PROJECT NARRATIVE

Contents

Page

INTRODUCTION

American Institutes for Research (AIR), in collaboration with Learning Forward, Teachstone, and five school district partners (SDPs), proposes a mid-phase grant focused on a strategy for scaling MyTeachingPartner-Secondary (MTP-S)—a video-based, Web-mediated teacher coaching program for middle and high school teachers that improves teacher classroom practice, student engagement, and student achievement. Teachers participate in the MTP-S program for 2 years, each including a series of six to 10 cycles of coaching, for a total of 15 to 23 hours annually. Two randomized controlled trials (RCTs) have tested MTP-S, in distinct settings, both showing positive impacts on student achievement (effect sizes = 0.22–0.48). The first study focused primarily on middle school teachers and was conducted in 12 schools in rural and small town districts (Allen, Pianta, Gregory, Mikami, & Lun, 2011). The What Works Clearinghouse (WWC; 2012, 2015) rated the first study as meeting WWC standards with reservations. The second study, not yet reviewed by WWC, focused primarily on high school teachers and was conducted in a large, "highly challenging" urban district with high levels of poverty (Allen, Hafen, Gregory, Mikami, & Pianta, 2015).

The project's goal is to refine and test a strategy for scaling MTP-S in diverse settings that serve high-need students and build a network to support continued scaling. As the provider for MTP-S, Teachstone typically provides coaching services directly to teachers. However, most districts prefer to deliver coaching with their own staff. Therefore, Teachstone's strategy for scaling MTP-S is to select and train local SDP staff as MTP-S coaches and to monitor and support them.

To facilitate iterative testing and refinement of the scaling strategy, Teachstone and the SDP partners will implement the 2-year MTP-S program in three successive cohorts: Cohort 1 schools will begin in 2018–19, Cohort 2 schools in 2019–20, and Cohort 3 schools in 2020–21,

as shown in Exhibit 1. Across these cohorts, Teachstone will use feedback routines, internal data, and evaluation data from AIR to improve their scaling strategy continuously and revise its manuals and other supporting materials for future implementations.

Exhibit 1. Number of Teachers in Each Cohort by Year

	2017–18	2018–19	2019–20	2020–21	2021–22
Cohort 1		50 (25 T)	50 (25 T)		
Cohort 2			100 (50 T)	100 (50 T)	
Cohort 3				150 (75 T)	150 (75 T)

To determine which teachers from each cohort will receive MTP-S, AIR will identify teachers within each school who want to participate in MTP-S and randomly assign half to treatment and half to control. These groups will be the basis for answering the impact questions for AIR's independent evaluation. The size of each successive cohort will increase such that Cohorts 1 and 2 combined, as well as Cohort 3 by itself, will be large enough to provide a sufficiently-powered test of impact on student achievement. The 5-year budget amounts to $159.46 for each year a student receives instruction from a teacher with MTP-S training.

Each cohort will include urban, suburban, and rural schools with high percentages of high-need students (see section B.1). While MTP-S is being delivered in these schools, Learning Forward will set up supports that facilitate the spread of MTP-S as part of coherent systems of teacher support during and after the project.

ABSOLUTE PRIORITIES

The project addresses **Absolute Priority 1—Supporting High-Need Students—**by providing MTP-S in high-poverty and high-minority schools, based on state definitions under the Every Student Succeeds Act. The project also addresses **Absolute Priority 5—Evidence-Driven Practices—**by focusing on MTP-S. MTP-S had statistically significant positive impacts on student achievement in two trials and addresses two **areas of critical national need:** the need for

interventions that improve teacher quality and for interventions that improve adolescent academic engagement.

A. SIGNIFICANCE

A.1. Severity of the Problem

The problem to be addressed by the project is that too many students, particularly students of color and students from low-income families, move through middle and high school without being adequately prepared for college. The 2016 scores on the National Assessment of Education Progress (NAEP; U.S. Department of Education, 2016) underscore the severity of the problem: 48% of Black students scored *below basic* on the 12th-grade NAEP reading assessment compared with 21% of White students. Among students whose parents did not graduate high school, 45% scored *below basic* versus 19% for children of college graduates.

These **severe inequities** in academic readiness play out in educational attainment. As The Education Trust (2017) reports, young adults from high-income families are more than three times as likely as those from low-income families to have earned a bachelor's degrees by age 24. Black students are less likely than Whites to graduate from high school on time (71% and 87%, respectively) and far less likely to hold a bachelor's degree as a young adult (20% and 39%, respectively). And among those who do manage to graduate high school and enter 2- or 4-year institutions, students who are disadvantaged are more likely to enroll in remedial classes and eventually drop out (Chen & Simone, 2016).

These gaps are certainly not immutable. Since 1999, the United States made progress in closing the academic achievement gaps for elementary age students. But going back as far as 1990, the gap for 17-year-olds has not changed, **underscoring the need to find ways to improve outcomes for high-need adolescents** (The Education Trust, 2017).

A.2. National Significance

The proposed project is significant because it has the potential to reduce achievement gaps and improve achievement and attainment for **high-need students** in secondary schools. It focuses on **high-poverty and high-minority schools exclusively** and addresses two **areas of critical national need**: teacher quality and adolescent academic engagement.

Teacher quality is seen by researchers and policymakers as a potential lever on student achievement, based on a large body of research showing that teachers vary in their effects on students' engagement, achievement, and later life outcomes (see, e.g., Chamberlain, 2013; Chetty, Friedman, & Rockoff, 2014; Gershenson, 2016; Rivkin, Hanushek, & Kain, 2005). But effective approaches to improving teacher quality are hard to find, especially teacher professional development (PD) programs. (See Blazar, Kraft, and Hogan [2017]; Desimone & Garet [2015]; Gersten, Taylor, Keys, Rolfhus, & Newman-Gonchar [2014]; and Kennedy [in press] for reviews.) The scarcity of effective PD programs is frustrating given the potentially large numbers of students who could benefit. Whereas reforms to preservice teacher preparation can improve the teacher workforce slowly at best, in individual graduating cohorts, teacher PD can be delivered to the current teacher workforce.

The second area of critical national need—adolescent academic engagement—is an urgent challenge in middle and high school. Following students across time reveals that engagement declines during middle and high school, beginning as early as age 11; in surveys, more than half of adolescents report no serious commitment to learning (Allen et al., 2015). Low engagement in adolescence is associated with a range of future problems, including dropout, depression, aggressive behaviors, delinquency, gang involvement, risky sexual behavior, and low earnings (Li & Lerner, 2011; Salmela-Aro, 2017; Voisin & Elsaesser, 2016).

MTP-S is designed to address both areas of critical national need, boosting teachers' skills and fostering adolescent academic engagement. To do so, it targets validated, measurable dimensions of classroom practice defined by the Classroom Assessment and Scoring System-Secondary (CLASS-S), which measures teacher-student interactions associated with adolescent engagement and learning, specifically the extent to which teacher-student interactions build a positive emotional climate and demonstrate sensitivity to adolescents' needs for autonomy, an active role in their learning, and a sense of the relevance of course content to their lives (see Allen et al., 2002; Allen, Hauser, Bell, & O'Connor, 1994; Allen, Kuperminc, Philliber, & Herre, 1994; Ryan & Deci, 2000). (See Appendix G.6 for domains and dimensions measured by CLASS-S.) All the MTP-S processes and resources are organized to support improvement on these aspects of instruction. Specifically, each year of the 2-year program starts with a half-day workshop for teachers focused on the principles of adolescent motivation and social interaction that are the basis of CLASS-S. All other aspects of the MTP-S coaching occur in 12–20 cycles—six to 10 per year—designed to stimulate improvement along the dimensions of teaching measured by CLASS-S (see Exhibit 2). Its focus on adolescent academic engagement also gives MTP-S its broad potential reach; it can help middle and high school teachers in any content area implement curriculum more effectively.

