

Teacher Professional Learning in Grade-Level Teams: The Implementation and Impact of a Program Delivered During the COVID-19 Pandemic

Final Evaluation Report

Kyle Neering, Cheryl Graczewski, Kristina Zeiser, Heleana Lally, Dory Seidel, Andrew Wayne, and Mengli Song

American Institutes for Research

JULY 2024



Contents

Acknowledgements.....	vii
Abstract.....	viii
Introduction	1
A Project to Improve, Test, and Scale a Promising Teacher PL Program	1
Research Questions	2
Program Overview	3
Strategy to Implement PLI at Scale	4
PLI Program and Its Components	5
Evaluation Overview	7
Implementation Evaluation	8
Measures and Data Sources	8
Analytic Samples for the Implementation Analysis	10
Findings	12
Impact Evaluation	28
Design of the Impact Evaluation	28
Findings	38
Cost Study	42
Introduction to the Cost Study	42
Total Cost of PLI	46
Cost-Effectiveness of PLI	53
Discussion.....	54
Limitations	54
Findings About the Implementation of the PLI Program.....	54
Findings About Impact	55
Through what mechanism did the program have an impact on student achievement?	56
Is the program scalable?	57
References	58
Appendix A. Additional Analytic Samples for the Implementation Analysis and Implementation Findings for All Four Analytic Samples.....	61

Appendix B. Additional Implementation Findings	68
Appendix C. Scoring That Defines Adequate Implementation of Each Key Component, by Indicator	70
Appendix D. Descriptive Information About the Sample of Schools Included in the Randomized Controlled Trial	79
Appendix E. Assessment of the Measurement Properties of Teacher Survey Scales	82
Appendix F. Attrition and Baseline Equivalence for Outcome Measures in Program Year 2 and Program Year 1	91
Appendix G. Impact of Professional Learning with Impact on Supplemental Outcomes From the Teacher Survey in Program Year 2 and Program Year 1	97
Appendix H. Impact of Professional Learning with Impact (PLI) on Supplemental Student Outcomes	99
Appendix I. Additional Information About the Cost Study	102

Exhibits

Exhibit 1. Professional Learning with Impact Theory of Change	3
Exhibit 2. Key Components of the Strategy to Scale the Professional Learning with Impact (PLI) Program	4
Exhibit 3. Key Components of the Professional Learning with Impact (PLI) Program.....	5
Exhibit 4. Five Stages of the Professional Learning with Impact Teacher Learning Team Cycles.....	6
Exhibit 5. Timeline of Professional Learning with Impact (PLI) Implementation, by Cohort	7
Exhibit 6. Implementation Evaluation Measures and Data Sources, by Component	9
Exhibit 7. Primary Analytic Samples for Findings About Implementation.....	10
Exhibit 8. Scoring of Adequate Implementation of Each Key Program Component, by Year	12
Exhibit 9. Participation in the Preparatory Teacher Workshop.....	13
Exhibit 10. Participation in the Teacher Learning Team Cycles.....	14
Exhibit 11. Participation in Individual Instructional Coaching	16
Exhibit 12. Participation in Direct Support for Teacher Learning Teams (TLTs).....	17
Exhibit 13. Scoring of Adequate Implementation of Each Key Strategy-to-Scale Component, by Year	19
Exhibit 14. Participation in Guidance for District Management of Professional Learning with Impact (PLI) Implementation	20
Exhibit 15. Participation in Professional Learning with Impact (PLI) Program Protocols and Guidance Documents.....	21
Exhibit 16. Participation in Professional Learning with Impact (PLI) Coach Workshops.....	21
Exhibit 17. Participation in Ongoing Professional Learning with Impact (PLI) Coach Support.....	23
Exhibit 18. Variation by District in Year 2 Principal Support for Common Meeting Time, Average Number of Learning Team Cycles Completed, and Average Number of One-on-One Meetings.....	26
Exhibit 19. Variation by School in Year 2 Principal Support for Common Meeting Time, Average Number of Learning Team Cycles Completed, and Average Number of One-on-One Meetings.....	27

Exhibit 20. Inter-Rater Reliabilities of Measures of Classroom Practice	30
Exhibit 21. Sample Sizes and Attrition Rates at the School and Teacher Levels for Program Year 2 Teacher Outcomes.....	34
Exhibit 22. Impacts of Professional Learning with Impact on Teacher Outcomes in Year 2	39
Exhibit 23. Impacts of Professional Learning with Impact on Fourth-Grade English Language Arts and Math Achievement in Year 2	40
Exhibit 24. Differential Impacts of Professional Learning with Impact (PLI) on Fourth-Grade English Language Arts and Math Achievement in Year 2, by Student Characteristics	40
Exhibit 25. Professional Learning with Impact Program Components	43
Exhibit 26. Upper-Bound Overall Cost Per Teacher of the 2-Year Professional Learning with Impact Program (2023 Dollars).....	47
Exhibit 27. Overall Per-Teacher Cost of Professional Learning with Impact in Year 1 and Year 2 (in 2023 Dollars).....	48
Exhibit 28. Lower-Bound Overall Cost Per Teacher of the 2-Year Professional Learning with Impact Program (2023 Dollars).....	51
Exhibit 29. Lower-Bound Overall Per-Teacher Cost of Professional Learning with Impact in Year 1 and Year 2 (in 2023 Dollars).....	52
Exhibit A1. The Four Analytic Samples for Findings About Implementation.....	62
Exhibit A2. Participation in the Preparatory Teacher Workshop, Teacher Learning Team (TLTs) Cycles, Individual Instructional Coaching, and Direct Support for TLTs.....	62
Exhibit A3. Scoring of Adequate Implementation of Each Key Program Component, by Year	64
Exhibit A4. Districts’ and Coaches’ Participation in the Strategy-to-Scale Components	65
Exhibit A5. Scoring of Adequate Implementation of Each Key Strategy-to-Scale Component, by Year	66
Exhibit B1. Variation by School in Year 2 Teacher Learning Team (TLT) Student and Teacher Learning Goals.....	68
Exhibit D1. Characteristics of Professional Learning with Impact (PLI) and Business-as-Usual (BAU) Schools.....	79
Exhibit E1. Survey Items Associated With Teacher Survey Outcome Measures, Their Corresponding Factor Loadings, and Values of Cronbach’s Alpha	84

