control group means were computed based on the unadjusted treatment group means and the estimated mean differences. Effect sizes were computed as Hedges' g. SD = standard deviation.

Exhibit 17 presents findings about the impacts of the MTP-S program on four outcome measures related to teachers' perceptions of their work and their mental health based on the Year 2 teacher survey. It shows that treatment teachers expressed a significantly higher level of enthusiasm about teaching on average than control teachers at the end of the 2-year program (effect size = 0.77, $p < 0.05$). The program also had a substantial positive impact on teachers' self-efficacy with an effect size of 0.70, which was marginally significant ($p < 0.10$). The program

---

[12] For each impact analysis sample, we conducted baseline equivalence analysis using a model similar to the main impact model, where a baseline measure is predicted by the treatment indicator, controlling for random assignment blocks but not other covariates. All baseline equivalence results are presented in Appendix E.

did not have a significant impact on the other two survey-based teacher outcomes examined (improvement in teaching practices and symptoms of depression) at the end of Year 2.

Program impacts on this set of survey-based teacher outcomes at the end of Year 1 tended to be much smaller in magnitude and were not statistically significant (see Appendix F for detailed Year 1 results).

Exhibit 17. Year 2 Impacts of the MTP-S Program on Survey-Based Teacher Outcomes

Teacher Survey Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	p-value
	Mean	SD	Mean	SD				
Self-efficacy	6.44	1.25	5.66	0.88	0.77	0.39	0.71	0.060
Improvement in teaching practices	2.74	0.57	2.57	0.65	0.17	0.20	0.27	0.414
Enthusiasm about teaching	2.84	0.65	2.29	0.74	0.55	0.24	0.77	0.032
Symptoms of depression	0.79	0.69	0.71	0.73	0.08	0.26	0.11	0.754

Note. Sample size = 18 to 19 teachers in each group for each measure (see Exhibit 14). The treatment group means are unadjusted means; the control group means were computed based on the unadjusted treatment group means and the estimated mean differences. Effect sizes were computed as Hedges' g. SD = standard deviation.

Impacts on Student Outcomes

Exhibit 18 presents findings about the impacts of the MTP-S program on student achievement outcomes measured at the end of the first program year based on subject-specific achievement data and achievement data pooled across both math and ELA. The impact estimates for all the achievement measures examined were relatively small, with effect sizes ranging from -0.09 to 0.05. None of those estimates were statistically significant.[13]

Exhibit 18. Year 1 Impacts of the MTP-S Program on Student Achievement Outcomes

Student Achievement Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	p-value
	Mean	SD	Mean	SD				
ELA EOG score	-0.11	0.83	-0.14	0.77	0.03	0.08	0.04	0.707
ELA (EOG+EOC) score	-0.18	0.94	-0.22	0.81	0.04	0.08	0.05	0.600
Math EOG score	-0.17	0.74	-0.16	0.79	-0.01	0.07	-0.01	0.929
Math (EOG+EOC) score	-0.14	0.73	-0.07	0.79	-0.07	0.07	-0.09	0.292

---

[13] Baseline equivalence analysis results presented in Exhibit E3 in Appendix E indicate that differences in prior achievement scores between the two study groups were all non-significant and in the "adjustable" range according to the WWC (2022) standard, with effect sizes ranging from -0.13 to 0.06.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Student Achievement Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	p-value
	Mean	SD	Mean	SD				
Overall EOG score	-0.15	0.84	-0.15	0.81	0.00	0.06	0.00	0.967
Overall (EOG+EOC) score	-0.16	0.89	-0.18	0.82	0.02	0.06	0.02	0.778

Note. See Exhibit 15 for sample size information. The treatment group means are unadjusted means; the control group means were computed based on the unadjusted treatment group means and the estimated mean differences. Effect sizes were computed as Hedges' g. SD = standard deviation.

Differential Impacts on Student Achievement

In addition to the main impacts of the MTP-S program on Year 1 student achievement outcomes, we examined whether those impacts varied significantly by student characteristics (prior achievement and demographic characteristics) or teacher characteristics (teaching experience and graduate degree). Results from these differential impact analyses are presented in Appendix G, which shows that the impacts of MTP-S on Year 1 student achievement did not vary significantly by any of the student and teacher characteristics examined.

Cost Study

Introduction

In addition to gaining an understanding of whether MTP-S was effective at improving student outcomes, the evaluation was designed to answer questions about the cost of the MTP-S program as implemented (RQ 10: What was the cost of the MTP-S program, and what was the cost-effectiveness of the program?). To estimate costs, we applied the ingredients approach (Levin et al., 2018), in which costs are estimated based on the personnel and nonpersonnel resources actually used to deliver the program. The ingredients approach is rooted in the economic concept of opportunity costs, which assumes that each resource used to implement a program cannot be concurrently used for another program. This cost study was conducted as if MTP-S was an "add-on" to existing professional development/coaching practices. In some cases, MTP-S replaced at least part of the work of existing district coaches while in others, new coaches were hired to implement MTP-S. In instances where coaches were re-assigned from other coaching responsibilities to implement the MTP-S coaching model, the net cost of MTP-S over and above the cost of "business-as-usual" coaching would likely be lower than the cost estimate produced in this study.

This cost study was conducted to help potential adopters determine if MTP-S is an affordable program for them, what the necessary resources and costs are to implement the program, and what is needed to sustain the program. Findings from this cost study are presented by program component to allow future adopters of MTP-S to understand the cost of the program for their district or school given existing coaching and resource structures in place and whether the program is to be implemented as an "add-on" or as a partial or full "replacement" program. Further methodological details of the cost study can be found in Appendix H.

We examined program resources based on the activities described in the Program Overview section, which refers to "program components" and "strategy-to-scale components." For the purpose of the cost study, we consolidate these under three headings we call program components. Exhibit 19 summarizes these program components and the program activities within each component that are intended to contribute to the outcomes of the MTP-S program. The main resource for implementing the MTP-S program is school-based personnel (coaches and teachers[14]) who carry out these activities (see Appendix Exhibit H5 for annual hours allocated toward training and coaching cycles).

Exhibit 19. MTP-S Program Components

Program Component	Description
Training	**Coach Pre-service Training:** One-time (Year 1 only), 5-day training on MTP-S and CLASS-S.
	Coach Recertification Training: Annual MTP-S recertification training (Year 2 only for this study) that takes approximately 15.5 hours based on Teachstone's estimated time to complete recertification.
	Teacher Training: 2-hour teacher orientation on MTP-S program requirements and the coaching cycle process in Year 1 and refresher teacher training in Year 2.
Coaching Cycles	**Coaching Cycle Process:** 2-week work cycles involving coaches and teachers videotaping lessons, clipping videos, writing and reviewing prompts, holding conference meetings, developing a post-meeting action plan, and carrying out the action plan.
	Additional Coaching Support: All additional time coaches or teachers spent outside of the coaching cycle process or contracted time. For coaches, this includes additional support for teachers, communications (e.g., emails), and logistics planning. For teachers, this includes exploration of MTP-S and myTeachstone resources and participating in the CLASS-S Community online discussion board.
	Monitoring and Support From Coach Specialist: Time coaches spend with coach specialists including biweekly, one-on-one calls; coaching cycle review calls; and monthly group check-in meetings with all MTP-S coaches.

---

[14] Most of the coaches participating in this study were school-based, and a few were central-office-based or consultants. For the purpose of this report, we refer to all coaches and teachers as school-based staff.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic:
Evidence From a Randomized Experiment

Program Component	Description
Program Fees	**MTP-S Program Fee:** One-time (Year 1 only) fees for the MTP-S program, including fees for certifying coaches in the MTP-S program, certifying coaches as a CLASS-S observer, and for teacher program materials.
	Coach Recertification Fee: After the first year of MTP-S programming, if coaches wish to maintain their certification, they must take an annual CLASS-S recertification assessment (in Year 2 only for this study).

Cost Study Results

Overall Cost. **The estimated, overall per-teacher cost of the 2-year MTP-S program was $13,648**. The overall cost includes all school-based staff time and program fees used to implement MTP-S. Exhibit 20 shows how the per-teacher cost was distributed across the key program components: training, coaching cycles, and program fees.[15] The cost of the coaching cycles represents the largest proportion of the overall cost (62%), with training and program fee proportions of 17% and 21%, respectively.

Exhibit 20. Overall Cost per Teacher of 2-Year MTP-S Program (in 2023 Dollars)

---
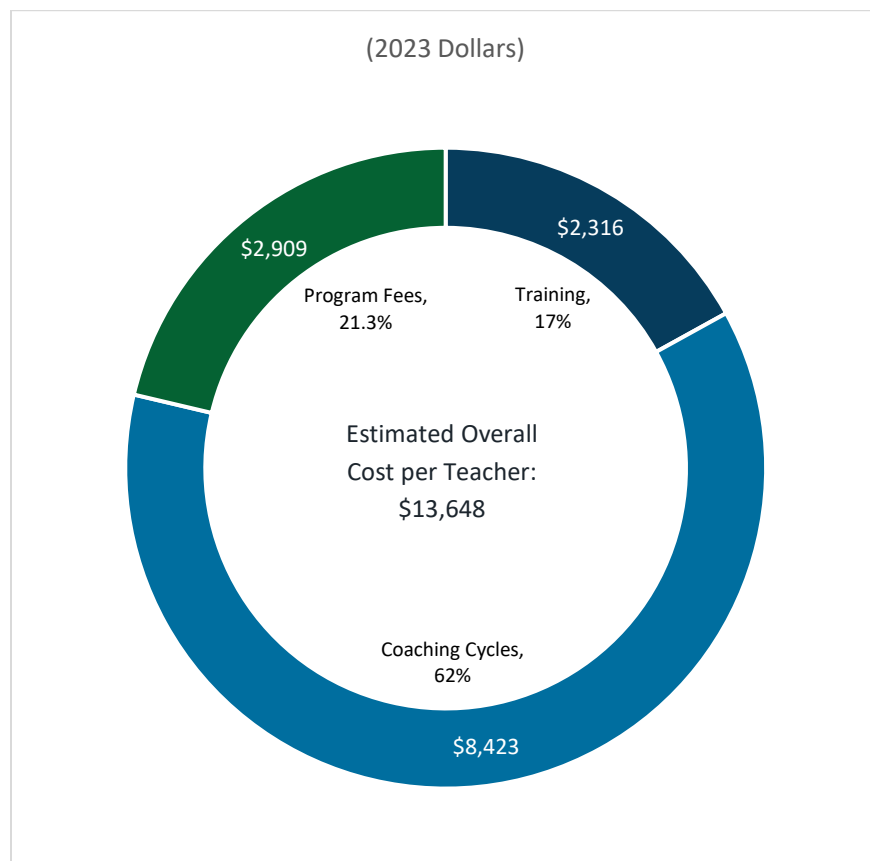
[15] The present values of all costs in Year 1 (2021–22) were calculated in 2023 dollars using a 3.5% discount rate.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Exhibit 21 shows how the overall per-teacher cost is broken down across the 2 years of the project. The decrease in costs from Year 1 to Year 2 for the "training" program component is likely due to a change in hours spent on initial training in Year 1 versus ongoing refresher training in Year 2. For example, each coach spent 66 hours on initial training of CLASS-S and MTP-S methods in Year 1 and 15.5 hours participating in recertification training in Year 2. The decrease in costs from Year 1 to Year 2 for the "coaching cycles" program component may be explained by increased familiarity with technology and material that greatly impacted the number of hours school-based staff spent on coaching cycle activities in Year 1 versus Year 2. For example, coaches reported spending double the time on "additional coaching support" activities in Year 1 (32 hours) relative to Year 2 (16 hours) because of content familiarity and technology issues. Additional information on the change in training and coaching cycle hours is provided in Appendix H.