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

Step 1.	The teacher videotapes 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 the CLASS-S as a lens. The coach selects three 1-minute clips, each involving examples of teacher-student 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 those interactions. Per the coach manual, some clips are provided to recognize effective teacher-student interactions and others to identify interactions that could be improved.

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 videoconference.
Step 4.	During a 20- to 30-minute videoconference, 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.
Step 6.	The teacher reads the conference summary and follows through on the action plan.

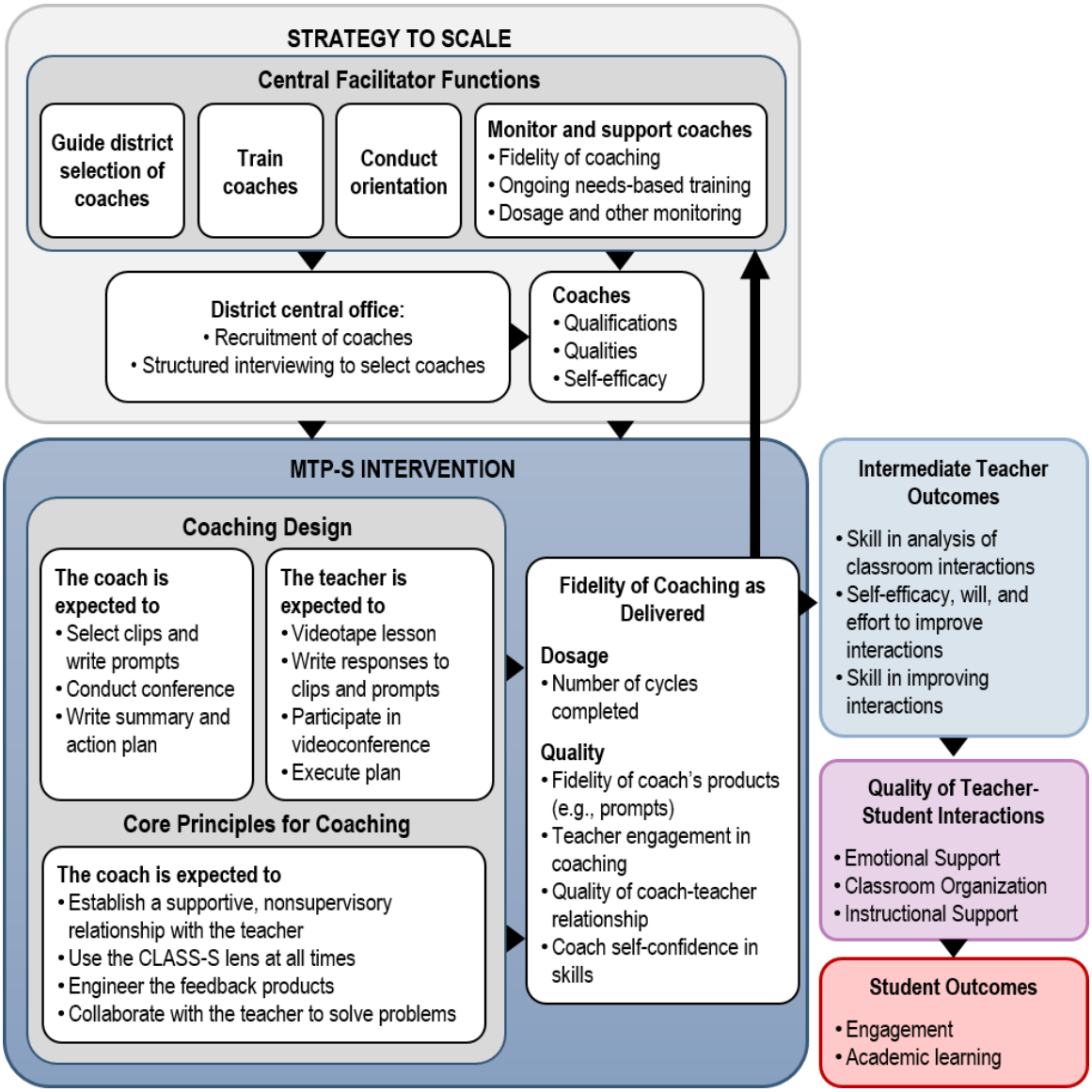
In sum, by targeting **two areas of critical national need**—teacher quality and adolescent academic engagement—MTP-S potentially can reduce persistent gaps in achievement and attainment for high-need adolescents across the United States, if we develop an effective strategy to scale it.

A.3. Exceptional Approach to Absolute Priority 5

This project represents an exceptional approach to Absolute Priority 5—Evidence-Driven Practices—because of the **exceptional depth of evidence and scholarship driving the MTP-S theory of change**. Teacher PD interventions generally link two theories of change (Wayne, Yoon, Zhu, Cronen, & Garet, 2008). One is a theory about how classroom practices and interactions affect student outcomes, what we call the *theory of student change.* As shown in the bottom right of Exhibit 3 in purple and pink, the MTP-S theory of student change is that increasing the *Quality of Teacher-Student Interactions* leads to improvement in two *Student Outcomes*: *student engagement* and, in turn, *academic learning*. **Decades of evidence** on the measurement of classroom interactions and adolescent engagement underlie this part of the MTP-S theory of change, as described earlier (Section A.2). Some more recent **evidence** comes from the first MTP-S trial: analysis of mediating variables in the Allen et al. (2011) found that half of MTP-S's total effect on student achievement was mediated by the quality of classroom

interactions as measured using CLASS-S. There also is **correlational evidence** supporting the link between CLASS-S scores and student engagement (Gregory, Allen, Mikami, Hafen, & Pianta, 2013) and between CLASS-S scores and student achievement (Allen et al., 2013).

Exhibit 3. Theory of Change for the Strategy to Scale and the MTP-S Intervention



The second part of the theory of change underlying MTP-S is called the *theory of teacher change.* It addresses how to get teachers to learn new practices and use them in the classroom (see the blue and purple parts of Exhibit 3). The *Coaching Design* processes are intended to build

Intermediate Teacher Outcomes, including teacher skills and self-efficacy, and in turn improve the *Quality of Teacher-Student Interactions,* as depicted in Exhibit 3. Like the MTP-S theory of student change, the MTP-S theory of teacher change is **based in evidence**. MTP-S coaches are trained to apply four evidence-based *Core Principles for Coaching:*

1. *Establish a supportive, nonsupervisory relationship with the teacher*. **Evidence** shows that teachers can benefit from feedback that does not have stakes attached (e.g., Steinberg & Sartain, 2015; Taylor & Tyler, 2012). Some hypothesize that teachers may be less motivated to change their practices if they view feedback as being used for accountability (e.g., Atwater et al., 2007; Danielson, 2008; Smither et al., 2005).
2. *Use the CLASS-S lens at all times.* **Evidence** shows that teachers' capacity to analyze classroom interactions is a factor in their contributions to student achievement (Downer, Pianta, Burchinal, Field, Hamre, & Scott-Little, 2014; Hamre et al., 2012). To develop teachers' skill in analyzing classroom interactions, coaches use the CLASS-S lens and lexicon at all times—selecting video clips, writing prompts, talking in the videoconference, and writing the conference summary. Doing so builds a common language and makes the feedback seem more objective (Sartain, Stoelinga, & Brown, 2011). **Evidence** from studies of narrative feedback shows that this coherence increases the information value of the feedback (Brutus, 2010; Rowan & Raudenbush, 2016).
3. *Engineer the feedback products (e.g., prompts).* Across the MTP-S process, coaches are trained to follow **evidence-based guidelines**. For example, coaches focus feedback on the observed practices and interactions, not on the teacher; they describe practice qualitatively, not using numeric ratings; and they keep the scope of feedback manageable, not more than

the teacher can process. Some of these same guidelines appear in the industrial and organizational psychology literature on formative feedback (see Shute, 2008).