Exhibit E2. Multigroup Confirmatory Factor Analysis Nested Model Comparisons for All Teacher Survey Outcome Measures.....	89
Exhibit E3. Correlation Matrix of Scale Means	90
Exhibit F1. Attrition for Teacher Survey Outcome Measures in Program Year 2 and Program Year 1	91
Exhibit F2. Baseline Observation Scores and Survey Measures of Teachers Included in the Analyses of Year 2 Impact on Teacher Outcomes, by Study Group	92
Exhibit F3. Baseline Equivalence of Survey Outcomes for Teachers Included in the Analyses of Year 2 Impact on Supplemental Survey Outcomes, by Study Group	93
Exhibit F4. Baseline Equivalence of Student Characteristics for English Language Arts and Math Analytical Samples from the Year 2 Achievement Impact Analysis.....	93
Exhibit F5. Baseline Equivalence of Student Characteristics for English Language Arts and Math Analytical Samples from the Year 2 Achievement Impact Analysis, Using All Four Districts	94
Exhibit F6. Baseline Equivalence of Student Characteristics for English Language Arts and Math Analytical Samples from the Year 1 Achievement Impact Analysis.....	96
Exhibit G1. The Effect of Professional Learning with Impact (PLI) on Supplemental Outcomes From the Teacher Survey in Program Year 2.....	97
Exhibit G2. The Effect of Professional Learning with Impact (PLI) on Supplemental Outcomes From the Teacher Survey in Program Year 1.....	98
Exhibit H1. Impacts of Professional Learning with Impact on Fourth-Grade English Language Arts and Math Achievement in Year 2, by English Learner Status	99
Exhibit H2. Impacts of Professional Learning with Impact on Fourth-Grade English Language Arts and Math Achievement in Year 2	99
Exhibit H3. Differential Impacts of Professional Learning with Impact (PLI) on Fourth-Grade English Language Arts and Math Achievement in Year 2, by Student Characteristics	100
Exhibit H4. Impacts of Professional Learning with Impact on Fourth-Grade English Language Arts and Math Achievement in Year 1	101
Exhibit H5. Differential Impacts of Professional Learning with Impact (PLI) on Fourth-Grade English Language Arts and Math Achievement in Year 1, by Student Characteristics	101
Exhibit I1. Compensation Rates for Personnel (in 2023 Dollars).....	103

Acknowledgments

This report is the culmination of the efforts of many individuals over several years. The authors wish to express appreciation for the contributions of the many people from school districts, the Danielson Group, Learning Forward, and the American Institutes for Research® (AIR®) who contributed to the success of the evaluation.

From school districts, we appreciate the work of teachers, coaches, and central office leaders, who made these efforts even during the difficult conditions caused by the COVID-19 pandemic. These school districts and staff allowed us to select the treatment and control schools by lottery, collect data in schools, and conduct interviews and administer surveys with staff. Without the support and collaboration of these individuals, our evaluation would not have been possible. For privacy reasons, we do not acknowledge school district staff by name.

We are especially indebted to staff at Learning Forward and the Danielson Group—in particular, Denise Borders, Tracy Crow, Charlotte Danielson, Jim Furman, Shirley Hall, Stephanie Hirsh, Lee Kappes, Jaqueline Kennedy, Michael Lanham, Tom Manning, Nikki Mouton, and Dara Stigdon—as well as those who served as program specialists to support local coaches and teachers: Wendy Anderson, Tanya Batzel, Jorie Ellis, John Eyolfson, Virginia Hall, Teresa Lien, Peggy Olcott, and Kellie Randall.

Many AIR staff contributed to the project effort. Jessica Giffin, Jane Coggshall, and Nayssan Safavian were instrumental to the success of the work. We are grateful to the numerous staff who also contributed significant amounts of their time: Connie Chandra, Nicol Christie, James Colyott, Marlene Darwin, Isabelle Edwards, Janice Esau, Jennifer Ford, Ellie Fulbeck, Alison Hauser, Anna Healy, Ursula Hill, Abigail Jefferys, Shannon Keuter, Vivian Le, Katelyn Lee, Varsha Menon, Danielle Midgyett, Sara Mitchell, Oluchi Moten, Caleb Perlman, Martha Ramirez, Amy Roach, Elizabeth Rohrback, Zipporah Sanders, Harriet Sidford, Honora Stagner, George Stifel, Nathalia Trujillo, and Edith Tuazon.

We also appreciate the expert advice of several AIR staff, including Carrie Scholz, Anja Kurki, Dean Gerdeman, Drew Atchison, and Samantha Neiman. Robin Gurley provided careful editorial assistance in the preparation of the report. We benefited from the advice of Anne Wolf at Abt Associates as well, which provides valuable evaluation guidance to all Education Innovation and Research (EIR) grantees.