Exhibit 21. Overall Cost per Teacher of MTP-S Program, by Year and Across 2 Years (in 2023 Dollars)

Program Component	Year 1 (2021–22) (N of teachers = 44)	Year 2 (2022–23) (N of teachers = 20)	Total
#1: Training	$1,957	$359	$2,316
#2: Coaching Cycles	$5,014	$3,410	$8,423
#3: Program Fees	$1,401	$1,508	$2,909
Total	**$8,372**	**$5,276**	**$13,648**

Because the coaching cycle component accounts for a substantial portion of the overall cost, we examined it in more detail. This component includes time spent on coaching cycles, additional coaching support, and monitoring and support from coach specialists. As shown in Exhibit 22, the cost of school-based staff time associated with the coaching cycle process is the highest proportion and is a substantial portion (40%) of overall costs.

Exhibit 22. Overall Cost per Teacher of MTP-S Coaching Cycle Program Component, Broken Out by Activities (in 2023 Dollars)

Program Activities for Program Component #2: Coaching Cycles	Total Cost per Teacher	Percentage of Total Cost per Teacher Relative to Estimated Overall Program Cost ($13,648)
Coaching Cycle Process	$ 5,427	40%
Additional Coaching Support	$ 2,141	16%
Monitoring and Support From Coach Specialist	$ 855	6%
Total	$ 8,423	61%

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Exploration of cost per student. The results above are reported in terms of per-teacher cost. However, information about per-student cost may be helpful to future adopters as that is how most school funding amounts are calculated (i.e., on a per-student basis) and program costs are reported. In this exploration of per-student cost, we assume that each teacher has a class size of 25 students and teaches five different classes each day. Thus, across the 2-year MTP-S program, a single teacher would serve an estimated 250 students.[16] To calculate the per-student cost in this study, given the data limitations, we divided the overall per-teacher cost of the program each year by the total number of students served per teacher. Exhibit 23 describes the per-student cost based on these assumptions. With an overall per-teacher program cost of $13,648 for the 2-year program, the per-student cost would be $55. Because the per-student cost information is based on assumed rather than actual class sizes, these results should be viewed as exploratory.

Exhibit 23. Cost per Student of MTP-S in Year 1 and Year 2

Program Year	Number of Students	Overall Cost	Per-Student Cost
Year 1	125	$8,372	$67
Year 2	125	$5,276	$42
Total (across both years)	250	$13,648	$55

Cost-effectiveness analysis. While this study intended to include a cost-effectiveness analysis, the study team found no statistically significant impacts on primary outcomes; thus, cost-effectiveness ratios would not be informative.

Discussion

This evaluation was designed to examine the implementation and impact of the MTP-S program when delivered by local coaches who were trained and supported by Teachstone. This approach to program delivery is potentially more scalable than direct delivery of coaching by the program provider, which was the method used in the two previous trials of MTP-S (Allen et al., 2011, 2015). So it is important to see if MTP-S can be effective when delivered by local coaches.

We report answers to all the research questions about program implementation and impact, but there are limitations on what we could learn. Most of the planned data collections were successful, but as discussed in this section, the conclusions we can reach are limited by challenges related to sample size and program implementation. Nevertheless, some of the findings on implementation and impact are important for the field.

---

[16] It is important to note that the per-student cost will vary based on the number of students served in a classroom. For example, some districts may have classroom caps at 20 students, while others allow up to 35 students in each class.

Study Limitations

The basis for the evaluation was a cohort of teachers who participated across 2 school years: 2021–22 and 2022–23 (Year 1 and Year 2). We chose this time window to avoid or at least minimize the challenges associated with the COVID-19 pandemic. One consequence of choosing this time window was that Year 2 student achievement data would not be available before project funding ended in September 2023. There was no way for the U.S. Department of Education to extend the project timeline further. So, one limitation of the study is the absence of student achievement data from the end of Year 2. That is important to note because the two previous trials of MTP-S lasted 2 years and did not find a statistically significant impact on student achievement until the second year.

The other major effect of the COVID-19 pandemic on the study design was to reduce the size of the study sample such that the study could only detect relatively large impacts. Originally, we planned to recruit 300 teachers across three cohorts and pool the cohorts for analysis. The pandemic rendered the first two cohorts unusable, leaving us with only one cohort as the basis for the evaluation. Recruitment of that cohort overlapped with the surge in COVID-19 cases associated with the Delta variant. Because of the difficult conditions for recruitment, the cohort included only 87 teachers. Then, after implementing the program and the evaluation activities in Year 1 during the Delta and Omicron variants, two of the six study districts decided to stop participating in the study. Thus, the number of teachers for estimating Year 2 impacts on teacher outcomes decreased from 87 to 39.

As a result of the sample size reductions, the evaluation was underpowered. For student achievement, which was measured only at the end of Year 1, the realized minimum detectable effect sizes (MDES) ranged from 0.20 to 0.27, based on EOG and EOC data combined.[17] Impact estimates for teacher outcomes at the end of Year 2 had realized MDES ranging from 0.58 to 1.05 standard deviations. The realized MDES were higher than the observed effect sizes for all the outcomes examined.

Findings About the Implementation of the Strategy-to-Scale and Program Components

The components of the strategy-to-scale—that is, the supports needed for local coaches to conduct the coaching—were implemented largely as intended but fell short in one important respect. The initial selection, training, and certification of local MTP-S coaches occurred as planned. In Year 1, however, some coaches did not fully participate in the ongoing coach monitoring and support events due to competing priorities and time constraints. These events

---

[17] In comparison, the target MDES for Year 2 impacts based on our original evaluation design were 0.15 for student achievement and 0.40 for measures of the quality of classroom interactions.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

were intended to help coaches successfully implement the program with quality. Missing these events may have reduced the ability of coaches to use the program to fully engage teachers in the way Teachstone intended. In Year 2, coaches attended the ongoing coach monitoring and support events at the expected rate.

Similarly, implementation data revealed gaps in what coaches were able to accomplish in Year 1. On the one hand, the products produced by the coaches (including video clips, reflection prompts, and summaries) exhibited virtually all the intended features. On the other hand, even though almost all teachers attended the orientations conducted by the coaches, the coaches were not confident that they had covered all topics they intended to address at those orientations. Perhaps most importantly, about one third of coaches and teachers did not complete the intended minimum number (six) of coaching cycles. The qualitative data from Year 1 illustrate how competing professional and personal priorities and time constraints were a factor hindering program participation for both coaches and teachers.

In Year 2, on the other hand, the coaches and teachers in the four districts that remained in the study implemented the program largely as planned. Coaches were confident that they covered all intended topics at the refresher teacher training at the beginning of Year 2. And 91% of teachers completed the minimum number of coaching cycles.

In sum, implementation did not go as planned in the six study districts in Year 1, but Year 2 implementation was stronger in the four districts that continued. As we report in the Implementation Findings section, implementation of the MTP-S program did not achieve fidelity in either year, based on the thresholds pre-established by Teachstone. However, it is conceivable that the program could have had some impact by the end of Year 2, so it was worthwhile to test for Year 2 impacts, even though the tests would be underpowered.

Findings About Program Impact

Given the limited sample size and weak program implementation in Year 1, it is not surprising that the study did not find any statistically significant impact on student achievement outcomes at the end of Year 1. Although implementation was stronger in Year 2, achievement data for that year were not available in time.

In addition to student achievement, this study assessed the impact of MTP-S on the quality of classroom interactions based on Year 2 observations and on a set of survey-based teacher outcomes at the end of each program year. However, as acknowledged in the earlier section on impact findings, impact estimates for observation-based teacher outcomes, which were all non-significant and mostly in the negative direction, have weak causal validity given high attrition combined with a lack of baseline equivalence (WWC, 2022). Below, we discuss impact findings

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

for survey-based teacher outcomes, which have clearer interpretations as most of the survey measures had low attrition according to WWC's attrition standard.

The study's teacher survey incorporated four supplemental teacher outcomes of broader interest to policy makers, practitioners, and researchers. Despite the smaller-than-intended sample size and the weaker-than-intended level of implementation, there was a statistically significant impact on one of these measures and a marginally significant impact on a second. These impacts were evident after 2 years of implementation; there were no statistically significant impacts on these measures in Year 1.

The four supplemental teacher outcomes focused on self-efficacy, perceived improvement in teaching practice, depression, and enthusiasm about teaching. The first two of these are directly relevant to potential mechanisms by which a teacher professional learning program might have a positive effect on teaching and learning.

- **Self-efficacy.** In studies of employee performance, there is evidence that repeated feedback on task performance can lead to improved self-efficacy, especially if the nature of the feedback is positive (e.g., Karl et al., 1993). Based on that general theory of action, instructional coaching programs should improve self-efficacy for teaching, especially if those programs emphasize positive feedback. Accordingly, we incorporated an established measure of self-efficacy for teaching on the spring surveys. Our findings provide some support for the hypothesis that instructional coaching affects teacher self-efficacy. After 2 years, there was a marginally significant, positive impact on self-efficacy for teaching that was large in magnitude (effect size = 0.71, $p < 0.10$).

- **Perceived improvement in teaching practice.** When a teacher professional learning program causes changes in teaching practice, it is likely that the teacher is aware that some change has taken place. To explore this hypothesis, the spring teacher surveys included questions about improvements in practice along the dimensions of classroom interactions emphasized in the MTP-S program. Although we did not find a statistically significant impact on this measure, the impact estimate was in the positive direction in both years, as we had expected.

The other two measures are more relevant to the well-being of the teacher workforce. Since the start of the pandemic, there has been increased concern about working conditions for teachers. Poor working conditions may increase teacher depression, reducing the quality of teaching that students receive. National survey data from 2021–22 show that depression symptoms were more common among teachers than those in other professions (23.1% and 18.0%, respectively; Kush et al., 2022). Moreover, difficult working conditions could lead teachers to become dissatisfied with teaching as a career and choose a different profession, further reducing the supply of teachers.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

- **Depression.** The MTP-S program gives teachers feedback that calls attention to their successes. Compared to programs that do not incorporate that feature, MTP-S seems relatively likely to reduce depressive symptoms. That said, we are not aware of any studies demonstrating that a professional learning program designed to improve instruction had an impact on teacher depression. And, contrary to our hypothesis, we did not find a statistically significant impact on teacher depression in either year.

- **Enthusiasm about teaching.** The impact of the MTP-S program on teachers' enthusiasm about teaching was not statistically significant at the end of Year 1, but was significant at the end of Year 2 with a large effect size (0.77). Many have hypothesized that teacher professional learning programs could improve enthusiasm about teaching indirectly, by giving teachers skills and confidence, but this is the first result of this kind from an impact study.