4. *Collaborate with the teacher to solve problems.* To engage the teacher, coaches are expected to help the teacher identify problems and resolve them, a practice supported by **correlational evidence** about the features of PD and changes in classroom practice (Garet, Porter, Desimone, Birman, & Yoon, 2001). In a systematic **review of evidence** on the impact of teacher PD, Kennedy (in press) highlights this feature of MTP-S as potentially crucial; she notes PD programs that simply tell teachers what to do sometimes show a negative impact.

In sum, MTP-S is exceptionally well grounded in theory and evidence. The project will extend that base of evidence, examining the implementation and impact of MTP-S when delivered using the scaling strategy described in Section B.

B. STRATEGY TO SCALE

B.1. Demand for Scalable, Effective Instructional Coaching

Scalable, effective PD programs are in **demand**, and instructional coaching programs are especially appealing because they provide support that is more individualized and contextualized than traditional programs (e.g., Desimone, Smith, & Phillips, 2013; Pianta, 2005). Districts already invest a significant share of their resources in teacher PD—at least 5% of annual spending in a recent study of three large districts—and want to make those resources count (Hasiotis, Jacobs, & McGovern, 2015). **Adding to the demand**, the Every Student Succeeds Act calls on districts to show that they use proven programs, while at the same time awareness is spreading that few teacher PD programs show impact on achievement when tested in rigorous studies (e.g., Garet, Heppen, Walters, Smith, & Yang, 2016; see Section A.2). The **unmet demand** for scalable, effective coaching programs is especially high in rural schools. Rural

schools often have just a few teachers in a building to share their expertise. In addition, the distance between schools makes it costly for rural districts to deploy coaches who drive from school to school (Glover et al., 2016). MTP-S can fill this need in rural areas because its processes are designed to work online; the exchange of video and text described in Exhibit 2 occurs through a Web portal, and the coach-teacher meetings can use the telephone or Skype.

These conditions led five districts from four states to volunteer to participate in the project. The districts include 223 schools (18 rural, 21 suburban, and 184 urban) defined by their states as high-poverty or high-minority, which is plenty for recruiting all three cohorts (see Appendix D and G.4). This excess of potential sites gives the project flexibility to fill each cohort strategically, balancing across settings in terms of urbanicity and other factors (e.g., size, improvement status, graduation rate) to generate valuable guidance for later replications. If more SDPs are needed during the project for any reason, AIR will use **multiple methods** to recruit more. First, AIR will ask administrators in the five SDPs to introduce AIR to their counterparts in other suitable districts, allowing decision makers in prospective districts hear first-hand from trusted contacts about what participation in the project would mean. Second, Learning Forward will tap into its national networks to identify potential partner districts to introduce to AIR. Third, AIR and Teachstone will conduct a session at Learning Forward's annual conference for districts interested in learning more about trying MTP-S.

B.2. A Specific Strategy to Scale That Addresses Past Barriers

The **barrier** that most prevents the scale-up of MTP-S is its dependence on centralized staff at Teachstone to serve as the coaches, interacting directly with each participating teacher. Districts far prefer to use their own staff or trusted, local consultants, instead of "outsourcing." To address that **barrier**, Teachstone will train staff from each SDP to serve as MTP-S coaches.

Ordinarily, one full-time coach can serve 15 teachers. We expect the study's 150 treatment teachers to be served by a combination of full- and part-time district-based staff. The budget includes resources to pay a share of these costs, with the SDPs paying the remainder.

This strategy to scale requires a central facilitator (CF) employed by Teachstone, with experience as an MTP coach and coach supervisor, to perform **specific** functions:

Guide District Selection of Coaches. A common **barrier** to successful coaching programs is poor selection of staff to serve as coaches (New Teacher Center, 2016). Therefore, a key feature of the MTP-S scaling strategy is the selection of coaches who can implement the coaching process and adhere to the core principles. The CF will provide the SDP's with Teachstone's list of guidelines for desired coach qualities, which include, for example, experience teaching in a middle or high school and skills in building supportive relationships with teachers.

Train Coaches. The CF conducts a 5-day, in-person training for the coaches, which includes the standard 2-day training on rating classroom practice using the CLASS-S plus 3 days on the MTP-S coaching model, as **specified** in the 79-page coaching manual.

Conduct Teacher Orientation. The CF will visit each district in-person to lead the half-day orientation, which will explain the value of coaching, the coaching cycle, and the benefits of participation, such as continuing education credits, to build teacher buy-in.

Monitor and Support Coaches—Fidelity of Coaching. To monitor fidelity of coaching, the CF will conduct a "cycle review" with each coach once every 2 weeks using GoToMeeting. To do so, a CF views the products (e.g., video clips) from a recent cycle and completes a 24-item checklist called the MTP-S Secondary Cycle Feedback Form (see Appendix Exhibit G.2.4), which contains detailed specifications for video clips (3 items), prompts (14 items), and the conference summary (7 items). The CF sends the coach the completed form as the basis for a

one-on-one discussion about the coach's work and examines the teacher's responses to the prompts and rates teacher engagement. Teachstone has found that, before finishing the first year of coaching, many coaches exhibit near-perfect fidelity. To minimize potential **barriers** related to cost, Teachstone will pilot a system whereby coaches who achieve a high fidelity level across cycles will be considered "certified" and therefore monitored for fidelity less frequently. As part of the independent evaluation, AIR will complete some cycle reviews independently, without feedback to the coach, to determine whether the certification system worked as expected.

Monitor and Support Coaches—Ongoing Needs-Based Training. The CF leads biweekly team meetings with all coaches to engage coaches in reflection in a supportive, group context. To help determine the focus of these meetings, the CF will review data from AIR surveys of the coaches in fall, winter, and spring to learn the coaching skills and strategies and MTP-S procedures on which they would like additional information or more support. This information can inform the biweekly meetings and also lead to potential improvements in the coaching resources.

Monitor and Support Coaches—Other. When a coach misses meetings or is struggling, the CF provides enhanced, individual attention and support, to solve problems and provide additional training as needed.

Finally, a **barrier** frequently encountered by coaches is teachers' motivation during the program. Specifically, teachers sometimes do not gather and send video when expected, even though that video is needed as the first step of every cycle. In addition to buy-in efforts at the orientation, steps will be taken to ensure teacher motivation: (1) Give teachers an option to opt-out before random assignment; (2) reimburse teachers for their time; and (3) obtain advanced approval for continuing education credits for participation. In addition, the CF will monitor coaches' progress using the online system and offer support when needed. The threshold for adequate participation for a teacher

is six cycles per year. That was the intended dosage for the most recently reported MTP-S trial, which found an impact on achievement (Allen et al., 2015). In practice, Teachstone plans to tell teachers to complete 10 cycles per year. Based on its experience, many teachers will find that feasible, and some will complete fewer but rarely fewer than six per year.

B.3. Feasibility of Successful Replication in a Variety of Settings and Populations

If the project's evaluation demonstrates that implementing MTP-S using the scaling strategy has an impact on student achievement, it will be feasible to use the scaling strategy to deliver MTP-S **in a variety of settings and populations**, for three reasons: (1) There will be manuals and other materials to support the roles of the CF, coaches, and teachers; (2) the materials will reflect lessons learned during project, including **lessons about challenges associated with particular settings and populations**; and (3) there will be a support network for existing and first-time users of MTP-S. To make replication **feasible**, the project will create all manuals and supporting materials for the CF, including all the CF activities outlined earlier (Section B.2). The materials needed to support the work of the coaches and teachers are already available. The 79-page manual for training coaches includes detailed guidance and resources for all aspects of MTP-S coaching (see Appendix G.2 for the table of contents and sample excerpts). The CLASS-S training, which is a component of the MTP-S coach training, also is fully manualized, and a test of skill at coding videos is available. A website contains the MTP-S online resources, which now include hundreds of videotaped segments of classroom interaction corresponding to the different dimensions of the CLASS-S and independently coded as exemplifying high quality (see Web screenshots in Appendix Exhibits G.2.6-2.8). The website has been designed to support fidelity of implementation by structuring the exchange of classroom videos, video clips, prompts, and responses between the teacher and coach.