In addition, we want to acknowledge the U.S. Department of Education's EIR grant program through the Office of Elementary and Secondary Education for its generous financial support. An EIR midphase grant (U411B180032) enabled Learning Forward and the Danielson Group to

implement Professional Learning with Impact (PLI) in the study schools and paid for all of our data collection, analysis, and reporting.

The authors take responsibility for any errors in this report.

Abstract

The purpose of this study is to examine the implementation and impact of a 2-year evidence-based teacher professional learning program, called Professional Learning with Impact (PLI), using a randomized controlled trial with a sample of 32 elementary schools across four districts, each in a different state. The program was designed to meet a need in the field for scalable, effective teacher professional learning programs, especially those that work in high-need schools. The implementation data show that although the program was not implemented as designed, each of the planned professional learning activities occurred to some extent. The implementation data also point to some challenges related to the scalability of the program, in addition to the COVID-19 conditions, which made the program difficult to implement. The impact analyses found the program had no statistically significant impact on teacher self-efficacy or classroom practice. However, we found the program had a statistically significant positive effect on English language arts student achievement and a positive but not statistically significant effect on math achievement.

Introduction

The COVID-19 pandemic worsened several persistent challenges in K–12 education. Setbacks in achievement levels and teacher job satisfaction were particularly dire, and disparities in achievement between disadvantaged students and their peers grew larger (National Center for Education Statistics [NCES], 2022a, 2022b; Will, 2023). Overcoming these challenges is critical and requires new efforts to develop, scale, and sustain solutions that are effective, especially in schools with high percentages of disadvantaged students.

For achievement and other student outcomes, research underscores the importance of teachers. Longitudinal data show that teachers play a major role in 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). Boosting teacher effectiveness may be especially important for students from low-income families and students of color, who face gaps in access to effective instruction (Cowan et al., 2017; Isenberg et al., 2016). In fact, the Every Student Succeeds Act calls on states and districts to ensure that such students are not taught disproportionately by ineffective teachers.

But how can we improve teacher effectiveness? Among the most widely used approaches is teacher professional learning (PL). Schools invest billions in teacher PL every year (U.S. Department of Education, 2014). However, it is difficult to design PL programs that are effective and scalable. Research finds that some PL programs work, and some do not (Garrett et al., 2021). One of the most promising PL strategies—instructional coaching—is expensive and hard to scale (Kraft & Blazar, 2018; Wayne & Coggshall, 2022).

A Project to Improve, Test, and Scale a Promising Teacher PL Program

In 2018, the American Institutes for Research® (AIR®) launched a project to improve, test, and scale a teacher PL program that builds on evidence of positive impacts from two multisite randomized controlled trials (RCTs) (Garet et al., 2017; Steinberg & Sartain, 2015) and one quasi-experimental study (Taylor & Tyler, 2012). In each study, providing teachers with feedback using Danielson's Framework for Teaching (FfT) led to improved student achievement. Researchers also have found positive relationships between teachers' classroom practice as measured by FfT scores and student achievement gains (e.g., Kane & Staiger, 2012).

The FfT is the basis for this project's focal teacher PL program, called Professional Learning with Impact (PLI), which was designed for elementary schools and includes two main types of PL activities, each led by a national organization with the capacity to deliver PL programs in multiple sites simultaneously. The Danielson Group led the implementation of the first PL

activity, which was instructional coaching grounded in the latest version of the FfT, called the FfT Clusters. Learning Forward led the second PL activity, establishing grade-level teams called Teacher Learning Teams (TLTs), which were to meet regularly to set and pursue goals for classroom practice and student learning. The program begins with a preparatory teacher workshop and includes ongoing support.

Research Questions

As part of the project, AIR led an evaluation of the implementation and impact of the PLI program. This report documents the evaluation and its findings.

The research questions (RQs) guiding this evaluation are based on an understanding of the PLI program and its theory of change. As detailed in the Program Overview section, which appears after this introduction, the program was designed to improve instruction, student engagement, and achievement. Thus, the study was designed to answer the following questions about implementation and impact.

The RQs related to the implementation evaluation of the PLI program are as follows:

1. To what extent were the key components of the program (i.e., the preparatory teacher workshop, TLT cycles, individual instructional coaching, and ongoing support) implemented with fidelity?
2. To what extent were the key components of the strategy to scale (i.e., guidance for district selection of PLI Coaches, coach workshops, and ongoing coach support and monitoring) implemented with fidelity?
3. What were the barriers to and facilitators of implementation?
4. To what extent does the implementation of the PLI program and scaling strategy vary across districts, and what contextual factors are perceived as affecting implementation?

The impact evaluation addresses the following RQs about the impacts of the PLI program on teacher and student outcomes:

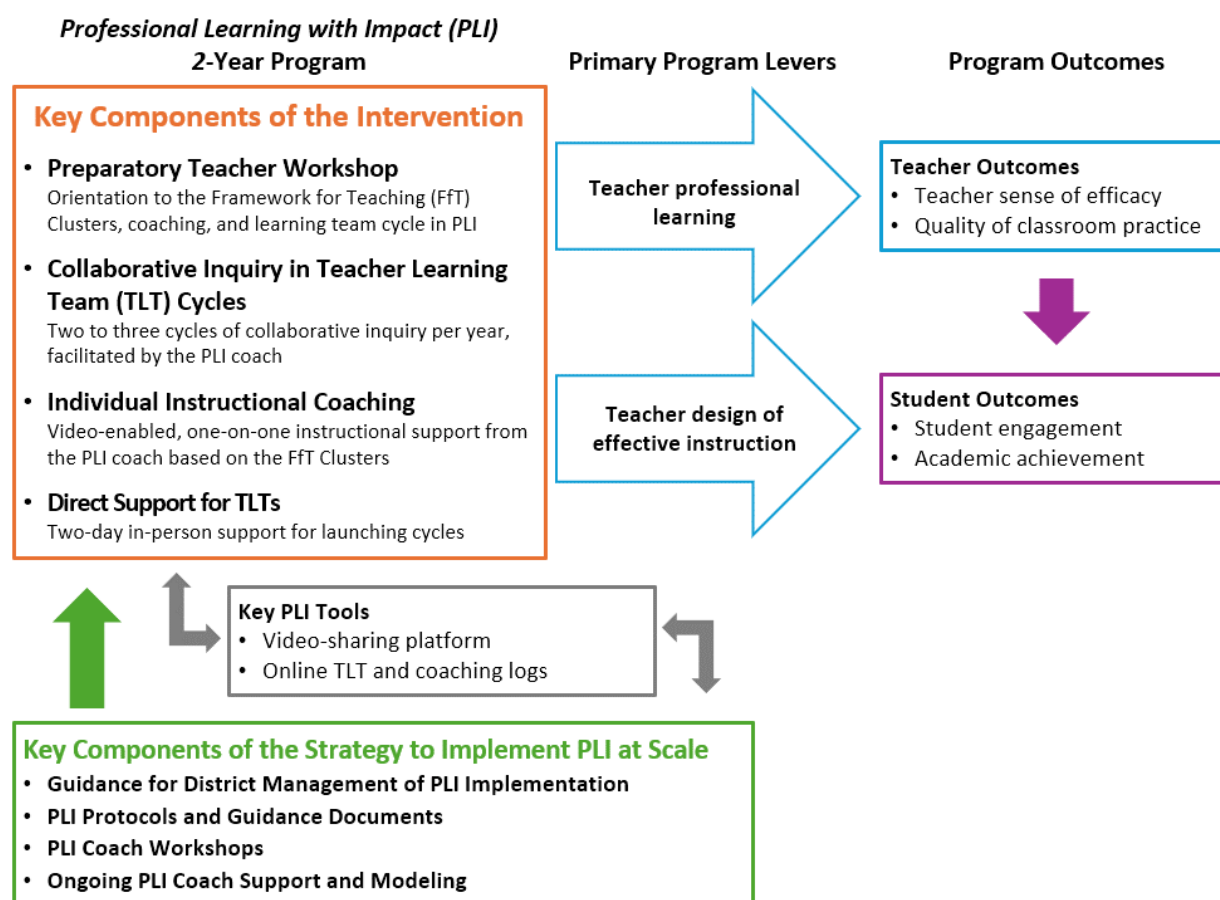
5. What is the impact of PLI on teachers' self-efficacy for improving classroom practice and the quality of teachers' classroom practice?
6. What is the impact of PLI on students' academic achievement and engagement?
7. To what extent is the impact of PLI on student achievement moderated by student background characteristics?

Program Overview

The PLI program includes 2 years of PL activities supported by the Danielson Group and Learning Forward using a *strategy to implement PLI at scale*. The theory of change in Exhibit 1 depicts that strategy's four components in the box on the lower left. Above the strategy components is the *PLI program*, which itself incorporates four elements. Moving to the right, the PLI program activates the two *primary program levers*: (1) teacher PL and (2) teacher design of effective instruction. Through these levers, the program is expected to have an impact on the *program outcomes*. Improvements in teacher sense of efficacy and classroom practice may affect student engagement and achievement.

This theory of change provides the context for more detailed descriptions of the strategy to implement PLI at scale and of the PLI program itself.

Exhibit 1: Professional Learning with Impact Theory of Change



Strategy to Implement PLI at Scale

The strategy to scale the PLI program consists of four components, detailed in tabular form in Exhibit 2. The main feature of PLI that makes it scalable is its use of trainers called program specialists—who are staff members at the Danielson Group and Learning Forward. The program specialists support local school or district staff to serve as PLI Coaches. The coaches, in turn, deliver the most intensive PLI program components to teachers (i.e., collaborative inquiry in TLT cycles and individual instructional coaching).

The four strategy-to-scale components include not only guidance for district management of PLI implementation but also specific supports that program specialists provide to coaches. The PLI protocols and guidance documents include the focus for the coach workshops and the ongoing PLI coach support and monitoring.

The strategy to scale leverages two key PLI tools: a video-sharing platform and online logs, which coaches complete. Using these tools, program specialists monitor the progress of the work and the work products in an effort to ensure that PLI is implemented with high fidelity and quality.

Exhibit 2. Key Components of the Strategy to Scale the Professional Learning with Impact (PLI) Program

Key Components of Strategy to Scale	Roles/Responsibilities
1. Guidance for District Management of PLI Implementation	
Coach selection criteria; explicit descriptions of coach, teacher, and principal responsibilities (including expected labor hours) and performance expectations; and guidance regarding communication of those expectations provided to districts.	School district partners commit to implementing guidance provided by program specialists.
2. PLI Program Protocols and Guidance Documents	
Standard protocols and guidance documents for coaches across sites, including stage protocols, PLI guidebooks for coaches and teams, Framework for Teaching (FFT) Clusters framework documents, and so on.	Program specialists provide and train coaches to use protocols.
3. PLI Coach Workshops	
- An initial 2 1/2-day, in-person preparatory workshop for all PLI coaches (which includes an introduction to using the FFT Clusters in PLI, Teacher Learning Team (TLT) cycles, and the video-sharing platform, and practicing facilitation and coaching strategies) as well as a half-day virtual or in-person follow-up in each district in Program Year 1 (PY1). - A 1 1/2-day in-person or virtual booster workshop in PY2.	Coaches participate in workshop; program specialists facilitate/lead workshops.