In sum, this study was conducted in challenging conditions with serious limitations, but produced some substantively important findings. As a test of the impact of MTP-S when implemented using local coaches, the study was underpowered, and implementation was weak—especially in Year 1. Unsurprisingly, there was no statistically significant impact on either student achievement outcomes at the end of Year 1 or classroom practice outcomes at the end of Year 2. Nevertheless, at least some impacts on important teacher outcomes based on survey data were detected. These impacts were positive and large—about three quarters of a standard deviation. The marginally significant impact on teachers' self-efficacy suggests that the program succeeded in boosting teachers' sense that they can successfully engage and teach students. And the statistically significant impact on teacher enthusiasm about teaching demonstrates that as policy makers and school system leaders seek ways to ensure that teachers feel energized about teaching and their teaching career, they should consider professional learning as one potential lever.

References

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*(6045), 1034–1037.

Allen, J. P., Hafen, C. A., Gregory, A. C., Mikami, A. Y., & Pianta, R. (2015). Enhancing secondary school instruction and student achievement: Replication and extension of the My Teaching Partner-Secondary intervention. *Journal of Research on Educational Effectiveness, 8*(4), 475–489.

American Institutes for Research. (2021). *Standards for the economic evaluation of educational and social programs: Cost analysis standards project*. https://www.air.org/sites/default/files/Standards-for-the-Economic-Evaluation-of-Educational-and-Social-Programs-CASP-May-2021.pdf

Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*.

Chamberlain, G. E. (2013). Predictive effects of teachers and schools on test scores, college attendance, and earnings. *Proceedings of the National Academy of Sciences, 110*(43), 17176–17182.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of Teachers II: Teacher value-added and student outcome in adulthood. *American Economic Review*, *104*, 2633–2679.

Clark, M., Max, J., James-Burdumy, S., Robles, S., McCullough, M., Burkander, P., & Malick, S. (2022). *Study of teacher coaching based on classroom videos: Impacts on student achievement and teachers' practices* (NCEE 2022-006r). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Doan, S., Greer, L., Schwartz, H. L., Steiner, E. D., & Woo, A. (2022). *State of the American Teacher and State of the American Principal Surveys: 2022 Technical documentation and survey results*. RAND. https://www.rand.org/pubs/research_reports/RRA1108-3.html

Early, D.M., Maxwell, K.L., Ponder, B.B., & Pan, Y. (2015). Improving teacher-student interactions: A randomized control trial of Making the Most of Classroom Interactions and MyTeachingPartner professional development models. *Early Childhood Research Quarterly, 38*, 57–70.

Foster, E. (2021). *Seeing teaching through a different lens: The MyTeachingPartner-Secondary coaching model*. Learning Forward. https://eric.ed.gov/?id=ED611275

Garet, M. S., Heppen, J., Walters, K., Smith, T., & Yang, R. (2016, November). *Does content-focused teacher professional development work? Findings from three Institute of Education Sciences studies*. NCEE Evaluation Brief. National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pubs/20174010/

Garrett, R., Citkowicz, M. & Williams, R. (2019). How responsive is a teacher's classroom practice to intervention? A meta-analysis of randomized field studies. *Review of Research in Education*, 43(1), 106–137. https://doi.org/10.3102/0091732X19830634

Garrett, R., Zhang, Q., Citkowicz, M., & Burr, L. (2021). *How Learning Forward's standards for professional learning are associated with teacher instruction and student achievement: A meta-analysis*. Center on Great Teachers and Leaders, American Institutes for Research.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, *11*(2), 125–149.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. National Comprehensive Center for Teacher Quality.

Isenberg, E., Max, J., Gleason, P., Johnson, M., Deutsch, J., & Hansen, M. (2016). *Do low-income students have equal access to effective teachers? Evidence from 26 districts* (NCEE 2017-4008). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Karl, K. A., O'Leary-Kelly, A. M., & Martocchio, J. J. (1993). The impact of feedback and self-efficacy on performance in training. *Journal of Organizational Behavior*, *14*(4), 379–394. http://www.jstor.org/stable/2488290

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980.

Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. *The Journal of Human Resources*, *54*(1), 1–36.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547–588.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2003). The patient health questionnaire-2: Validity of a two-item depression screener. *Medical Care, 41*(11), 1284–1292.

Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. *Educational Researcher*, *51*(7), 500–506. https://doi.org/10.3102/0013189X221109178

Kush, J. M., Badillo-Goicoechea, E., Musci, R. J., & Stuart, E. A. (2022). Teachers' mental health during the COVID-19 pandemic. *Educational Researcher, 51*(9), 593–597. https://journals.sagepub.com/doi/10.3102/0013189X221134281

Levin, H. M., McEwan, P. J., Belfield, C., Brooks Bowden, A., & Shand, R. (2018). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis*. SAGE Publications, Inc.

Lieberman, M. (2022). All teaching shortages are not equal: 4 takeaways from new research. *Education Week*. https://www.edweek.org/leadership/all-teaching-shortages-are-not-equal-4-takeaways-from-new-research/2022/12

National Center for Education Statistics (NCES). (2022a). NAEP Report Card: 2022 NAEP mathematics assessment. U.S. Department of Education. https://www.nationsreportcard.gov/highlights/mathematics/2022/

NCES. (2022b). NAEP Report Card: 2022 NAEP reading assessment. U.S. Department of Education. https://www.nationsreportcard.gov/highlights/reading/2022/

NCES. (2017). *Teacher questionnaire: National Teacher and Principal Survey, 2017–18 school year*. U.S. Department of Education.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Seidman, S. A., & Zager, J. (1987). The Teacher Burnout Scale. *Educational Research Quarterly, 11*(1).

Steiner, E. D., Doan, S., Woo, A., Gittens, A. D., Lawrence, R. A., Berdie, L., Wolfe, R. L., Greer, L., & Schwartz, H. L. (2022). *Restoring teacher and principal well-being is an essential step for rebuilding schools: Findings from the State of the American Teacher and State of the American Principal surveys.* RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1108-4.html

Stuhlman, M., Mikami, A. Y., Hofkens, T., Allen, J., Pianta, R., & Smit, S. (2022). Integrating research-supported coaching practices into secondary teachers' team meetings: Early indications of potential to impact collaborations, classroom interactions, and student engagement. *Frontiers in Education, 7*, 883226. https://www.frontiersin.org/articles/10.3389/feduc.2022.883226/full

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

U.S. Department of Education. (2014). *Fiscal year 2014 budget summary and background information*. http://www2.ed.gov/about/overview/budget/budget14/summary/14summary.pdf

Wayne, A. J., & Coggshall, J. G. (2022, January 24). How to ensure high-quality instructional coaching at scale. *Phi Delta Kappan*, *103*(5), 42–46.

Wells, K., & Foster, E. (2022, April). Strengths-based coaching supports teachers during COVID. *The Learning Professional*. https://learningforward.org/journal/coaching-for-change/strengths-based-coaching-supports-teachers-during-covid/

What Works Clearinghouse. (2018, February). *Teacher Training, Evaluation, and Compensation intervention report: MyTeachingPartner™ Pre-K*. U.S. Department of Education, Institute of Education Sciences. https://whatworks.ed.gov

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf

West, M., & Lake, R. (2021). *How much have students missed academically because of the pandemic? A review of the evidence to date*. Center on Reinventing Public Education. https://eric.ed.gov/?id=ED614130

Will, M. (2022, December 16). The teaching profession in 2022 (in charts). *Education Week*. https://www.edweek.org/teaching-learning/the-teaching-profession-in-2022-in-charts/2022/12

Appendices

Appendix A. Domains and Dimensions Measured by the Classroom Assessment and Scoring System–Secondary and the Framework for Teaching

CLASS-S	FFT[a]
Domain 1: Emotional Support	**Domain 2: Classroom Environment**
• Positive climate	• Creating an environment of respect and rapport
• Teacher sensitivity	• Establishing a culture for learning
• Regard for adolescent perspectives	• Managing classroom procedures
Domain 2: Classroom Organization	• Managing student behavior
• Behavior management	**Domain 3: Instruction**
• Productivity	• Communicating with students
• Negative climate	• Using questioning and discussion techniques
Domain 3: Instructional Support	• Engaging students in learning
• Instructional learning formats	• Using assessment in instruction
• Content understanding	
• Analysis and inquiry	
• Quality of feedback	
• Instructional dialogue	
Domain 4: Student Engagement	
• Student engagement	

[a] The full FFT instrument includes two additional domains (Domain 1. Planning and Preparation, and Domain 4. Professional Responsibilities) and two additional dimensions (Organizing Physical Space in Domain 1 and Demonstrating Flexibility and Responsiveness in Domain 3), which are not included in this study as they are not readily amenable to video-based classroom observations.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Appendix B. Scoring That Defines Adequate Implementation of Each Key Component, by Indicator

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Coach Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Implementation Component 1: Annual Teacher Trainings					
Teacher attendance rate in teacher orientation and refresher teacher trainings attended (Measured by coach survey)	Teacher	**Year 1** • 0 = Attended less than 60 minutes of the prep training or did not attend a makeup webinar • 1 = Attended more than 60 but less than 100 minutes and/or partially attended the makeup webinar • 2 = Attended more than 100 minutes or the full makeup webinar **Year 2** • 0 = Attended less than 30 minutes of the refresher or did not attend a makeup webinar • 1 = Attended more than 30 but less than 45 minutes and/or partially attended the makeup webinar • 2 = Attended more than 45 minutes or the full makeup webinar Adequate implementation = 2.	N/A	**Year 1** • 0 = < 98% teachers with a score of 2 • 1 = > 98% teachers with a score of 2 **Year 2** • 0 = < 90% teachers with a score of 2 • 1 = > 90% teachers with a score of 2	N/A
Teacher training and refresher teacher training topic coverage (Measured by coach survey)	Coach	**Years 1 and 2** • 0 = Fewer than 3 topics covered at least "mostly" as planned as reported by the coach or AIR observer • 1 = 3 to 4 topics covered at least "mostly" as planned (i.e., no significant deviations made, or coach made up for anything partially covered) • 2 = 3 to 4 topics covered "fully" as planned Adequate implementation = 2.	N/A	**Years 1 and 2** • 0 = one or more coaches with a score of 0 • 1 = all coaches with a score of 1 or more	N/A
Key Program Component 1 Total Score: Annual Teacher Trainings			N/A	Sum of district-level indicator scores (range 0–2): **Adequate = 2**	Percentage of districts with a score of 2 **Adequate sample score = 100%**

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Coach Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Implementation Component 2: Coaching Cycles					
Number of MTP-S coaching cycles completed (Measured by counts applied to online system records)	Teacher	• 0 = 5 or fewer cycles completed per year • 1 = 6–7 cycles completed per year • 2 = 8 or more cycles completed per year Adequate implementation = 1.	**Years 1 and 2** • 0 = average of < 6 cycles per coach • 1 = average of 6–7 cycles • 2 = average of ≥ 8 cycles	N/A	N/A
Fidelity of coach's products (Measured by fidelity checklist applied to online system records for a randomly selected sample of coaching cycles for each coach)	Coach	**Years 1 and 2** • 0 = Average checklist score ≤ 15 • 1 = Average checklist score 15–19 • 2 = Average checklist score > 19 Adequate implementation = 1.	N/A	N/A	N/A
Teacher responsiveness to coaching (Measured by coach survey)	Teacher	Average score across all 10 survey items	• 0 = mean score < 2.2 • 1 = mean score 2.2–3.0 • 2 = mean score > 3 Adequate implementation = 2.	N/A	N/A
Quality of coach–teacher relationship (Measured by coach survey)	Teacher	Average score across all 5 survey items	• 0 = mean score < 2 • 1 = mean score 2–3 • 2 = mean score > 3 Adequate implementation = 2.	N/A	N/A