Successful use of these materials at the end of the project will be **feasible** because the materials will be iteratively refined across cohorts that include **a variety of settings and populations**. The project will thus generate up to three rounds of lessons about challenges associated with distinct settings and populations. These lessons will be captured through feedback routines built into the project plan. (For details, see Section C.3.) To further increase the **feasibility** of replication, Learning Forward will launch a national network designed to capture and disseminate implementation lessons about using the scaling strategy and MTP-S—including lessons derived with **diverse settings and populations**. The network will engage users of MTP-S, such as teachers, local coaches, and central office staff, to help them support each other with successful implementation. In addition, the network will support nonusers of MTP-S who are considering adopting MTP-S and want to hear current users' perspectives. Learning Forward will also help Teachstone develop a strategy and materials to support districts' efforts to make MTP-S part of a coherent system of support for teachers. For example, in settings where many PD initiatives are underway, it may be useful for the CF and coaches to learn about those initiatives and how they are affecting teachers. To inform Teachstone's strategy for implementing MTP-S in such settings, Learning Forward will solicit input from its existing 22-district community of practice focused on coherent systems of support for teachers and continuous improvement.

C. QUALITY OF THE PROJECT DESIGN AND MANAGEMENT PLAN

C.1. Clearly Specified and Measurable Goals, Objectives, and Outcomes

The project's goal is to test and refine a strategy for scaling MTP-S in diverse settings that serve high-need students and build a network to support continued scaling. Exhibit 4 **specifies the objectives, strategies, and outcomes** to be achieved and indicates **how outcomes will be measured**. Unless otherwise specified, the strategies are to be used for all three cohorts.

Exhibit 4. Objectives, Strategies, Outcomes, and Measures

Strategies	Outcomes	Measures
Objective 1. Implement strategy to scale while continuously using feedback and fidelity data for project improvement.		
Strategy 1.1. Train SDP interviewers to develop and execute plans to recruit and select local staff as coaches.	Local interviewers who are trained for their role in coach recruitment and selection.	Measure 1.1. Attendance records for SDP interviewers; districts' plans for coach recruitment and selection.
Strategy 1.2. Recruit and select local staff to serve as coaches.	Coaches committed to implementing MTP-S with sufficient qualifications and skills to be trained.	Measure 1.2. Coach applicant screening and interview records; agreements for each coach's time commitment.
Strategy 1.3. Train local staff as MTP-S coaches. (All coaches travel to training in Charlottesville.)	Fully trained coaches for each district.	Measure 1.3. Coach training attendance records; coach training fidelity form; coach post-training needs assessment/feedback form.
Strategy 1.4. Monitor and support MTP-S coaches through biweekly one-on-one meetings and biweekly teleconferences using GoToMeeting.	Coaches receive support tailored to individual needs and support for needs that are common across coaches.	Measure 1.4. Coach support attendance records; coach confidence survey; coach-teacher relationship survey; coach interview; CF interview; fidelity of implementation (FOI) checklist; # of cycles completed.
Strategy 1.5. Refine materials and procedures for each strategy under Objective 1.	Improved materials: CF manual; SDP point-of-contact manual; Coach manual; Teacher manual.	Measure 1.5. Biannual memo summarizing revisions made to materials and procedures for Objective 1.
Objective 2. Implement MTP-S while continuously using feedback and fidelity data for project improvement.		
Strategy 2.1 Implement MTP-S orientation. (CF travels to each district.)	Treatment teachers, coaches, and district points of contact understand MTP-S and commit to participation. Treatment teachers are prepared to complete first MTP-S cycle (e.g., videotape classroom).	Measure 2.1. Teacher orientation attendance records; teacher orientation fidelity form; teacher post-training feedback form.

Strategies	Outcomes	Measures
Strategy 2.2. Implement 6–10 MTP-S cycles per teacher each year for 2 years.	Coaches complete intended number of cycles with each treatment teacher with fidelity; treatment teachers report more coaching than control; impacts occur on quality of teacher-student interactions, student engagement, and student achievement.	Measure 2.2. Online system records of number of cycles completed; fidelity of MTP-S coaching checklist; teacher survey items on the frequency and content of PD and coaching; video of classrooms coded using CLASS-S and FFT; student engagement surveys; student achievement records.
Strategy 2.3. Refine materials and procedures for each strategy under Objective 2.	Improved materials: CF manual; District point-of-contact manual; Coach manual; Teacher manual.	Measure 2.3. Biannual memo summarizing revisions made to materials and procedures for Objective 2.
Objective 3. Conduct an RCT to test the impact of MTP-S provided by trained local coaches on (a) the quality of classroom interactions and (b) student outcomes.		
Strategy 3.1. Identify and recruit eligible teachers in the participating schools.	Eligible teachers interested in MTP-S and consent to random assignment.	Measure 3.1. Teacher consent forms.
Strategy 3.2. Randomly assign teachers to treatment and control.	Samples of treatment and control teachers and students with likely equivalence at baseline.	Measure 3.2. Random assignment memo; baseline equivalence memo.
Strategy 3.3. Measure and analyze fidelity of implementation.	Data on fidelity of implementation collected and analyzed.	Measure 3.3. Data collection update; fidelity memo for each strategy for scaling (1.1–1.4) and each strategy for implementing MTP-S (2.1, 2.2).
Strategy 3.4. Measure and analyze treatment-control contrast in teachers' PD experiences.	Data on frequency and content of PD and coaching collected and analyzed.	Measure 3.4. Data collection update; service contrast memo
Strategy 3.5. Determine the impact of MTP-S on classroom practice, student engagement, and student achievement.	Data on outcome measures collected and analyzed.	Measure 3.5. Data collection update; impact memo; study report that meets WWC standards without reservation.
Strategy 3.6. Refine materials and procedures for each strategy under Objective 3.	Improved materials and procedures for teacher recruitment, random assignment, data collection, and analysis.	Measure 3.6. Biannual memo summarizing revisions made to materials and procedures for Objective 3.
Objective 4. Develop sample and infrastructure for continued scaling of MTP-S.		

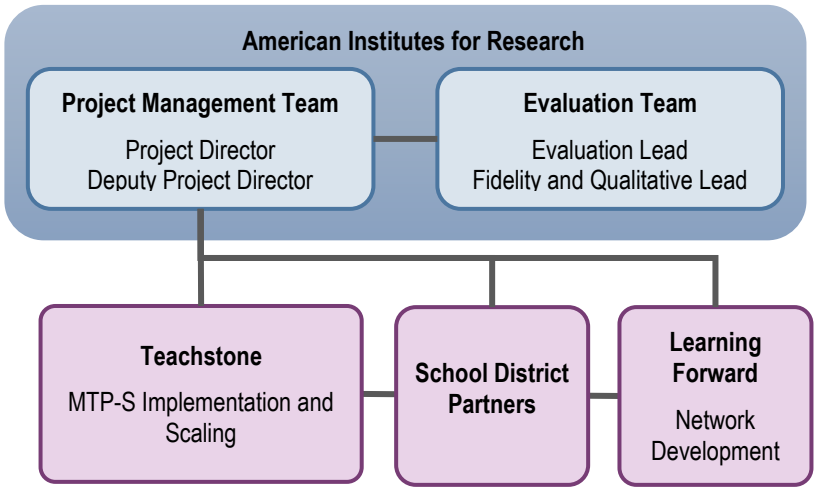
Strategies	Outcomes	Measures
Strategy 4.1. Identify more districts if needed.	Commitment of districts to participate in project.	Measure 4.1. Signed MOUs from additional districts.
Strategy 4.2. Establish and operate network to support and sustain work of participating districts and other districts with interest in MTP-S (Yrs 3, 4, 5).	A network that supports MTP-S users and potential users.	Measure 4.2. Network website; network convening materials; network roster of network participants.
Strategy 4.3. Develop and refine approach to help districts integrate MTP-S into existing systems of support (Yrs 1 & 3).	Tools and procedures for implementing MTP-S as part of coherent system of teacher support.	Measure 4.3. Committee agenda; Summary of committee input.