Key Components of Strategy to Scale	Roles/Responsibilities
4. Ongoing PLI Coach Support	
- Monthly 1-hour individual monitoring and support calls with PLI coaches in PY1; bimonthly in PY2. - Monthly 1-hour group support calls with PLI coaches in PY1; bimonthly in PY2. - To inform the support calls, program specialists monitor the video-sharing platform and coach logs to provide targeted feedback to coaches. Coaches also may upload videos of their coaching and TLT facilitation practice to the Whetstone platform to receive feedback from their assigned program specialist.	Coaches participate; program specialists implement ongoing monitoring and support.

PLI Program and Its Components

The implementation of the program spans 2 years, beginning with a preparatory teacher workshop led by program specialists. As shown in Exhibit 3, after the workshop, coaches begin delivery of the core components of the program. Specifically, they meet with the TLTs to launch the first collaborative inquiry cycle, which provides the structure and focus for team’s work across meetings. The TLTs continue meeting regularly across the 2-year period, completing multiple cycles. In addition, at a specific stage of each cycle, the coaches provide individual coaching aligned to the goals adopted by the TLTs and the individual teachers’ goals. To further help the coach and enhance teachers’ learning experiences, program specialists provide additional direct support to TLTs.

Exhibit 3. Key Components of the Professional Learning with Impact (PLI) Program

Key Component of Program	Roles/Responsibilities
1. Preparatory Teacher Workshop	
A 6-hour-long workshop (spread across two sessions virtually or offered in-person during a single day) introduces the learning team cycle activities, the Framework for Teaching (FFT) Clusters, and one-on-one coaching rounds.	Teachers participate; program specialists lead/facilitate.
2. Collaborative Inquiry in Teacher Learning Team (TLT) Cycles	
In a five-stage curriculum- and/or practice-focused cycle of inquiry, each TLT reviews student work to select a problem of practice, engages in learning to find solutions, and adjusts their practice based on feedback. Teams meet three to four times per month, completing two to three cycles per program year (PY). FFT Clusters are used throughout.	Teachers participate; PLI coaches facilitate.
3. Individual Instructional Coaching	
During stages 3 and 4 of each TLT cycle, the PLI coach meets with each teacher individually for a coaching round using the FFT Clusters. The coaching round consists of a planning conversation, then two to three video reflections (in which teachers video-record themselves teaching, and coaches support their reflection on the video through questioning and an in-person debrief), followed by a summarizing reflection conversation.	Teachers participate; PLI coaches facilitate.

Key Component of Program	Roles/Responsibilities
4. Additional Direct Support for TLTs	
Two in-person or virtual district-based workshops (one 3-hour and one 6-hour) are held near the beginning of each learning team cycle in PY1 only.	Teachers and coaches participate; program specialists facilitate.

Through the TLT cycles, the PLI coaches facilitate a structured process designed to empower teachers to solve important student learning challenges that they have identified through a close review of existing data. Each learning challenge provides the basis for sustained inquiry in a five-stage cycle, outlined in Exhibit 4. These five stages are designed to put student learning and the curriculum at the center of the teacher’s efforts and professional learning, while the FfT Clusters help teachers codevelop and refine instructional practices.

Exhibit 4. Five Stages of the Professional Learning with Impact Teacher Learning Team Cycles

Teams examine student and educator learning challenges by analyzing student work, data, curricular materials, learning expectations, and other information.
Teams examine their findings from Stage 1 through the lens of the Framework for Teaching (FfT) Clusters, and identify shared goals for student and educator learning.
Teams collaborate to explore or design solutions to the challenges using the FfT Clusters and other resources provided by the PLI coach or one another, sometimes simulating those solutions with their colleagues to get additional FfT-based feedback.
Teachers implement their new learning and instructional strategies in classrooms, and collect information on students’ responses.
Teams examine students’ responses and teachers’ observations, and adjust their classroom practice as a result, documenting their learning as they go.

Concurrently with the learning team cycles, just after the second stage, PLI coaches begin one-on-one coaching rounds with teachers to provide personalized support. Teachers share classroom videos with their coach using the platform (provided by an external vendor, Whetstone Education, Inc.), and both make comments directly on the video. Teachers also may share classroom videos with their teams as part of learning team cycles.

Coaches document and track both their TLT and coaching activities using an online log (also supported by Whetstone Education, Inc.). Coaches can upload stage protocols and meeting agendas, document teachers’ learning goals and next steps, and input other information into the platform, helping them stay organized and remember earlier decisions.

Evaluation Overview

This evaluation focuses on a cohort of schools that began in fall 2021 and ended in spring 2023, but the overall PLI project consisted of three successive cohorts of schools, occurring in the timeline shown in Exhibit 5. The first two cohorts were both treatment-only pilot cohorts, with five schools in Cohort 1 and three schools in Cohort 2.¹ The evaluation team collected program implementation data from the eight pilot schools to refine data collection instruments and inform continuous improvement of the PLI program.