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Coach Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Component 2 Total Score: **Coaching intensity and quality**	N/A	Sum of 4 indicator scores (range 0–8): • 0 = 0–4 • 1 = 5–6 • 2 = 7–8 Adequate implementation = 2.	N/A	Percentage of coaches with a score of 2 **Adequate sample score = > 90%**	
Key Strategy-to-Scale Component 1: Guiding District Coach Selection					
Program provides support to districts to select high-quality coaches. (Measured by coach interviews—Coach has 5 or more years of teaching experience, has worked in the district for 2 or more years, and has any amount of coaching/ supervisory experience)	District	**Years 1 and 2** • 0 = No coach in the district fully meets requirements. • 1 = At least one coach in the district does not fully meet requirements. • 2 = All coaches in the district fully meet requirements. Adequate implementation = 2.	N/A	N/A	N/A
Key Component 1 Total Score: **District selection of coaches**	N/A	N/A	N/A	Percentage of districts with a score of 2 **Adequate sample score = 100%**	

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Coach Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Strategy-to-scale Component 2—Coach Training					
CLASS-S certification obtained and maintained (Measured by online completion records)	Coach	**Year 1** • 0 = Did not pass certification assessment • 1 = Passed certification assessment **Year 2** • 0 = Did not pass recertification assessment • 1 = Passed recertification assessment Adequate implementation = 1.	N/A	N/A	N/A
MTP-S training attended (Measured by attendance records)	Coach	**Year 1 Only** • 0 = Did not attend at least 2.5 days of session or makeup session • 1 = Attended makeup session • 2 = Attended full session Adequate implementation = 1.	N/A	N/A	N/A
MTP-S coach training topics covered as intended (Measured by Coach Training Fidelity and Continuous Improvement Form)	Coach	**Year 1 Only** • 0 = Training fidelity form score is less than 15 • 1 = Training fidelity form score is 16–17 • 2 = Training fidelity form score is 18–19 Adequate implementation = 2.	N/A	N/A	N/A
Coach self-confidence in skills (Measured by coach survey)	Coach	**Years 1 and 2** • 0 = Mean score < 4 • 1 = Mean score 4–5 • 2 = Mean score > 5 Adequate implementation = 1.	N/A	N/A	N/A

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Coach Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Component 2 Total Score **Coach training intensity**	N/A	**Year 1** Sum of 4 indicator scores (range 0–5) **Year 2** Sum of indicator 1 and 4 scores (range 0–2)	N/A	**Year 1:** Percentage of coaches with a score of 5 **Adequate sample score = > 90%** **Year 2:** Percentage of coaches with a score of 2 **Adequate sample score = 100%**	
Key Strategy-to-Scale Component 3: Coach Monitoring and Support					
Individual support sessions attended (Measured by support monitoring tool with attendance records)	Coach	**Year 1** • 0 = Attended fewer than 5 sessions per year • 1 = Attended 5–8 sessions per year • 2 = Attended more than 8 sessions **Year 2** • 0 = Attended fewer than 2 sessions per year • 1 = Attended 2–3 sessions per year • 2 = Attended all 4 sessions Adequate implementation = 2.	**Year 1** Adequate implementation = 90% of coaches attend > 8 sessions **Year 2** Adequate implementation = 90% of coaches attend > 3 sessions	N/A	N/A
Monthly group support sessions attended (Measured by support monitoring tool with attendance records)	Coach	**Only Year 1** • 0 = Attended fewer than 3 sessions per year • 1 = Attended 4–6 sessions per year • 2 = Attended more than 6 sessions per year Adequate implementation = 2.	**Only Year 1** Adequate implementation = 90% of coaches attend > 6 sessions	N/A	N/A

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Coach Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Individual problem solving provided as needed (Measured by coach interviews)	Coach Specialist	**Years 1 and 2** • 0 = 75% or fewer coaches reported timely support provided when needed nearly all the time • 1 = Between 75% and 99% of coaches reported timely support provided when needed most of the time • 2 = 100% of coaches reported timely support provided when needed most of the time Adequate implementation = 2.			
Key Component 3 Total Score **Support and training for district coaches**	**Year 1** • Adequate fidelity on 3 indicators **Year 2** • Adequate fidelity on indicators 1 and 3	N/A	N/A	**Year 1:** Adequate fidelity on 3 indicators **Year 2:** Adequate fidelity on indicators 1 and 3 **Adequate sample score = 100%**	

Appendix C. Additional Implementation Findings

Teachers reported that having a motivating, likable, and knowledgeable coach was instrumental to their continued participation in the MTP-S program and their professional growth. Exhibit C.1 displays how teachers perceived their MTP-S coaches in Year 1. These perceptions of coaches created conditions that increased the likelihood of teacher receptiveness.

Exhibit C.1. Extent to Which Treatment Teachers Agreed With Specific Statements About Their Coaches
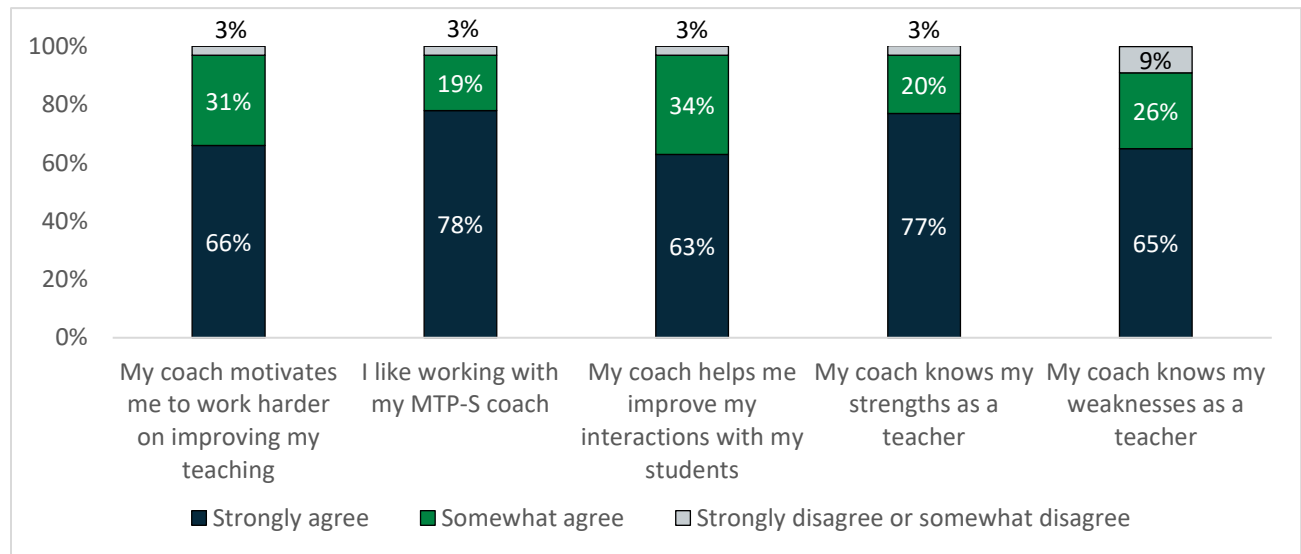
Source: Year 1 teacher survey.

Note. *N* = 32 treatment teachers. The relevant survey question was not included in the Year 2 teacher survey.

Teachers found one-on-one conferences with their MTP-S coaches provided helpful feedback that increased their confidence in teaching and focused their attention on classroom interactions. As shown in Exhibit C.2, the vast majority of teachers perceived the feedback provided by their MTP-S coach as relevant, specific, and informative about their instruction and classroom interactions.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Exhibit C.2. Teacher Perceptions of Feedback Provided by Their Coach

The feedback provided by my MTP-S coach…

	was relevant to my instruction in all of my reading/ELA or mathematics classes	provided specific ideas about how I could improve my performance	changed how I feel about my instruction	changed how I think about my interactions with students
Strongly disagree or somewhat disagree	3%	12%	6%	6%
Somewhat agree	28%	35%	38%	38%
Strongly agree	69%	53%	56%	56%

■ Strongly agree ■ Somewhat agree ☐ Strongly disagree or somewhat disagree
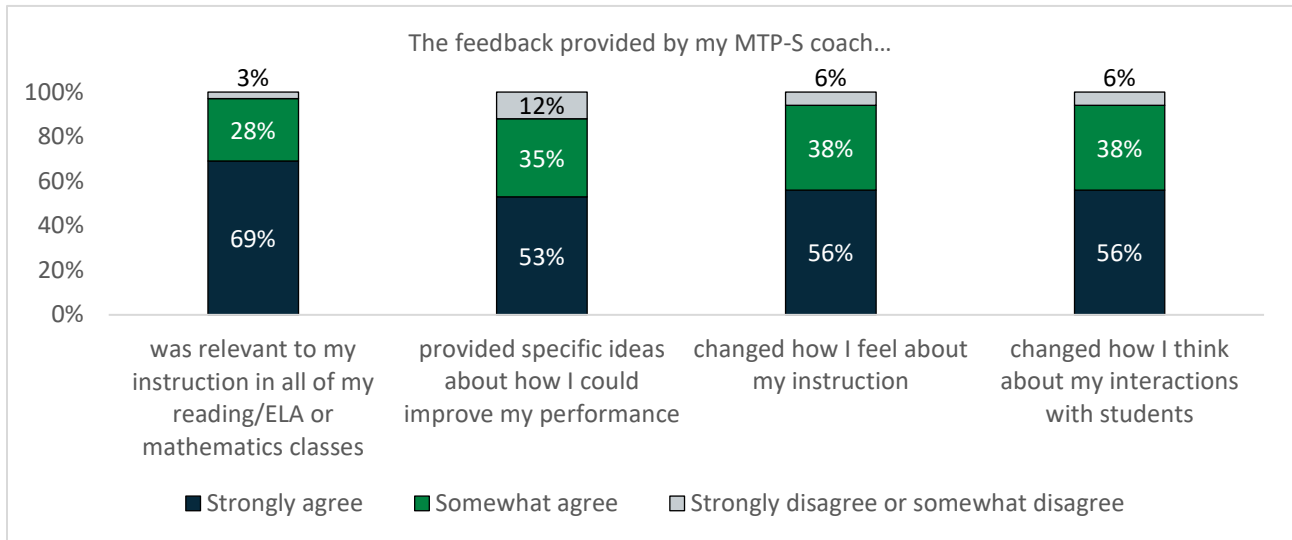
Source: Year 1 teacher survey.

Note. *N* = 32 treatment teachers. This survey question was not included in the Year 2 teacher survey.