C.2. A Management Plan That Defines Responsibilities, Timelines, and Milestones

The management plan establishes the reporting relationships for the partner organizations (see Exhibit 5). The plan is **more than adequate** in part because each partner organization is highly qualified for a clear and specific role involving **execution of the strategies** (see Section C.1) **at each milestone on the project's 5-year timeline** (see Exhibit 7).

AIR is the lead organization for the project, responsible to the U.S. Department of Education for grant performance. AIR's role is to (1) oversee the subgrants to Teachstone, the SDPs, and Learning Forward, ensuring coordination across the partners to achieve the project objectives; (2) recruit eligible teachers

Exhibit 5. Organizational Chart



from the SDPs; and (3) conduct the independent evaluation. The AIR evaluation team will be separate from AIR's role overseeing the study partners; the evaluation team will have no role in the implementation of MTP-S or the strategy to scale except to provide implementation data as feedback. This structure meets a higher standard for independence than was typical for Investing in Innovation grants and is the model the Institute of Education Sciences (IES) requires research contractors to use in similar projects involving RCTs and an intervention provider. AIR is uniquely qualified for this role, having successfully led four such projects for IES in the last decade focused on teacher PD interventions, coordinating across subcontracted organizations, including an intervention provider and several school districts (see Exhibit 6). AIR's experience monitoring intervention providers and providing feedback on fidelity for continuous improvement also helps ensure relevant, actionable feedback from the evaluation team, which will draw on instruments and methods that AIR has refined across several studies. ***AIR Lead Staff:*** Andrew Wayne, Project Director (PD); Sarah Caverly, Deputy Project Director (DPD); Mengli Song, Evaluation Lead (EL); Jane Coggshall, Evaluation Lead for Fidelity & Qualitative.

Exhibit 6. AIR-Led IES Contracts to Evaluate Teacher PD Interventions

Project Name	Subcontractor(s) That Provided the Intervention	Number of Districts and Schools
Focusing on Mathematical Knowledge: The Impact of Content-Intensive Teacher PD	Intel Math (summer institute), Mathematics Learning Community along with district-based coaches (school-year meetings), Harvard University along with district-based coaches (video-based coaching)	6 districts 73 schools
The Impact of Providing Performance Feedback to Teachers and Principals	Danielson Group, Teachscape, Teachstone, University of Virginia, Discovery Education (performance feedback)	8 districts 127 schools
Middle School Mathematics Professional Development Impact Study	America's Choice & Pearson Achievement Solutions (summer institute, school-year meetings, coaching)	12 districts 77 schools

Project Name	Subcontractor(s) That Provided the Intervention	Number of Districts and Schools
The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement	Language Essentials for Teachers of Reading and Spelling (summer institute/school-year meetings), district-based staff trained by the Consortium on Reading Excellence (coaching)	6 districts 90 schools

Teachstone will (1) lead the implementation of the strategy to scale and (2) provide all the necessary supervision, training, tools, and support to enable the SDPs to provide MTP-S to the treatment group teachers. Teachstone is uniquely qualified for this role as the licensed vendor for MTP-S and other services based on CLASS-S. To date, more than 500 professionals have been trained as MTP coaches, and more than 3,000 teachers have participated in MTP. Teachstone employs more than 90 professionals, 10 of whom focus on content development and design, which allows the organization to continuously improve its products and services. ***Teachstone Lead Staff:*** Victoria Kintner-Duffy, CF; Robert Pianta, Quality Assurance (QA).

Learning Forward (LF) will (1) lead outreach to identify any additional districts needed to provide schools for Cohort 3, (2) build and sustain a national network of active and prospective MTP-S users, and (3) advise Teachstone on methods and tools for ensuring that districts adopting MTP-S integrate it coherently with other programs. These activities leverage LF's unparalleled state, regional, and national networks of educators, central office leaders, consultants, and others focused on improving teacher PD (see learningforward.org). ***Learning Forward Lead Staff:*** Tom Manning, Project Leader (PL).

The SDPs will be supported by the other partner organizations to (1) recruit and select coaches who will be trained, (2) implement MTP-S, and (3) participate in all data collections.

Exhibit 7. Group Responsible, Time Frame, and Milestones for Each Strategy

		Project Year (October 1–September 30)				
Milestones	**Responsible**	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Objective 1. Implement strategy to scale while continuously using feedback and fidelity data for project improvement.						
Strategy 1.1	Teachstone CF, SDP	√	√	√		
Strategy 1.2	Teachstone CF, SDP	√	√	√		
Strategy 1.3	Teachstone CF, SDP	√	√	√		
Strategy 1.4	Teachstone CF, SDP		√	√	√	
Strategy 1.5	Teachstone CF	√	√	√	√	
Objective 2. Implement MTP-S while continuously using feedback and fidelity data for project improvement.						
Strategy 2.1	Teachstone CF, SDP	√	√	√	√	
Strategy 2.2	SDP		√	√	√	√
Strategy 2.3	Teachstone CF		√	√	√	√
Objective 3. Conduct an RCT to test the impact of MTP-S provided by trained local coaches on (a) the quality of classroom interactions and (b) student outcomes.						
Strategy 3.1	AIR Evaluation lead, SDP	√	√	√		
Strategy 3.2	AIR Evaluation lead, SDP	√	√	√		
Strategy 3.3	AIR Evaluation lead		√	√	√	√
Strategy 3.4	AIR Evaluation lead		√	√	√	
Strategy 3.5	AIR Evaluation lead		√	√	√	√
Strategy 3.6	AIR Evaluation lead		√	√	√	√
Objective 4. Develop sample and infrastructure for continued scaling of MTP-S.						
Strategy 4.1	LF PL, Teachstone CF, AIR DPD			√		
Strategy 4.2	LF PL, Teachstone CF			√	√	√
Strategy 4.3	LF PL, Teachstone CF, AIR DPD	√		√		

C.3. Procedures Ensuring Feedback and Continuous Improvement

The project is designed to ensure **feedback and continuous improvement** through (1) its sequenced cohort structure and (2) routines for using feedback. Each cohort's participation will create **feedback** that informs real-time improvements or improvements for the next cohort. The routines for gathering feedback and deciding on improvements are integrated into the meetings planned for the operation of the project, as shown in Exhibit 8. Each meeting will include a

regular agenda item **to discuss feedback and implications for improving the project's materials, strategies, and procedures**. For example, during a *Monthly implementation review meeting* just after the launch of Cohort 1, the Teachstone CF may report that all coaches attended the *Biweekly coaching quality meeting* (part of Strategy 1.4) but only half were engaged. The review team could decide that the CF should use the *Biweekly one-on-one cycle reviews* to clarify expectations and ask each disengaged coach about making the *Biweekly coaching quality meetings* more engaging. The CF could then try modifying the meeting content or format as suggested. If that is successful, the review team would discuss implications for future instances of the *Biweekly coaching quality meetings* and for materials for future implementations.