This report focuses on the randomized controlled trial we completed using Cohort 3, which included 32 schools spread across four districts, each in a different state. Within each of the four districts, we randomly assigned the eligible schools to treatment and control groups, resulting in a sample of 17 treatment and 15 control schools.

The evaluation examined the implementation and impact of the PLI program delivered to either Grade 4 teachers or Grade 5 teachers in a given school. The choice of the focal grade (either Grade 4 or 5) for the program is based largely on the preference of each study school.

The COVID-19 pandemic created significant challenges with recruitment and school attrition. For statistical power, we wanted as many as 80 schools in Cohort 3. School systems were under stress due to COVID-19, making it difficult to find districts and schools willing to participate in RCTs.

COVID-19-related challenges continued during the 2-year study period, especially during the 2021–22 school year when the Delta and Omicron variants struck. We were able to gather achievement data for all of the randomized schools. However, we were not able to gather teacher outcome data from all schools because some stopped participating part way through the study, as discussed in more detail in the sections on implementation findings and impact findings.

Exhibit 5. Timeline of Professional Learning with Impact (PLI) Implementation, by Cohort

Cohort	Number of Schools	PLI Program for Study Teachers in Treatment Schools				
		Study Year 1 (2018–19)	Study Year 2 (2019–20)	Study Year 3 (2020–21)	Study Year 4 (2021–22)	Study Year 5 (2022–23)
Cohort 1	5 treatment		PY1	PY2		
Cohort 2	3 treatment			PY1		
Cohort 3	17 treatment, 15 control				PY1	PY2

Note. PY1 = Program Year 1; PY2 = Program Year 2.

¹ According to our original evaluation plan, Cohort 2 would be an impact study cohort. However, due to disruptions to the project caused by COVID-19 (particularly due to the lack of spring 2020 state test data), we decided to treat Cohort 2 as a second pilot cohort rather than an impact study cohort.

Implementation Evaluation

In this section, we describe the implementation evaluation. The implementation evaluation addresses the first four RQs, which examine the extent to which the key components of the program and strategy to scale were implemented with fidelity, the barriers to and facilitators of implementation, and the contextual factors perceived as affecting implementation.

To organize this section, we first describe the measures and data sources used to determine fidelity of implementation. Second, we define the analytic samples. Third, we present the evaluation findings, including the extent to which PLI was implemented with fidelity in Years 1 and 2 for Cohort 3 schools (RQs 1 and 2) and the barriers and facilitators that coaches, teachers, and program specialists perceived as affecting implementation (RQ3). Last, we discuss implementation differences across districts and the contextual factors that may have affected implementation (RQ4). These findings provide context for interpreting the impact findings.

Measures and Data Sources

As outlined in the Program Overview section, the PLI program has four key components—the preparatory teacher workshop, collaborative inquiry in TLT cycles, individual instructional coaching, and additional direct support for TLTs. There also are four strategy-to-scale components—guidance for district management of PLI implementation, PLI program protocols and guidance documents, PLI coach workshops, and ongoing PLI coach support. Program specialists from the Danielson Group and Learning Forward carried out all of these components except for the TLT cycles and individual coaching, which were conducted in each school by trained local coaches.

For each component, the project team collaborated to develop measurable indicators of implementation fidelity and set thresholds for expected levels of implementation. Exhibit 6 outlines the specific indicators and data sources. As shown in the right-side column, the project team used a variety of methods to collect implementation information, including observations, attendance, interviews, surveys, work product reviews, and coach logs.²

² PLI used an online platform for PLI coaches to track TLT meetings, TLT cycle progress, one-on-one coaching meetings, and teacher classroom video reflections and comments. Some coaches reported difficulty in using the platform while others found the platform to be a useful tool. Due to the challenges with the platform experienced by some coaches, we believe coach log data may be an underestimate of implementation in some cases. The data sources column in Exhibit 6 shows the implementation measures that use coach log data.

Exhibit 6. Implementation Evaluation Measures and Data Sources, by Component

Key Components	Indicators	Data Sources
Program Components		
1. Preparatory Teacher Workshop	a. Proportion of summer team workshop sessions attended b. Summer team workshop topic coverage c. Proportion of participants who demonstrate intellectual engagement with the content d. Proportion of participants who perceive value in their engagement with Professional Learning with Impact (PLI)	a. Attendance records b. American Institutes for Research (AIR) observer report c. Workshop feedback survey d. Workshop feedback survey
2. Teacher Learning Team (TLT) Cycles	a. Number of TLT meetings convened b. Number of TLT meetings attended c. Number of learning team cycles completed d. Fidelity of learning team work products (i.e., stage protocol text) e. Teacher perceptions of team engagement	a. Coach log (online) b. Coach log (online) c. Coach stage protocols (online) d. Fidelity scoring rubric applied to online record (stage protocol) for a randomly selected sample stage protocol for each coach e. Teacher spring survey (Program Year [PY] 1, PY2)
3. Individual Instructional Coaching	a. Number of PLI coaching rounds completed b. Fidelity of coach's products (i.e., video reflection text) c. Teacher perceptions of coach supportiveness	a. Coach log (online) b. Fidelity scoring rubric applied to online work product (video reflection text) for a randomly selected sample video reflection text for each coach c. Teacher spring survey (PY1, PY2)
4. Direct Support for TLTs	a. Proportion of fall and winter team workshops attended b. Fall and winter team workshop topic coverage	a. Attendance records b. AIR observer report
Strategy-to-Scale Components		
1. Guidance for District Management of PLI Implementation	a. Program provides written guidance for successful implementation to district partners (i.e., district booklet). b. District commits to ensuring conditions for successful implementation (i.e., signed services agreement). c. District hires/assigns qualified coaches.	a. Recruitment/implementation records b. Signed services agreements c. Program specialist survey (PY1, PY2)
2. PLI Program Protocols and Guidance Documents	a. Program provides cross-site guidance materials. b. Program specialists provide training to coaches on key PLI program materials.	a. Implementation records b. AIR observer reports