Evaluation of MTP-S Delivered Using Local Coaches During the COVID-19 Pandemic: Evidence From a Randomized Experiment

Appendix D. Item Composition and Reliabilities of Survey-Based Teacher Outcomes

Scales	Survey Items	Reliability
Self-efficacy	Please indicate your opinions about each of the following statements . How much can you do to... • Get through to the most difficult students • Promote student learning when there is a lack of support at home • Keep students on task on difficult assignments • Remind students of what they have been taught in previous lessons • Motivate students who show low interest in schoolwork • Control disruptive behavior in the classroom (Response options: 1 = Nothing; 2, 3 = Very little; 4, 5 = Some; 6, 7 = Quite a bit; 8, 9 = A great deal)	α = 0.85
Improvement in Teacher Practices	Thinking back to the beginning of the 2022–23 school year, please indicate the extent to which you feel you were able to improve in each of the following areas over the course of this school year. • Classroom organization (e.g., routines for distributing materials, behavior management, instructional formats such as group work) • Emotional support for students (e.g., how to establish a culture for learning, how to demonstrate sensitivity to student concerns) • Instructional support (e.g., how to use questioning and discussion techniques, how to assess and support student understanding) • Student engagement (e.g., how to increase classroom participation) (Response options: 1 = I did not improve; 2 = I improved a little; 3 = I improved a moderate amount; 4 = I improved a large amount)	α = 0.72
Enthusiasm about teaching	Please indicate your level of agreement with the following statements about teaching: • The stress and disappointments involved in teaching aren't really worth it. (reverse coded) • I don't seem to have as much enthusiasm now as I did when I began teaching. (reverse coded) • I look forward to teaching in the future. • I am glad I selected teaching as a career. (Response options: 1 = Strongly disagree; 2 = Somewhat disagree; 3 = Somewhat agree; 4 = Strongly agree)	α = 0.84
Symptoms of Depression	Over the last 2 weeks, how often have you been bothered by the following problems? • Little interest or pleasure in doing things • Feeling down, depressed, or hopeless (Response options: 0 = Not at all; 1 = Several days; 2 = More than half of the days; 3 = Nearly every day)	α = 0.89

Appendix E. Supporting Exhibits for Baseline Equivalence Analyses

Exhibit E1. Baseline Observation Scores of Teachers Included in the Analyses of Year 2 Impact on the Quality of Classroom Interactions, by Study Group

Baseline Measure	Treatment Group Mean	SD	Control Group Mean	SD	Treatment – Control Difference	Standard Error	Effect Size	p-value
CLASS-S Overall Score	4.91	0.75	4.23	0.76	0.68	0.25	0.88	0.010
CLASS-S Domain Scores:								
Emotional support	4.81	0.93	4.29	1.24	0.52	0.32	0.46	0.116
Classroom organization	5.96	0.80	5.49	0.90	0.47	0.26	0.54	0.087
Instructional support	4.19	1.12	3.27	0.91	0.92	0.32	0.89	0.008
Student engagement	5.63	0.96	4.98	1.15	0.65	0.31	0.60	0.049
FFT Overall Score	2.81	0.43	2.47	0.53	0.34	0.14	0.69	0.023
FFT Domain Scores:								
Classroom environment	2.97	0.43	2.62	0.63	0.35	0.14	0.63	0.021
Instruction	2.64	0.51	2.30	0.61	0.34	0.18	0.59	0.080

Note. Sample size = 19 teachers and 19 observations for the treatment group; 20 teachers and 20 observations for the control group. The treatment group means are unadjusted means; the control group means were computed based on the unadjusted treatment group means and the estimated mean differences. Effect sizes were computed as Hedges' g. SD = standard deviation.

Exhibit E2. Background Characteristics of Teachers Included in the Analyses of Year 2 Impacts on Survey-Based Teacher Outcomes, by Study Group

Teacher Characteristics	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	p-value
	N of Teachers	Mean	N of Teachers	Mean				
Outcome: Symptoms of Depression								
Teaching 4–10 years	19	42.1%	18	39.0%	3.1%	0.67	0.08	0.849
Teaching 11–20 years	19	31.6%	18	17.8%	13.8%	0.84	0.46	0.364
Teaching 21+ years	19	15.8%	18	32.3%	-16.5%	0.85	-0.57	0.270
Graduate degree	19	47.4%	18	66.3%	-18.9%	0.68	-0.47	0.253
Outcome: Enthusiasm about teaching								
Teaching 4–10 years	19	42.1%	19	36.9%	5.2%	0.67	0.13	0.744
Teaching 11–20 years	19	31.6%	19	22.7%	8.9%	0.78	0.27	0.562
Teaching 21+ years	19	15.8%	19	30.9%	-15.1%	0.84	-0.53	0.304
Graduate degree	19	47.4%	19	68.1%	-20.7%	0.68	-0.52	0.202
Outcome: Improvement in Teaching Practices								
Teaching 4–10 years	18	44.4%	18	39.5%	5.0%	0.68	0.12	0.763
Teaching 11–20 years	18	27.8%	18	17.7%	10.0%	0.84	0.35	0.492
Teaching 21+ years	18	16.7%	18	32.4%	-15.8%	0.84	-0.53	0.299
Graduate degree	18	44.4%	18	66.0%	-21.5%	0.69	-0.54	0.202
Outcome: Self-efficacy								
Teaching 4–10 years	18	44.4%	19	36.9%	7.5%	0.67	0.19	0.643
Teaching 11–20 years	18	33.3%	19	22.5%	10.8%	0.79	0.33	0.492
Teaching 21+ years	18	11.1%	19	32.8%	-21.7%	0.96	-0.83	0.155
Graduate degree	18	50.0%	19	68.1%	-18.1%	0.68	-0.46	0.268

Note. The treatment group means are unadjusted means; the control group means were computed based on the unadjusted treatment group means and the estimated mean differences. Effect sizes were computed as the Cox Index.

Exhibit E3. Prior Achievement Scores for Year 1 Student Achievement Outcomes, by Study Group

Prior (Spring 2021) Achievement Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	p-value
	Mean	SD	Mean	SD				
ELA EOG score	-0.14	0.96	-0.26	0.93	-0.12	0.190	-0.13	0.514
ELA (EOC+EOG) score	-0.21	1.01	-0.16	0.97	0.05	0.185	0.05	0.801
Math EOG score	-0.16	0.92	-0.28	0.93	-0.12	0.108	-0.13	0.265
Math (EOC+EOG) score	-0.09	0.92	-0.07	0.94	0.03	0.112	0.03	0.810
Overall EOG score	-0.13	0.97	-0.22	0.96	-0.09	0.10	-0.10	0.355
Overall (EOC+EOG) score	-0.13	0.99	-0.07	0.98	0.06	0.11	0.06	0.581

Note. See Exhibit E4 for sample size information. The treatment group means are unadjusted means; the control group means were computed based on the unadjusted treatment group means and the estimated mean differences. Effect sizes were computed as Hedges' g. SD = standard deviation.

Exhibit E4. Sample Sizes for Baseline Equivalence Analyses for Year 1 Student Achievement Outcomes, by Study Group

Prior (Spring 2021) Achievement Measure	Treatment Group		Control Group	
	N of Teachers	N of Students	N of Teachers	N of Students
ELA EOG score	13	772	10	630
ELA (EOC+EOG) score	17	1,283	12	721
Math EOG score	11	927	7	1,214
Math (EOC+EOG) score	13	1,252	7	1,313
Overall EOG score	25	1,346	28	1,535
Overall (EOC+EOG) score	30	2,152	31	1,699

Exhibit E5. Additional Information for Year 1 Student Achievement Outcome and Baseline Measure With Imputed or Missing Data in the Analytic Sample—ELA EOG Score

	Treatment Group			Control Group		
Sample	# Students	Mean of Baseline Measure	Mean of Outcome Measure	# Students	Mean of Baseline Measure	Mean of Outcome Measure
Analytic sample	967	0.08	-0.11	793	0.21	-0.10
Subsample of students with non-missing values for outcome and baseline measure	772	-0.13	-0.07	630	0.02	-0.08
Subsample of students with non-missing outcome measure and missing baseline measure	195	Not applicable	-0.28	163	Not applicable	-0.15
Subsample of students with non-missing baseline measures and missing outcome measure	0	Not applicable	Not applicable	0	Not applicable	Not applicable

Note. Correlation between the baseline and outcome measures (calculated using only non-imputed data): 0.70

Exhibit E6. Additional Information for Year 1 Student Achievement Outcome and Baseline Measure With Imputed or Missing Data in the Analytic Sample—ELA (EOC+EOG) Score

	Treatment Group			Control Group		
Sample	# Students	Mean of Baseline Measure	Mean of Outcome Measure	# Students	Mean of Baseline Measure	Mean of Outcome Measure
Analytic sample	1,594	0.03	-0.18	955	0.12	-0.16
Subsample of students with non-missing values for outcome and baseline measure	1,283	-0.20	-0.16	721	-0.07	-0.10
Subsample of students with non-missing outcome measure and missing baseline measure	311	Not applicable	-0.41	234	Not applicable	-0.29
Subsample of students with non-missing baseline measures and missing outcome measure	0	Not applicable	Not applicable	0	Not applicable	Not applicable

Note. Correlation between the baseline and outcome measures (calculated using only non-imputed data): 0.49

Exhibit E7. Additional Information for Year 1 Student Achievement Outcome and Baseline Measure With Imputed or Missing Data in the Analytic Sample—Math EOG Score

Sample	Treatment Group			Control Group		
	# Students	Mean of Baseline Measure	Mean of Outcome Measure	# Students	Mean of Baseline Measure	Mean of Outcome Measure
Analytic sample	1,146	0.03	-0.17	1,560	0.01	-0.16
Subsample of students with non-missing values for outcome and baseline measure	927	-0.16	-0.14	1,214	-0.18	-0.13
Subsample of students with non-missing outcome measure and missing baseline measure	219	Not applicable	-0.30	346	Not applicable	-0.27
Subsample of students with non-missing baseline measures and missing outcome measure	0	Not applicable	Not applicable	0	Not applicable	Not applicable

Note. Correlation between the baseline and outcome measures (calculated using only non-imputed data): 0.57

Exhibit E8. Additional Information for Year 1 Student Achievement Outcome and Baseline Measure With Imputed or Missing Data in the Analytic Sample—Math (EOC+EOG) Score

Sample	Treatment Group			Control Group		
	# Students	Mean of Baseline Measure	Mean of Outcome Measure	# Students	Mean of Baseline Measure	Mean of Outcome Measure
Analytic sample	1,547	0.08	-0.14	1,707	-0.01	-0.16
Subsample of students with non-missing values for outcome and baseline measure	1,252	-0.09	-0.11	1,313	-0.20	-0.12
Subsample of students with non-missing outcome measure and missing baseline measure	295	Not applicable	-0.29	394	Not applicable	-0.29
Subsample of students with non-missing baseline measures and missing outcome measure	0	Not applicable	Not applicable	0	Not applicable	Not applicable

Note. Correlation between the baseline and outcome measures (calculated using only non-imputed data): 0.54

Exhibit E9. Additional Information for Year 1 Student Achievement Outcome and Baseline Measure With Imputed or Missing Data in the Analytic Sample—Overall EOG Score

Sample	Treatment Group			Control Group		
	# Students	Mean of Baseline Measure	Mean of Outcome Measure	# Students	Mean of Baseline Measure	Mean of Outcome Measure
Analytic sample	1,669	0.07	-0.15	1965	0.11	-0.12
Subsample of students with non-missing values for outcome and baseline measure	1,346	0.08	-0.10	1535	0.14	-0.09
Subsample of students with non-missing outcome measure and missing baseline measure	323	Not applicable	-0.34	430	Not applicable	-0.24
Subsample of students with non-missing baseline measures and missing outcome measure	0	Not applicable	Not applicable	0	Not applicable	Not applicable

Note. Correlation between the baseline and outcome measures (calculated using only non-imputed data): 0.63

Exhibit E10. Additional Information for Year 1 Student Achievement Outcome and Baseline Measure With Imputed or Missing Data in the Analytic Sample—Overall (EOC+EOG) Score

Sample	Treatment Group			Control Group		
	# Students	Mean of Baseline Measure	Mean of Outcome Measure	# Students	Mean of Baseline Measure	Mean of Outcome Measure
Analytic sample	2,639	0.07	-0.16	2237	0.07	-0.13
Subsample of students with non-missing values for outcome and baseline measure	2,152	0.08	-0.12	1699	0.10	-0.10
Subsample of students with non-missing outcome measure and missing baseline measure	487	Not applicable	-0.34	538	Not applicable	-0.23
Subsample of students with non-missing baseline measures and missing outcome measure	0	Not applicable	Not applicable	0	Not applicable	Not applicable

Note. Correlation between the baseline and outcome measures (calculated using only non-imputed data): 0.65

Appendix F. Supporting Exhibits for Year 1 Impacts on Survey-Based Teacher Outcomes

Exhibit F1. Year 1 Impacts of the MTP-S Program on Survey-Based Teacher Outcomes

Outcome Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	p-value
	Mean	SD	Mean	SD				
Self-efficacy	5.55	2.02	5.37	2.14	0.18	0.52	0.09	0.737
Improvement in teaching practices	2.74	0.50	2.67	0.64	0.07	0.14	0.13	0.605
Enthusiasm about teaching	2.77	0.43	2.74	0.34	0.03	0.16	0.07	0.87
Symptoms of depression	0.93	0.99	1.03	0.95	-0.1	0.22	-0.10	0.655

Note. See Exhibit F2 for sample size information. The treatment group means are unadjusted means; the control group means were computed based on the unadjusted treatment group means and the estimated mean differences. Effect sizes were computed as Hedges' g. SD = standard deviation.