Exhibit 8. Routines for Using Feedback

Meeting Name and Frequency (When Active)	Participants	Feedback Data Sources	Strategies to be Improved
Monthly implementation reviews	AIR: PD, DPD, EL Teachstone: CF, QA	All fidelity data collections (see Appendix G)	1.1, 1.2, 1.3, 1.4, 2.1, 2.2
Biweekly coaching quality meetings	Teachstone: CF SDP: Coaches	Coach confidence survey; coach-teacher relationship survey; coach logs; MTP-S FOI checklist; number of cycles completed	2.2
Biweekly one-on-one "cycle reviews"	Teachstone: CF SDP: Coach	MTP-S FOI checklist; number of cycles completed	1.4
Monthly partnering network team meetings	Learning Forward: PL Teachstone: CF AIR: PD	Partner engagement records	4.1, 4.2
Weekly evaluation team meetings	AIR: EL & evaluation team	Data collection and analysis update memo	3.1, 3.2, 3.3, 3.4, 3.5
Bimonthly evaluation reviews	AIR: Evaluation lead, PD, Vice P.	Data collection and analysis update memo	3.1, 3.2, 3.3, 3.4, 3.5
Monthly individual SDP point-of-contact check-ins (separately by district)	AIR Evaluation lead, SDP point-of-contact	Data collection update	3.1, 3.2, 3.3, 3.4, 3.5

Meeting Name and Frequency (When Active)	Participants	Feedback Data Sources	Strategies to be Improved
Biannual Teachstone design team meetings	Teachstone: CF, QA	All fidelity data collections (see Appx. G)	1.1, 1.2, 1.3, 1.4, 2.1, 2.2

C.4. Ongoing Work Beyond the End of the Grant

The project's benefits will continue into the future through the partners, who each see the work as aligned to their mission. Teachstone wants MTP-S and all its services to be effective, affordable, and widely used. **Teachstone plans** to use the project's final materials and procedures in future implementations. Learning Forward's mission includes helping states and districts identify scalable, effective PD that can be integrated into a coherent system of teacher support. **Learning Forward plans** to continue the network with support from Teachstone. The **SDPs** will have trained, experienced, local MTP-S coaches and resources earmarked for teacher PD that can support **ongoing delivery** of MTP-S. In addition, the cadre of teachers in each SDP, trained through the project, have the **potential** to have an impact on new classes of students every year. To maximize impact during and after the project **AIR plans** to seek foundation support for delivery of MTP-S via the scaling strategy to the control teachers in each cohort, delayed two years from the treatment group. **AIR** also plans to identify broader lessons from the evaluation to inform the development and enhancement of other PD programs. Finally, AIR will continue to pursue opportunities to build partnerships to scale, refine, and test teacher PD for high-need students, consistent with its mission to conduct and apply the best-possible research toward improving people's lives with a special emphasis on the disadvantaged.

D. QUALITY OF THE PROJECT EVALUATION

D.1. Evaluation Methods Designed to Meet WWC Evidence Standards Without Reservations

AIR will conduct an independent evaluation of MTP-S, as delivered using the scaling strategy, to address six research questions (RQs) about impact and implementation fidelity:

(RQ1) What is the impact of MTP-S on the quality of teachers' interactions with students? (RQ2) What is the impact of MTP-S on student engagement and academic achievement? (RQ3) Is the impact of MTP-S on student achievement moderated by student, teacher/classroom, and school characteristics? (RQ4) Is the impact of MTP-S on student achievement mediated by the quality of teachers' interaction with students? (RQ5) Are the scaling strategy and MTP-S implemented with fidelity? (RQ6) What are the obstacles and success factors to the scaling of MTP-S?

These six questions will be addressed through a blocked cluster RCT, with teachers randomly assigned to the treatment or control condition within blocks defined by subject (mathematics or English language arts [ELA]) and school. Teachers in the treatment and control conditions will be subject to the districts' normal PD requirements and opportunities, but treatment teachers will additionally participate in MTP-S. Teachers are the appropriate unit of assignment because the MTP-S activities (e.g., one-on-one meetings between the teacher and coach) are carried out by each teacher individually and because teacher-level random assignment has a clear statistical power advantage over school-level randomization.

One commonly voiced concern about within-school teacher-level random assignment is the risk of contamination, which occurs when some control teachers also receive some of the treatment. Although contamination will not affect the internal validity of impact estimates based on intent-to-treat analyses, it may reduce the service contrast and lead to an underestimate of the treatment effect. For the proposed project, the amount of contamination is likely to be limited, for several reasons. First, when the districts select staff to serve as MTP-S coaches (see Section B), staff whose role includes coaching mathematics or ELA teachers in the participating high-poverty and high-minority schools will be considered ineligible to ensure there is no interaction between the MTP-S coaches and control teachers. Second, most materials used for

MTP-S coaching will be stored in a secure, password-protected website that cannot be accessed by control teachers. Third, to minimize contamination, the treatment teachers will be instructed not to share what they learn from the MTP-S program with other teachers in their school before the study ends. No evidence of contamination has been found in prior studies of MTP-S (Allen et al., 2011; Allen et al., 2015) or in a recent AIR study of a teacher PD program using a similar design (Garet et al., 2016). Nevertheless, we will check for evidence of contamination by including teacher survey items about the sharing of MTP-S-related information between treatment and control teachers (as we did in Garet et al., 2016).

Based on a statistical power analysis (see Appendix G.3 for details), the sample for the proposed RCT will include 150 mathematics teachers and 150 ELA teachers who meet the following eligibility criteria: (1) teach in a regular high-poverty or high-minority middle or high school, as defined by the state's Teacher Equity Plan, (2) teach a regular yearlong mathematics or ELA class with an end-of-course (EOC) examination in Grades 6–12; and (3) do not participate in an induction program involving regular mentoring. Assuming an average of six participating teachers (three per subject) per school, the sample will include approximately 33–34 schools, which will participate in the evaluation in three cohorts. Treatment teachers in Cohort 1 schools will start to receive the 2-year intervention in 2018–19, and treatment teachers in Cohorts 2 and 3 schools will start in 2019–20 and 2020–21, respectively (see Introduction).

For the proposed RCT, the main threat to internal validity is potential selection bias resulting from sample attrition during the 2 years of the intervention. We plan to use multiple strategies to minimize attrition. First, during recruitment, we will advertise that only teachers who plan to stay in their school during the two intervention years should volunteer. Second, we will carry out random assignment as late as possible—at the beginning (fall) of the first intervention year, rather

than the spring of the prior school year as is often the case with impact studies of PD programs. This will eliminate attrition resulting from teacher turnover during the summer before the intervention starts. Given that the proposed evaluation is based on an RCT that is free of confounding factors and is expected to demonstrate baseline equivalence with low attrition, it is expected to produce strong evidence about MTP-S's effectiveness that will meet the WWC evidence standards without reservations.

D.2. Generation of Guidance About Effective Strategies Suitable for Replication

The proposed evaluation will generate useful guidance about effective strategies for implementing and scaling MTP-S in diverse settings by (a) including a large sample representing diverse settings, including suburban settings, where the program has not yet been tested; (b) deliberately assessing whether the impact of MTP-S differs for different types of students, teachers, classrooms, and schools (i.e., moderators); and (c) analyzing data on program implementation in different settings, as detailed next. The evaluation also will include a cost analysis, which will provide valuable information about the cost effectiveness of the program.

(a) Diverse Settings. The high-need schools included in this project will come from districts representing diverse settings. The commitment of five district partners gives us flexibility to define each cohort based on urbanicity and potentially other variables considered important for generating guidance for future replications, as explained in Section B.1. (See Appendix D for letters of support and Appendix G.4 for the numbers of schools considered high-poverty and high-minority.)

(b) Differential Impact Analyses. The proposed evaluation will generate useful guidance about the relative effectiveness of MTP-S for different types of students and settings through differential impact analyses, which will assess the extent to which the impact of the program is

moderated by the characteristics of students, teachers/classrooms, and schools (see Exhibit 9).[1] These results will be crucial in guiding later efforts to scale MTP-S, as they may identify settings and populations for which the program is not well suited.