Key Components	Indicators	Data Sources
3. PLI Coach Workshops	a. Proportion of coach workshops attended (summer coach workshop [PY1] and refresher workshop [PY2]) b. Coach workshop topics covered as intended c. Coach reported comfort with intended content	a. Attendance records b. AIR observer report c. Workshop feedback surveys
4. Ongoing PLI Coach Support	a. Program specialist review of coaching conversation and TLT facilitation videos b. Number of monthly one-to-one coaching call support sessions attended (by coach) c. Number of monthly virtual all-coaches professional learning community attended (by coach) d. Coach satisfaction with program specialist support	a. Coach interview (PY1, PY2) b. Program specialist log c. Program specialist log d. Coach interview (PY1, PY2)

Analytic Samples for the Implementation Analysis

We defined the analytic samples for the implementation analysis to give readers clear information about the program activities associated with the reported impacts. Specifically, we defined samples for Year 1 (2021–22) and Year 2 (2022–23), as is conventional for studies of 2-year programs. Readers can thus understand Year 1 implementation when interpreting Year 1 impacts, and consider implementation across both years when interpreting Year 2 impacts, which reflect the cumulative impact of 2-year programs.

Exhibit 7 shows the count of districts, schools, teachers, PLI coaches, and program specialists for the two primary implementation analysis samples in Years 1 and 2. For example, the primary implementation sample for Year 1 included 12 treatment schools and 56 treatment teachers. For Year 2, there were still 12 treatment schools but only 38 of the original treatment teachers, due to teacher mobility.

Exhibit 7. Primary Analytic Samples for Findings About Implementation

Participants	Year 1 (2021–22) in Three 2-Year Districts	Year 2 (2022–23) in Three 2-Year Districts
Districts	3	3
Schools	12	12
Teachers	56	38
PLI Coaches	12	12
Program Specialists	5	5

The primary implementation samples included only the treatment teachers who were part of the analyses of impact on teacher outcomes, who were the teachers who remained in their original assignments. For Year 1 and Year 2, these samples span three districts, even though four districts participated in the study. The fourth study district experienced teacher mobility such that no study teachers remained in one of the two treatment schools in Year 2. With the loss of all study teachers in one of the treatment schools in Year 2, we did not have the teacher outcome data that would be needed to include this district in the Year 2 teacher impact findings.

To generate implementation findings that include teachers in the fourth district, we defined supplemental implementation samples. Specifically, in Appendix B, for each year, we report implementation findings for “all four districts” to supplement findings for the “three 2-year districts” presented in this section. The findings based on these supplemental implementation samples are the most relevant to consider when interpreting the primary analyses of impact of student achievement because these intent-to-treat analyses account for all students at the target grade levels, regardless of teacher mobility.

The number of coaches serving the teachers in the primary implementation samples remained the same across the 2 years (12 in each year), but there was some coach mobility. There was no PLI coach mobility in Year 1; however, in Year 2, three schools in two of the three 2-year districts experienced a change in PLI coaches.

Specifically, in one district in Year 2, a PLI coach moved from one treatment school to another treatment school to replace a PLI coach who left the school. Also, in this district in Year 2, a new PLI coach was trained to replace the PLI coach who moved to the treatment school with the vacant PLI coach position. In a second district in Year 2, a new PLI coach was trained to replace a PLI coach who left after Year 1.

PLI coach mobility does not necessarily lead to reduced coaching dosage. However, it is possible that having a PLI coach replaced could lead to lower quality coaching or coach-teacher relationships.

Findings

Implementation Fidelity for the Program Components (RQ1)

PLI program specialists, PLI coaches, and teachers implemented each of the four program components—Preparatory Teacher Workshop, TLT Cycles, Individual Instructional Coaching, and Direct Support for TLTs—to varying degrees, as detailed in this section.

Overall, implementation of the Preparatory Teacher Workshop component and Direct Support for TLTs component had high levels of implementation, but that was not true for the TLT Cycles component and the Individual Instructional Coaching component. Due to the uneven implementation both within and across districts in both implementation years, the implementation of the four program components did not meet the component-level thresholds for adequate implementation fidelity that AIR, the Danielson Group, and Learning Forward had defined for the evaluation, which are depicted in Exhibit 8.

Implementation Fidelity for the Four Program Components—Preparatory Teacher Workshop, TLT Cycles, Individual Instructional Coaching, and Direct Support for TLTs

Each component was implemented, but the dosage of two components—TLT Cycles and Individual Instructional Coaching—was limited.

Implementation did not meet the thresholds for fidelity for any of the four components. Even for components that were implemented at nearly the intended dosage, there were individual districts that failed to meet the threshold for dosage or adherence to the program model.

Exhibit 8. Scoring of Adequate Implementation of Each Key Program Component, by Year

Key Program Components	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Preparatory Teacher Workshop	Fidelity Not Met 33% of districts (one of three) met threshold	N/A
Teacher Learning Team (TLT) Cycles	Fidelity Not Met 75% of schools (nine of 12) met threshold	Fidelity Not Met 83% of schools (10 of 12) met threshold
Individual Instructional Coaching	Fidelity Not Met 8% of schools (one of 12) met threshold	Fidelity Not Met 25% of schools (three of 12) met threshold
Direct Support for TLTs	Fidelity Not Met 33% of districts (one of three) met threshold	N/A

In this section, we describe the level of implementation for each program component, as shown in Exhibits 9–12. We also discuss the extent to which each program component met the threshold for adequate implementation.