Exhibit F2. Sample Sizes at Randomization and in Analytic Samples for Teacher Outcomes Based on Year 1 Teacher Survey, by Study Group

Teacher Outcomes	Treatment Teachers		Control Teachers	
	# Randomized	# Analytic Sample	# Randomized	# Analytic Sample
Self-efficacy	44	42	43	39
Improvement in teaching practices	44	42	43	40
Enthusiasm about teaching	44	42	43	40
Symptoms of depression	44	42	43	40

Appendix G. Supporting Exhibit for Differential Impacts Analyses

Exhibit G1. Differential Impacts of MTP-S on Year 1 Student Achievement Outcomes, by Student and Teacher Characteristics

Student/Teacher Characteristics	Impact Estimate for X = 0	Impact Estimate for X = 1	Difference in Impact	Standard Error	p-value
Outcome: ELA EOG Score					
Prior achievement	0.03	0.04	0.02	0.080	0.654
Gender	0.03	0.09	0.06	0.080	0.261
Eligibility for free- or reduced-price lunch	0.03	0.04	0.01	0.080	0.819
English learner status	0.03	0.12	0.10	0.079	0.259
Special education status	0.04	-0.10	-0.14	0.079	0.098
Race	0.03	0.03	0.00	0.080	0.956
Teaching experience	0.03	-0.02	-0.05	0.080	0.811
Graduate degree	0.05	0.34	0.28	0.081	0.170
Outcome: ELA (EOG+EOC) Score					
Prior achievement	0.06	0.07	0.02	0.085	0.614
Gender	0.06	0.10	0.04	0.085	0.437
Eligibility for free- or reduced-price lunch	0.06	0.05	0.00	0.085	0.941
English learner status	0.06	0.07	0.01	0.085	0.889
Special education status	0.06	-0.01	-0.08	0.085	0.360
Race	0.06	0.05	0.00	0.085	0.949
Teaching experience	0.06	0.09	0.03	0.087	0.895
Graduate degree	0.06	0.19	0.13	0.085	0.508
Outcome: Math EOG Score					
Prior achievement	-0.01	-0.03	-0.03	0.073	0.417
Gender	-0.01	-0.02	-0.01	0.074	0.839
Eligibility for free- or reduced-price lunch	-0.01	0.01	0.02	0.074	0.756
English learner status	-0.01	-0.10	-0.09	0.073	0.221
Special education status	-0.01	-0.07	-0.06	0.074	0.439
Race	-0.01	-0.02	-0.02	0.074	0.756
Teaching experience	-0.01	0.19	0.19	0.073	0.176
Graduate degree	0.01	0.15	0.13	0.078	0.467

Student/Teacher Characteristics	Impact Estimate for X = 0	Impact Estimate for X = 1	Difference in Impact	Standard Error	p-value
Outcome: Math (EOG+EOC) Score					
Prior achievement	0.00	-0.02	-0.02	0.068	0.580
Gender	0.00	0.00	0.00	0.068	0.945
Eligibility for free- or reduced-price lunch	0.00	0.05	0.05	0.068	0.294
English learner status	-0.01	-0.07	-0.06	0.068	0.325
Special education status	0.00	-0.07	-0.07	0.068	0.356
Race	0.00	0.00	0.00	0.068	0.932
Teaching experience	0.00	0.15	0.15	0.068	0.261
Graduate degree	0.01	0.20	0.19	0.069	0.228
Outcome: Overall EOG Score					
Prior achievement	0.00	0.01	0.01	0.056	0.635
Gender	0.00	0.03	0.03	0.056	0.388
Eligibility for free- or reduced-price lunch	0.00	0.05	0.05	0.055	0.328
English learner status	0.00	-0.11	-0.11	0.056	0.096
Special education status	0.00	-0.10	-0.10	0.055	0.157
Race	0.00	-0.03	-0.02	0.056	0.584
Teaching experience	0.00	0.13	0.13	0.055	0.262
Graduate degree	0.01	0.17	0.16	0.057	0.236
Outcome: Overall (EOG+EOC) Score					
Prior achievement	0.01	0.03	0.01	0.056	0.620
Gender	0.02	0.07	0.05	0.056	0.151
Eligibility for free- or reduced-price lunch	0.02	0.08	0.07	0.055	0.122
English learner status	0.02	-0.07	-0.08	0.056	0.114
Special education status	0.02	-0.08	-0.09	0.055	0.142
Race	0.02	0.02	0.00	0.056	0.978
Teaching experience	0.02	0.12	0.10	0.056	0.413
Graduate degree	0.02	0.20	0.18	0.055	0.134

Note. See Exhibit 15 for sample size information. For analyses of differential impact by students' prior achievement score, X = 1 for students with prior achievement scores that were 1 standard deviation above the district mean, and X = 0 for students with prior achievement scores at district mean. For analyses of differential impact by the other student characteristics, X = 1 for female students, students eligible for free- or reduced-price lunch, English learners, special education students, non-White students, teachers with no more than 3 years of teaching experience, and teachers without a graduate degree; X = 0 otherwise.

Appendix H. Cost Study Methodology

In this appendix we describe the methodology and limitations of the cost study. As mentioned in the Cost Study section, the analysis was designed to answer RQ 10 (What was the cost of the MTP-S program, and what was the cost-effectiveness of the program?). The cost study illustrates the cost of implementing MTP-S in terms of per-teacher cost and provides insight on the affordability, feasibility, and sustainability of implementing MTP-S based on costs of programming.

MTP-S Resources and Program Components

The MTP-S theory of change was used to identify the key program components that combine to produce the impact observed in this study. In this section, we describe the economic evaluation approach the study team used for the cost study, as well as the MTP-S resources and program components used to implement the program.

In this cost study, we applied the ingredients approach to measure program costs (Levin et al., 2018). In general, the ingredients approach involves three key steps: (1) identify and specify ingredients (i.e., program resources), (2) identify appropriate monetary values for each ingredient (i.e., prices), and (3) estimate overall cost in such a way that the cost results relate to the theory of change for the program (see Exhibit 1).

The ingredients approach is rooted in the economic concept of opportunity costs, in that each resource used to implement a program (like MTP-S) cannot be concurrently used for another program. Additionally, each input used in program implementation has a value that must be considered when estimating the true economic cost of a program. With this comprehensive approach, cost estimates include the value of all resources used by the program, even those that are shared across multiple programs and program components, like teacher time. Our data collection efforts were informed by the MTP-S theory of change (see Exhibit 1) and focused on all the activities and corresponding resources (i.e., non-personnel items) involved in implementing the MTP-S program.[18] By using the ingredients approach to delineate all the resources used across the various MTP-S components, we were able to capture a comprehensive view of the program costs and ensure that resources were not double counted (preventing overestimates of the costs).

School-based Personnel. A key resource for the MTP-S program is school-based personnel: coaches and teachers.[19] As described previously in the report, coaches and teachers carry out the main program component of MTP-S, the coaching cycles. This includes teachers videotaping

---

[18] Note that this particular implementation was disrupted by the COVID-19 pandemic. Details and potential implications of these disruptions are described in the Limitations section.

[19] Most of the coaches in this program were school-based, however, unique to this implementation, some were central office-based or consultants. For the purpose of this report, we refer to all coaches and teachers as *school-based staff*.

classroom lessons for review with coaches and developing and executing an action plan for improvements. The value of school-based personnel time was captured using national average salaries, inclusive of benefits (explained in detail in the section Determination of Prices for School-Based Staff).

Coach Specialists. Another essential personnel resource involved in the implementation of MTP-S was that of a coach specialist; however, coach specialists are included in the costs analysis differently than teachers and coaches. Coach specialists work for Teachstone and provide training and support to coaches to implement the program. This includes conducting the MTP-S and CLASS-S pre-service trainings for coaches prior to the first year and holding check-in meetings with the coaches one on one, as a group, and ad hoc throughout the coaching cycles. The Teachstone program fees (described in detail in the Program Fees section) include a salary and benefits for coach specialists. Therefore, the costs of the coach specialist's time are not estimated separately for the purposes of the cost study. In other words, we do not attempt to calculate and attach a value to the labor time expended by coach specialist in the manner that we do for teachers and coaches. Instead, the costs associated with coach specialists are assumed to be fully covered by Teachstone program fees.

Exhibit H1 summarizes school-based personnel positions and their roles in the MTP-S program.[20]

Exhibit H1. MTP-S Program Personnel

Personnel Position	Description	Duties
Coach	Personnel who have worked in the respective school district for 2 or more years, have 5 or more years of teaching experience, and have at least some coaching or supervisory experience.	Provide assistance and training to teachers as they carry out the program activities (i.e., coaching cycles). Coaching cycle activities: • Preparation work prior to coaching cycle meetings (i.e., Step 2 of Exhibit 2) • Coaching cycle conference (i.e., Step 4 of Exhibit 2) • Post-meeting work (i.e., Step 5 of Exhibit 2) • Additional coaching support

---

[20] Staff from Teachstone (coach specialists) are also an integral part of the program, as they provide coaches with training and support throughout the coaching cycle process. They are employed by Teachstone and have experience as both an MTP-S coach and a coach supervisor. However, the value of their time is included in the program fees and therefore not calculated separately in this personnel section.

Personnel Position	Description	Duties
Teacher	Personnel who work at the eligible schools, have taught a yearlong mathematics or ELA class in Grades 6–12, have not participated in an induction program involving regular mentoring, and plan to remain in the current school for at least the next 2 school years	Participate in MTP-S programming. Coaching cycle activities: • Preparation work prior to coaching cycle meetings (i.e., Steps 1 and 3 of Exhibit 2) • Coaching cycle conference (i.e., Step 4 of Exhibit 2) • Post-meeting work (i.e., execute plan) • Additional coaching support
Coach Specialist*	Personnel who are employed by Teachstone and have experience as both an MTP-S coach and a coach supervisor	Provide assistance and training to coaches as they carry out the program activities.