Exhibit 9. Potential Moderators at the Student, Teacher, and School Levels

Student-Level Moderators	Teacher-Level Moderators	School-Level Moderators
Race/ethnicity, eligibility for free or reduced-price lunch, English language learner status, prior achievement scores	Teacher experience, class size, and classroom average prior achievement, and subject	School level, school size, locale (urban/suburban/rural), and demographic composition (e.g., percentage of minority/low-income students)

(c) Analyses of Implementation Data From Multiple Sources. To provide guidance and lessons learned for future replication or testing of MTP-S in other settings, using Teachstone's strategy to scale, the evaluation team will collect and analyze implementation data from multiple sources throughout the two intervention years. In addition to implementation-related information tracked by the MTP-S online system (e.g., level of participation), we will examine implementation fidelity (RQ5) based on data from coach surveys, coach logs, implementation fidelity checklists completed for a random sample of coaching cycles, and teacher surveys.[2] Analysis of the fidelity of implementation of the scaling strategy will focus on issues such as coach selection and coach training, monitoring, and support. Analysis of implementation fidelity of the MTP-S coaching itself will focus on the dosage and quality of coaching received.

[1] In addition to differential impact analyses, we also will estimate MTP-S's impact within each key student subgroup separately, particularly subgroups of high-need students (e.g., minority students and low-income students).

[2] The teacher survey will be administered to both treatment and control teachers each spring, which will allow us to gather data on control teacher's coaching experience as well to assess "service contrast."

To further explore implementation-related issues, we will conduct phone interviews with all coaches each fall and spring to gather information about their experiences with (1) the training and support they receive as MTP-S coaches and (2) the delivery of coaching to teachers. We also will interview the CF each spring on the CF's experience in providing training and support to coaches. These interviews will pay particular attention to factors that facilitate or hinder the implementation of the scaling strategy and the MTP-S program (RQ6) and examine issues such as whether a sufficient pool of promising candidates exists for the coach positions, whether coach training and support is adequate, and how well the online MTP-S system functions. (See Appendix G.5 for a description of the different types of implementation data to be collected.)

Cost Analysis. To provide information about whether MTP-S is a cost-effective investment, we will conduct a cost analysis using the Resource Cost Model (RCM), which has been used extensively by AIR.[3] Focusing on both personnel and nonpersonnel resources used in the MTP-S program, we will populate the RCM using the *CostOut* tool and generate cost-effectiveness estimates based on the cost estimates and results from the impact analyses.[4]

D.3. Valid and Reliable Performance Data on Relevant Outcomes

Teacher Outcomes. According to the theory of change presented in Exhibit 3, the primary teacher outcome for the MTP-S program is the quality of teachers' interactions with students, which will be measured through video observations of classroom instruction coded using CLASS-S. For each study teacher, we plan to video-record one lesson in early fall of the first intervention year (as baseline), two lessons in the spring of the first year, and two lessons in the spring of the second year. For teachers teaching multiple sections, we will randomly select one

[3] See http://www.air.org/topic/p-12-education-and-social-development/school-finance.

[4] A tool created by the Center for Benefit-Cost Studies in Education at Teachers College, Columbia University.

section for the fall observation and two for the spring observations. The video-recoded lessons will be coded by certified CLASS-S observers at AIR. A subset (10%) of lessons will be double-coded by independent coders for reliability check. For teacher outcomes, we will use the CLASS-S overall score as the primary measure. The CLASS-S also provides three domain scores, which we will use as secondary measures (i.e., *Emotional Support, Classroom Organization,* and *Instructional Support*). To check the robustness of findings about impact on teacher outcomes not as directly aligned with the intervention, we plan to code video-recorded lessons from the spring Year 2 observations using an additional instrument: Charlotte Danielson's Framework for Teaching (FFT). The FFT is a non-subject-specific classroom observation instrument that focuses on dimensions of teacher practice similar to CLASS-S and has evidence of validity (Bill & Melinda Gates Foundation, 2012; Goe, Bell, & Little, 2008). (See Appendix G.6 for domains and dimensions measured by CLASS-S and FFT.).

Student Outcomes: Engagement and Achievement. Student engagement, which is hypothesized to be a key mediator in the theory of change, will be measured at both classroom and student levels. At the classroom level, we will use the CLASS-S *Student Engagement* dimension score from the classroom observations described previously. It captures the degree to which all students in the class are focused and participating. At the student level, we will administer a short survey in the fall of the first intervention year (as baseline) and each spring to measure the quality and nature of students' participation in learning activities based on measures developed by Skinner, Furrer, Marchand, and Kindermann (2008), and Skinner, Kindermann, and Furrer (2009). The survey will include items associated with four engagement scales with alpha reliability ranging from 0.70 to 0.79 (see Appendix G.7). The survey sample will include students from one randomly selected section taught by each study teacher in the fall of each year.

Our primary measures of student achievement are students' scores on state EOC examinations in mathematics and ELA, which are available in our partner districts for students in the target grades (6–12) for the study and will be requested along with student demographic data from the districts. To supplement the EOC achievement measures, we also will examine students' performance on the state end-of-grade tests in mathematics and ELA for students taking these tests (e.g., students in grades 6-8 and certain high school grades). Students from all sections of mathematics and ELA taught by study teachers in the fall of each intervention year will be included in the impact analyses. All our measures of student achievement will be based on statewide standardized tests and thus expected to have sufficient validity and reliability.

D.4. Clear Articulation of Components, Mediators, and Outcomes and Measurable Threshold

The design of the proposed evaluation is informed by clearly articulated key components, mediators, and outcomes of the MTP-S program as depicted in the theory of change presented in Exhibit 3. As Exhibit 3 shows, the central component of the MTP-S intervention is a coaching process designed based on core principles for coaching. The coaching process involves a set of carefully planned, manualized activities for both coaches and teachers. The theory of change also specifies key outcomes for teachers (i.e., quality of teacher-student interactions) and students (i.e., engagement and achievement) as detailed in Section D.3. The quality of teacher-student interactions serves also as a key mediator for the impact of the MTP-S on student outcomes.

In addition to outcome data, we will collect implementation data from multiple sources (see Appendix G.5). The **measurable threshold for acceptable implementation** of the scaling strategy is that a coach must complete the 5-day in-person coach training and participate in at least two thirds of the biweekly one-on-one meetings with the CF and at least two thirds of the biweekly team meetings led by the CF with all coaches. The **measurable threshold for**

acceptable implementation of the MTP-S coaching is that: (1) a teacher must complete the half-day workshop on CLASS-S each year and complete at least 12 (six per year) of the planned coaching cycles; and (2) a coach must demonstrate successful implementation of at least 90% of the key elements of MTP-S based on the fidelity of implementation checklist completed for a randomly selected sample of coaching cycles.

D.5. Clearly Specified Analytic Approach to Addressing the Research Questions

Our main impact analyses will be fixed-effects intent-to-treat analyses. Analyses of impact on teacher outcomes (RQ1) will be based on a two-level model that takes into account the clustering of lessons within teachers; analyses of impact on student outcomes (RQ2) will be based on a three-level model that takes into account the clustering of students within sections and sections within teachers. Both impact models will incorporate school fixed effects and covariates at the student and teacher levels, as appropriate. We will conduct analyses modeling schools as random effects as sensitivity analyses. We will adapt the main student impact model to assess the differential impact of MTP-S (RQ3) by incorporating a treatment-by-moderator interaction term, where the moderator is a characteristic of the student, teacher/classroom, or school, as explained in Section D.2.