*In the cost study, the value of coach specialist time is captured in the program fees while teacher time is valued using national standardized prices and benefit information.

MTP-S Program Components. Based on MTP-S theory of change, program components included coach and teacher pre-service trainings and certifications (e.g., MTP-S and CLASS-S trainings for coaches and annual trainings for teachers); coaching cycles, which include additional coach support (i.e., time school-based staff spent outside of coaching cycles); coach specialist monitoring and support (e.g., facilitating coach trainings and conducting 1:1 and group check-ins); and MTP-S program fees. Exhibit H2 summarizes the critical program components, and the program activities within each component, that contribute to the outcome of the MTP-S program.

Exhibit H2. MTP-S Program Components

Program Component	Description
Training	**Coach Pre-service Training:** One-time (Year 1 only), 5-day training on MTP-S and CLASS-S
	Coach Recertification Training: Annual MTP-S recertification training (Year 2 only for this study) that takes approximately 15.5 hours based on Teachstone's estimated time to complete recertification.
	Teacher Training: Annual, orientation on MTP-S program requirements and the coaching cycle process in Year 1 and refresher teacher training in Year 2.
Coaching Cycles	**Coaching Cycle Process:** 2-week work cycles involving coaches and teachers videotaping lessons, clipping videos, writing and reviewing prompts, holding conference meetings, developing a post-meeting action plan, and carrying out the action plan.

Program Component	Description
	Additional Coaching Support: All additional time coaches or teachers spent outside of the coaching cycle process or contracted time. For coaches, this includes additional support for teachers, communications (e.g., emails), and logistics planning. For teachers, this includes exploration of MTP-S and myTeachstone resources and participating in the CLASS-S Community online discussion board.
	Monitoring and Support From Coach Specialist: Time coaches spend with the coach specialist, including biweekly, one-on-one calls; coaching cycle review calls; and monthly group check-in meetings with all MTP-S coaches.
Program Fees	**MTP-S Program Fee:** One-time (Year 1 only) fees for MTP-S program, including fees for certifying coaches in the MTP-S program model, certifying coaches as a CLASS-S Observer, and for teacher program materials.
	Coach Recertification Fee: After the first year of MTP-S programming, if coaches wish to maintain their certification, they must take an annual CLASS-S recertification assessment. (Year 2 only for this study)

Training. Initial and ongoing training were both critical to implementing the MTP-S program. In the first year (Year 1), coaches participated in a one-time, 5-day MTP-S and CLASS-S pre-service training and assessment led by coach specialists (i.e., Teachstone staff). Coaches then led an orientation training for teachers on MTP-S program requirements and the coaching cycle process. In the following program year (Year 2), coaches participated in a recertification training to continue as an MTP-S coach, and teachers were given a refresher training.

Coaching Cycles. Throughout both program years, coaches and teachers conducted the coaching cycles based on the process described in Exhibit H2. "Additional coaching support" includes activities that may have contributed to the outcome of the program but are not formal/required steps in the coaching cycle process. These steps include, for example, exploring myTeachstone and CLASS-S resources, providing additional 1:1 support for teachers, unanticipated additional effort to review lessons or create action plans, communications, and logistics planning. Teachstone aims for each teacher to complete 10 coaching cycles per school year (8 months), but teachers are only required to complete six cycles per school year, as this is the threshold for adequate participation. In this analysis, we used the average number of coaching cycles completed in Year 1 (seven cycles per teacher) and Year 2 (eight cycles per teacher). Additionally, throughout the 8 months of the coaching cycle process, coaches receive support from Teachstone coach specialists through monthly, 1:1 cycle review calls and group meetings with all MTP-S coaches.

Program Fees. Program fees are another component of the MTP-S program. The MTP-S Program Fee covers the cost of the pre-service training, including coach specialist time as they support coaches throughout the program, coach certification fees for the MTP-S program model and CLASS-S observer, and program materials needed to carry out coaching cycles (e.g.,

recording device for teachers to use to record their lessons, program booklet, access to the online Teachstone dashboard). To maintain the CLASS-S observer certification, coaches must pay an annual recertification fee. Each program fee is summarized in Exhibit H3.

Exhibit H3. MTP-S Program Fees*

Program Fees	Description
MTP-S Program Fee	One-time program pre-service training fee to certify MTP-S coaches (maximum of 10 coaches per training).
CLASS-S Program Fee	One-time program pre-service training fee to certify MTP-S coaches as CLASS-S observer (maximum of 15 coaches per training).
Materials Fee	One-time fee for materials packet used by teachers to carry out coaching cycle activities. Materials packet includes program booklet, print materials, audio/visual kits (recording device), access to Teachstone dashboard, and other materials needed to carry out coaching cycle activities.
Recertification Fee	After the first year of MTP-S programming, if coaches wish to maintain their certification, they must take an annual CLASS-S recertification assessment.

Since all MTP-S activities were conducted virtually, our analysis does not include any in-person costs such as travel time, transportation, or in-person programming costs from Teachstone. Future implementation that includes in-person programming will vary in cost.

Data Sources

To understand the resources used in MTP-S implementation and their associated costs, the study team gathered data from interviews and surveys as well as extant data from Teachstone. Interviews and surveys were used to capture the self-reported measures of time coaches and teachers spent on each program activity, Teachstone extant data helped us understand the *estimated* amount of time coaches and teachers spent on program activities and cost of program fees.[21] Additionally, public pricing data helped us value and determine the costs associated with staff time across all program components. Exhibit H4 aligns each data source with the associated program components, followed by a detailed description of each data source.

Exhibit H4. Data Sources for MTP-S Program Components

Program Component	Data Source		
	Interview	Survey	Extant Data
Training: Coach	X	X	X
Training: Teacher	X		

---

[21] Teachstone extant data for time coaches and teachers spent on program activities was an estimation of how much time staff *should* spend on each activity. These data do not represent how much time each activity actually took staff to complete in real time as it relates to this implementation study.

| | Data Source | | |
Program Component	Interview	Survey	Extant Data
Coaching Cycle	X	X	
Program Fees			X

Interviews. AIR conducted interviews with 18 participating coaches and coach specialists (13 in spring 2022 of Year 1 and five in spring 2023 of Year 2) to gather information across both years of the implementation study.[22] More details about interviews are described in the Implementation Evaluation section of the report. Interview data applicable to the cost study included the average number of hours that coaches spent on coaching cycle activities and additional coaching support.[23]

Surveys. As outlined in the Implementation Evaluation section of the report, AIR created and distributed electronic surveys to all teachers to understand the average amount of time spent on each activity and to receive feedback on the program process (used in the implementation study). The survey included questions about resources used, such as number of hours each teacher spent on coaching cycle and additional coaching support activities, such as preparation work prior to coaching cycle meetings (i.e., videotape lesson, write response to video clips and prompts); coaching cycle conference; executing the action plan; and exploring myTeachstone and CLASS-S resources.[24]

Extant Data. Teachstone extant data included a description of program fees (summarized in Exhibit H3) and "District Service Agreements" (i.e., agreements between AIR and each participating district). The District Service Agreements included information on the *estimated* number of coaching hours that Teachstone expected each program activity to take in Year 1 and Year 2.

Data Analyses

Estimation of Time Spent on Training and Coaching Cycles

For most program components, the number of hours that school-based staff spent on a program component activity were obtained from the interview and survey data. However, in some instances, no data were available. In those instances, *estimated hours* from District Service Agreements were applied rather than *self-reported hours*, thereby potentially biasing our results (described more in the Limitations section). For example, coaches were not

---

[22] Only 13 coaches and coach specialists participated in the interviews in Year 1; however, there was a total of 16 coaches and three coach specialists who participated in the implementation of the program in Year 1.
[23] The interview protocol is available upon request.
[24] Data collection instruments are available upon request.

interviewed about the number of hours spent attending the pre-service training or monthly 1:1 and group meetings with the coach specialist; thus, Teachstone's estimated hours of how much time school-based staff *should* spend on an activity as outlined in the District Service Agreements were applied.

All time allocated by school-based staff for MTP-S programming activities was cataloged and converted to annual quantities (hours per coaching cycle × number of coaching cycles) to obtain total annual number of hours per person for each program component in Year 1 and Year 2. Exhibit H5 summarizes the annual activity hours allocated by each school-based staff member, including how those hours are spread across program components. For example, a single coach spent an average of 79 hours on the "coaching cycle process" across all their assigned teachers in Year 1 and about 81 hours in Year 2.

Exhibit H5. Annual Hours Allocated Toward Training and Coaching Cycle MTP-S Program Components by School-Based Staff in Year 1 and Year 2

Program Component	Program Activity	Year 1 Hours per Personnel		Year 2 Hours per Personnel	
		Coach (n = 16)	Teacher (n = 44)	Coach (n = 5)	Teacher (n = 20)
Training	Coach pre-service training and recertification	66*	--	16*	--
	Annual Teacher Trainings	6	2	2	1
TOTAL TRAINING		**72**	**2**	**18**	**1**
Coaching Cycles	Coaching cycle process	79	12**	81	12
	Additional coaching support	32	15	13	3
	Monitoring and support from coach specialist	20*	--	20*	--
TOTAL COACHING CYCLES		**131**	**27**	**114**	**15**

Note. All hours represented here are rounded to the nearest whole number.

*For this program, activity estimated hours from Teachstone data were substituted for self-reported hours, as self-reported data were not available.

**The number of hours for this program activity are based on Year 2 Teacher Survey responses because there was a survey delivery error in Year 1, and data were not available. Self-reported hours in Year 2 for this activity were similar to those estimated by Teachstone.

Time allocation toward "Additional Coaching Support" varied widely for school-based staff in Year 1. For example, teachers reported spending between 0 and 200 hours (an average of 15 hours) on additional coaching support in Year 1, while coaches reported spending between 2 and 66 hours (an average of 32 hours) in Year 1 on additional coaching support. Most teacher respondents said they spent fewer than 50 hours on additional coaching support; however, one

respondent stated that they spent 200 hours across the year on this subtask. Based on the written response from this individual, this respondent did experience exceptional hardship with program implementation. While this extreme value inflates the self-reported time spent outside of the coaching cycle process, we included this single input in the calculation because it is part of time spent on the program implementation that led to the production of the outcome observed in this study. Additionally, this may reflect future implementations where one or more individuals have substantial hardship with program implementation.

Note that the "additional coaching support" program activity is only one variable of the coaching cycle program component and includes activities like exploring MTP-S and myTeachstone resources and participating in the CLASS-S Community online discussion board. The variation in teacher- and coach-reported time may be explained by initial technical difficulties with the program platform, familiarity with the material, and time needed to accomplish the required tasks. The variation could also be due to the varying levels of support needed by teachers in this study. With the change in participant sample size from Year 1 to Year 2, there is no definitive way to conclude why self-reported time was lower in Year 2.