To assess the extent to which the impact of MTP-S on student achievement is mediated by the quality of teacher-student interactions (RQ4), we will use two alternative models: (1) a single-mediator model that estimates the overall mediating effect of the quality of teacher-student interactions, as measured by the CLASS-S overall score (averaged across both intervention years), and (2) a multiple-mediator model that estimates the unique mediating effect of each CLASS-S domain. (See Appendix G.8 for technical details of all analytic models.) Finally, we will conduct descriptive analyses of implementation fidelity data and qualitative analyses of coach and CF interview data to address RQ5 and RQ6 (see Section D.2 for details).

REFERENCES

Allen, J. P., Gregory, A., Mikami, A. Y., Lun, J., Hamre, B. K., & Pianta, R. C. (2013). Observations of effective teaching in secondary school classrooms: Predicting student achievement with the CLASS-S. *School Psychology Review, 42*(1), 76–98.

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.

Allen, J. P., Hauser, S. T., Bell, K. L., & O'Connor, T. G. (1994). Longitudinal assessment of autonomy and relatedness in adolescent-family interactions as predictors of adolescent ego development and self-esteem. *Child Development, 65*(1), 179–194.

Allen, J. P., Kuperminc, G., Philliber, S., & Herre, K. (1994). Programmatic prevention of adolescent problem behaviors: The role of autonomy, relatedness, and volunteer service in the teen outreach program. *American Journal of Community Psychology, 22*(5), 617–638.

Allen, J. P., Marsh, P., McFarland, C., McElhaney, K. B., Land, D. J., Jodl, K. M., & Peck, S. (2002). Attachment and autonomy as predictors of the development of social skills and delinquency during midadolescence. *Journal of Consulting & Clinical Psychology, 70*(1), 56–66.

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.

Atwater, L., Brett, J., & Charles A. (2007). Multisource feedback: lessons learned and implications for practice. *Human Resource Management, 46*(2), 285–307.

Bill & Melinda Gates Foundation. (2012). *Gathering Feedback for Teaching: Combining High-Quality Observations With Student Surveys and Achievement Gains*. Seattle, WA: Author. Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf.

Blazar, D., Kraft, M. A., Hogan, D. (2017). *The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence.* Paper presented at the spring 2017 conference of the Society for Research on Educational Effectiveness, Washington, DC.

Brutus, S. (2010). Words versus numbers: A theoretical exploration of giving and receiving narrative comments in performance appraisal. *Human Resource Management Review, 20*(2), 144–157.

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. Retrieved from http://www.pnas.org/content/110/43/17176.full

Chen, X., & Simone, S. (2016). *Remedial coursetaking at U.S. public 2- and 4-year institutions: Scope, experience, and outcomes statistical analysis report.* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED568682.pdf

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.

Danielson, C. (2008). *The handbook for enhancing professional practice: Using the framework for teaching in your school.* Alexandria, VA: Association for Supervision and Curriculum Development.

Desimone, L. M., & Garet, M. S. (2015). Best practices in teacher's professional development in the United States. *Psychology, Society, & Education, 7*(3), 252–263.

Desimone, L., Smith, T. M., & Phillips, K. J. R. (2013). Linking student achievement growth to professional development participation and changes in instruction: A longitudinal study of elementary students and teachers in Title I schools. *Teachers College Record, 115*(5), 1–46.

Downer, J. T., Pianta, R. C., Burchinal, M., Field, S., Hamre, B. K., & Scott-Little, C. (2014). *Coaching and coursework focused on teacher-child interactions during language/literacy instruction: Effects on teacher outcomes and children's classroom engagement.* Manuscript submitted for publication.

The Education Trust. (2017). *Achievement and Opportunity in America: Where are We? What Can We Do?* Retrieved from https://edtrust.org/resource/achievement-opportunity-america-can-2/

Funk, S., LoCasale-Crouch, J., & Hamre, B. K. (2009). *Consultant rating of teacher* (Unpublished measure). University of Virginia.

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., Garrett, R., Yang, R., & Borman, G. D. (2016). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development* (NCEE 2016-4010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Garet, M. S., Heppen, J. B., Walters, K., Smith, T.M., & Yang, R. (2016). *Does content-focused teacher professional development work? Findings from Three Institute of Education Sciences studies* (NCEE 2017-4010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from https://ies.ed.gov/ncee/pubs/20174010/pdf/20174010.pdf

Garet, M., Porter, A., Desimone, L., Birman, B., & Yoon, K. (2001). What makes professional development effective? Analysis of a national sample of teachers. *American Education Research Journal, 38*(3), 915–945.

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

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). Summary of research on the effectiveness of math professional development approaches (REL 2014–010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Education Research Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2014010.pdf

Glover, T. A., Nugent, G. C., Chumney, F. L., Ihlo, T., Shapiro, E. S., Guard, K, Bovaird, J. (2016). Investigating rural teachers' professional development, instructional knowledge, and classroom practice. *Journal of Research in Rural Education, 31*(3), 1–16.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis.* National Comprehensive Center for Teacher Quality.

Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2013). Effects of a professional development program on behavioral engagement of students in middle and high school. *Psychology in Schools, 51*(2), 143–163.

Hasiotis, D., Jacobs, A., & McGovern, K. (2015). *The mirage: Confronting the hard truth about our quest for teacher development.* Washington, DC: The New Teacher Project. Retrieved from https://tntp.org/publications/view/evaluation-and-development/the-mirage-confronting-the-truth-about-our-quest-for-teacher-development

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J. L., Downer, J. T., Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49*(1), 88–123.

Kennedy, M. M. (In press). How does professional development improve teaching? *Review of Educational Research.*

Li, Y., & Lerner, R. M. (2011). Trajectories of school engagement during adolescence: Implications for grades, depression, delinquency, and substance use. *Developmental Psychology, 47,* 233–247.

LoCasale-Crouch, J., Downer, J. T., & Hamre, B. K. (2009a). *Coach confidence* (Unpublished measure). University of Virginia.

LoCasale-Crouch, J., Downer, J. T., & Hamre, B. K. (2009b). *Perceptions of professional development* (Unpublished measure). University of Virginia.

New Teacher Center. (2016). *High quality mentoring and induction practices.* Chicago, IL: Author. Retrieved from https://newteachercenter.org/wp-content/uploads/high-quality-mentoring_induction-resource.pdf

Pianta, R. (2005). Standardized observation and professional development: A focus on individualized implementation and practices. In M. Zaslow & I. Martinez-Beck (Eds.), *Critical issues in early childhood professional development* (pp. 231–254). Baltimore, MD: Brookes.

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

Rowan, B., & Raudenbush, S. (2016). Teacher evaluation in American schools. In D. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed.; pp. 1159–1216). Washington, DC: American Educational Research Association.

Ryan, M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.

Salmela-Aro, K. (2017). Dark and bright sides of thriving—School burnout and engagement in the Finnish context. *European Journal of Developmental Psychology*, *14*, 337–349. doi:10.1080/17405629.2016.1207517

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation. Research Report.* Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology, 100*(4), 765–781.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*, 493–525.

Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology, 58*(1), 33-66.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Steinberg, M., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy, 10*(4), 535–572.

Taylor, E., & Tyler, J. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628–3651.

U.S. Department of Education. (2016*). The nation's report card: 2015 mathematics & reading at Grade 12.* Washington, DC: Author. Retrieved from https://www.nationsreportcard.gov/reading_math_g12_2015/#reading/acl

Voisin, D. R., & Elsaesser, C. (2016). Brief report: The protective effects of school engagement for African American adolescent males. *Journal of Health Psychology, 21*, 573–576. doi:10.1177/1359105314531607

Wayne, A. J., Yoon, K. W., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher, 37*, 469–279.

What Works Clearinghouse. (2012). *WWC quick review of the report "An interaction-based approach to enhancing secondary school instruction and student achievement."* Washington, DC: Author. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/quick_reviews/myteachingpartner_022212.pdf

What Works Clearinghouse. (2015). *WWC intervention report "MyTeachingPartner-Secondary."* Washington, DC: Author. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_mtp_061615.pdf