Determination of Prices for School-Based Staff

To value the time of coaches and teachers (both of whom are school-based staff), we calculated the weighted averages of the number of years of experience and education of the school-based staff sample in this study. We then used the National Center for Education Statistics (NCES) annual national average salaries for full-time teachers in public elementary and secondary schools in 2021–22 dollars with similar qualifications and experience as our sample population (defined in the Methods section).[25] We adjusted the NCES annual national average salaries for inflation using the U.S. Bureau of Labor Statistics Consumer Price Index for 2022 (2021–22 school year) and 2023 (2022–23 school year) (1.053% inflation rate).[26] Next, we applied a 35% benefit rate to the annual national average salaries to account for overall compensation (pay and benefits).[27] We assumed the benefit rate for school-based staff in this study is equal to the average rate for instructional staff in the U.S. Bureau of Labor Statistics, Employer Costs for Employee Compensation data.[28] We then converted school-based staff annual national overall

---

[25] U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public School Teacher Data File," 1990-91, 1999-2000, 2007-08, and 2011-12; "Charter School Teacher Data File," 1999-2000; and National Teacher and Principal Survey (NTPS), "Public School Teacher Data File," 2015-16, 2017-18, and 2020-21. (Table prepared in September 2022.) https://nces.ed.gov/programs/digest/d22/tables/dt22_211.20.asp
[26] U.S. Bureau of Labor Statistics Consumer Price Index for All Urban Consumers (CPI-U), 2022-23. https://www.bls.gov/cpi/
[27] Benefit rate source: https://www.bls.gov/opub/btn/volume-8/teachers-job-requirements.htm
[28] Benefit rate source: https://www.bls.gov/opub/btn/volume-8/teachers-job-requirements.htm

compensation to hourly rates by dividing by 1,440 hours worked per year.[29,30,31,32] Exhibit H6 reports the annual and hourly compensation rates used for the cost study.

Exhibit H6. Compensation Rates for Personnel (in 2023 Dollars)

Position	Annual National Overall Compensation	National Hourly Rate
Coach	$98,273	$68.25
Teacher	$87,402	$60.70

Estimation of Program Fees

The program fees charged by Teachstone are an important part of the program as they include training, materials, and coach specialist (i.e., Teachstone staff) time.[33] Teachstone coach specialists provided the MTP-S and CLASS-S pre-service trainings to the district coaches, as well as continued monitoring and support throughout the coaching cycles, at $9,225 per 10 coaches (MTP-S training) and $8,460 per 15 coaches (CLASS-S training). In this implementation, Year 1 had 16 coaches, so the total annual cost of MTP-S and CLASS-S pre-service trainings was $18,450 ($1,153 per coach) and $16,920 ($1,058 per coach), respectively. Additionally, teachers required equipment and materials, such as a recording device and online access to the Teachstone dashboard, to carry out coaching cycle activities. All materials are included in the MTP-S teacher packet, which has a one-time fee of $550 per teacher. That is, $24,200 in total were incurred in Year 1 for 44 MTP-S teacher packets. Lastly, the recertification fee was only applied in Year 2 (2022–23) for this analysis. As previously mentioned, because of the COVID-19 pandemic, several teachers dropped out of the program and fewer coaches were needed in Year 2. In Year 2, the recertification fee of $6,030 per coach was applied across five coaches,

---

[29] See Shand, R., & Brooks, B. (2021). *Empirical support for establishing common assumptions in cost research in education* (p. 9, Exhibit 3). University of Pennsylvania, Scholarly Commons, Center for Benefit-Cost Studies of Education. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1001&context=cbcse

[30] See Shand, R., & Brooks, B. (2021). Research recommends a preferred value of 1,440 hours per year (8 hours per day at 180 days per year) for number of hours worked by educators in a K–12 academic year. This value is a mid-level representation of the number of hours worked in a K–12 school year and is based on national survey data and extensive review of literature, as outlined in the report.

[31] Based on the District Service Agreement, Teachstone estimated that coach support hours would occur over an 8-month period to ensure that enough time is allocated to complete 10 coaching cycles. Although training was done outside of the school year, total hours did not amount to 40 hours per week. In this study, coach work hours are assumed to be equivalent to time worked during the school year: 8 hours per day at 180 days per year equals 1,440 annual work hours.

[32] For each year of the program, teachers worked approximately 8 months (1 standard school year), or 8 hours per day at 180 days per year equals 1,440 annual work hours.

[33] Program costs for MTP-S vary based on in-person versus virtual implementation. Due to the disruption and adverse effects of the COVID-19 pandemic, activities in this program were conducted virtually. While AIR received a discount price for the program, this cost study reflects Teachstone's full retail price for virtual programming that would be typical in future implementation.

resulting in an overall Year 2 cost of $30,150. Exhibit H7 outlines the per-coach and per-teacher costs of program fees.[34]

Exhibit H7. Costs of MTP-S Program Fees per School-Based Staff Across Year 1 and Year 2

Program Fee	Year 1 Prices per Personnel		Year 2 Prices per Personnel	
	Coach (n = 16)	Teacher (n = 44)	Coach (n = 5)	Teacher (n = 20)
MTP-S Pre-service Training	$1,153	--	--	--
CLASS-S Pre-service Training	$1,058	--	--	--
CLASS-S Annual Certification	--	--	$6,030	--
MTP-S Teacher Packet	--	$550	--	--

Note. We adjust for the present value of program fees, along with all resource prices, in the next section once costs have been aggregated for Year 1 and Year 2.

Cost Analysis

Coach and teacher estimates of time allocation (overall and per program component) and the numbers of staff involved in each program year (i.e., Year 1 and Year 2) were paired with salary pricing information (described above) to estimate the overall cost of school-based staff in Year 1 and Year 2. For each year, the total cost of school-based staff time was summed for each program component and then divided by the number of teachers served in the respective year to estimate the year-specific staffing cost on a per-teacher basis. These year-specific, per-teacher costs of each program component were then summed together in each year to provide an annual per-teacher cost. The Year 1 costs were adjusted to future values for the base year of Year 2 (2023 dollars) using a 3.5% discount rate so that the costs reflect those at the time of program completion.[35] These annual per-teacher program costs were then aggregated across Year 1 and Year 2 to generate the overall per-teacher cost associated with school-based staff time devoted to program implementation. These calculations are expressed in the equations that follow, where *i* represents a given MTP-S program component; *Coach Time* and *Teacher Time* are the total amount of hours school-based staff spent on program component *i*; *Coach Compensation* and *Teacher Compensation* are the fixed, overall school-based staff compensation (pay and benefits); and *Number of Teachers* is the total number of treatment teachers in a given year (Year 1 or Year 2) of the program.

$$Per\ Teacher\ Staff\ Time\ Cost_{Y1} =$$
$$\frac{[(\sum_{i=1}^{I} Coach\ Time_{i,Y1} \times \overline{Coach\ Compensation}) + (\sum_{i=1}^{I} Teacher\ Time_{i,Y1} \times \overline{Teacher\ Compensation})]}{Number\ of\ Teachers_{Y1}}$$

$$Per\ Teacher\ Staff\ Time\ Cost_{Y2} =$$
$$\frac{[(\sum_{i=1}^{I} Coach\ Time_{i,Y2} \times \overline{Coach\ Compensation}) + (\sum_{i=1}^{I} Teacher\ Time_{i,Y2} \times \overline{Teacher\ Compensation})]}{Number\ of\ Teachers_{Y2}}$$

$$Overall\ Per\ Teacher\ Staff\ Time\ Cost = Per\ Teacher\ Cost_{Y1} + Per\ Teacher\ Cost_{Y2}$$

Similarly, total program fees from Year 1 and Year 2 were divided by the number of teachers served in the respective year to produce the year-specific program fee on a per-teacher basis. Then we applied the discount rate of 3.5% to the program fees in Year 1 to adjust for future value of program fees in 2023 dollars.[36] Next, the annual per-teacher cost of program fees for Year 1 and Year 2 were summed together to generate the overall per-teacher cost associated with program fees in this 2-year MTP-S program implementation.

$$Total\ Per\ Teacher\ Cost_{Y1} = \frac{Program\ Fees_{Y1}}{Number\ of\ Teachers_{Y1}}$$

$$Total\ Per\ Teacher\ Cost_{Y2} = \frac{Program\ Fees_{Y2}}{Number\ of\ Teachers_{Y2}}$$

$$Overall\ Per\ Teacher\ Cost\ of\ Program\ Fees$$
$$= Total\ Per\ Teacher\ Cost_{Y1} + Total\ Per\ Teacher\ Cost_{Y2}$$

Lastly, the overall per-teacher cost of school-based personnel time and the overall per-teacher cost of program fees for the 2-year program were summed to produce the estimated overall per-teacher cost of this MTP-S implementation.

$$Estimated\ Overall\ Per\ Teacher\ Cost\ for\ Two\ Year\ MTP\ S$$
$$= Overall\ Per\ Teacher\ Staff\ Time\ Cost$$
$$+ Overall\ Per\ Teacher\ Cost\ of\ Program\ Fees$$

Cost-Effectiveness Analysis

While this study intended to include a cost-effectiveness analysis, the study team found no statistically significant impacts on primary outcomes; thus, cost-effectiveness ratios would not

---

[36] See American Institutes for Research, 2021.

be informative. The Cost Analysis Standards Project (CASP) recommends that "[f]ollowing the reporting of costs and effects, a cost-effectiveness ratio should be reported for primary outcome(s)" but "[w]hen effect sizes for primary outcomes are statistically indistinguishable from zero, it is not appropriate to calculate a ratio except when the null effect is precisely estimated to be zero" (CASP, 2021, pp. 46).

Limitations

There are a few limitations to this study that help contextualize the findings and highlight areas for consideration and additional research. In addition to issues with data collection, the key limitation for this study was interruptions due to the COVID-19 pandemic that affected typical school activity as well as teacher capacity to participate in a study.

Survey delivery error. We encountered a survey delivery error that impacted data availability from the teacher survey in Year 1—specifically, the number of hours teachers spent on coaching cycle activities. This required us to substitute the self-reported hours in the Year 2 teacher survey for Year 1 (1.77 hours per coaching cycle). We found the self-reported hours in Year 2 to be comparable to Teachstone's estimate of 1.75 hours per teacher per coaching cycle.

Survey responses. On average, the response rate for the cost study questions was 60% in Year 1 and 100% in Year 2. We did not have a full response rate from teachers for the surveys in Year 1. For some questions, we only received responses from a proportion of all teachers. Since a proportion of answers were not included in the average self-reported hours for teacher time, the values may be skewed relative to actual observed hours.

Additionally, there were some extreme values in the teacher and coach responses related to the "additional coach support" program component. The extreme values skewed the average hours worked so that they are slightly higher relative to Teachstone expected hours. Upon review of written responses in the survey, we thought these extremes were valid responses based on individuals' experiences and chose to include them in this analysis.

Data limitations. The cost study was designed asynchronously with the impact analysis, resulting in survey and interview questions not linked directly to ingredients/resources in the cost model. This creates an inherent limitation in our results due to lack of alignment with best practices for capturing personnel time based on actual observed data. Ultimately, some of the reported values in the cost analysis are estimated hours rather than actual hours worked.

COVID-19 pandemic constraints. As this program took place during the COVID-19 pandemic, several unexpected changes occurred, such as moving the program fully virtual and causing extra strain on school-based staff during times of high variability. We acknowledge that this may have impacts that we were unable to control for in this study.

About the American Institutes for Research®

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of education, health, and the workforce. AIR's work is driven by its mission to generate and use rigorous evidence that contributes to a better, more equitable world. With headquarters in Arlington, Virginia, AIR has offices across the U.S. and abroad. For more information, visit AIR.ORG.

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

21881_08/23