Preparatory Teacher Workshop. The first program component is the Preparatory Teacher Workshop, which took place in Year 1 only, as planned. We examined four indicators as shown in Exhibit 9. Most teachers attended the workshop, which covered all of the intended topics, as shown in the exhibit. In addition, 85% of teachers reported a high level of intellectual engagement with the workshop content, and a similar share reported perceived value from engaging in PLI at the time of the workshop.

Exhibit 9. Participation in the Preparatory Teacher Workshop

Key Program Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: Preparatory Teacher Workshop		
Percentage of teachers attending preparatory teacher workshop	87.5%	N/A
Percentage of teacher workshops with topics covered as planned	100%	N/A
Percentage of workshop respondents reporting a high level of intellectual engagement with the workshop content	85.4%	N/A
Percentage of workshop respondents reporting perceived value from engagement with Professional Learning with Impact	90.2%	N/A

Though implemented nearly as planned, the implementation of the Preparatory Teacher Workshop component did not meet the threshold for adequate fidelity. To meet it, all districts would have had to meet the threshold for attendance (95% of teachers), the threshold for adequate topic coverage (all but one topic fully covered), the threshold for participant intellectual engagement with the workshop content (95% of respondents), and the threshold for participants’ perceived value from their engagement (95% of respondents). Not all districts met these thresholds, as shown in the first row of Exhibit 8. A few teachers from two of the three districts did not attend in-person and instead completed an asynchronous make-up session or attended for less time than the threshold required (86%). In addition, about 15% of teachers (six teachers) in two of the three districts disagreed that they “engaged in lively, intellectually stimulating conversation during the workshop,” and about 10% of teachers (four teachers) disagreed that they “participated fully in the breakout room activities during the

workshop.” One of the three districts met the overall threshold for the Preparatory Teacher Workshop, as shown in Exhibit 8.

TLT Cycles. The second program component is the TLT Cycles, which took place in Years 1 and 2. We examined five indicators, as shown in Exhibit 10.

The dosage was expected to average 24 TLT meetings per year (i.e., three times per month for 8 months), with each meeting lasting 45 minutes. In practice, PLI coaches implemented the TLT cycles with widely different dosages and approaches in different schools. Some schools convened the TLT meetings at the intended pace, some at a lesser pace, and some not at all. The mean number of meetings was a little less than half of the intended number, and the range varied across schools, from four to 24 TLT meetings convened, as shown in Exhibit 10.

Additional indicators of implementation show that most teachers attended and were engaged in the TLT meetings, but not all teams followed the program model. The program model called for each TLT to complete two full learning cycles per year using a tool called a Stage Protocol (i.e., completing two stage protocols with five stages each per implementation year). The mean number of learning team cycles completed in Year 1 was about half of what was expected, with a mean of 1.3 learning team cycles completed for the three districts. In Year 2, the mean number of learning team cycles completed was higher, with a mean of 1.8 learning team cycles completed for the three districts.

To further understand implementation of the TLT Cycles component, we examined PLI coaches’ TLT work products using a 17-item checklist. The expected level of implementation was that PLI coaches would achieve a score of at least 16 on the 17-item checklist. The mean number of items achieved by PLI coaches on the 17-item check list was about 12, as shown in Exhibit 10. Although not all TLTs followed the program model, 81% of teachers reported positive engagement with their TLT.

Exhibit 10. Participation in the Teacher Learning Team Cycles

Key Program Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: Teacher Learning Team (TLT) Cycles		
Mean number of TLT meetings convened	11.67 (4–24)	9.42 (1–22)
Mean number of TLT meetings attended	10.73 (0–24)	7.29 (0–20)

Key Program Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Mean number of learning team cycles completed	1.33 (0–3)	1.83 (0–4)
Mean number of items that PLI coaches achieved on 17-item fidelity stage protocol checklist	11.7 (0–16)	12.0 (0–17)
Percentage of items about team engagement rated <i>agree</i> or <i>strongly agree</i>	80.8%	80.7%

Based on limited dosage and minimal adherence to the program model learning team cycle protocol, the implementation of the TLT cycles did not meet the overall threshold for adequate fidelity in either year. To meet the overall threshold, 85% of schools would have had to meet a combination of the threshold for the TLT meetings convened (meeting more than 90 minutes per month on average), the threshold for TLT meetings attended (75% or more of teachers attending 66% or more of meetings), the threshold for learning team cycles completed (one or more cycles per year), the threshold for scoring on the 17-item fidelity stage protocol checklist (a score of 13 or higher), and the threshold for teachers’ perception of team engagement (a score of 19 or higher out of 36 points). As shown in Exhibit 8, 75% of schools in the three districts met the overall threshold for TLT cycles in Year 1. In Year 2, a higher percentage of schools met the overall threshold than in Year 1.

Individual Instructional Coaching. The third program component is individual instructional coaching. We examined three indicators in Years 1 and 2, as shown in Exhibit 11. The expected level of implementation for individual instructional coaching was that teachers would complete two rounds of coaching per year and that each round of coaching would include a planning meeting, two or more video debriefs, and a reflection meeting.

Overall, PLI coaches implemented individual coaching with a much lower dosage than expected. The mean number of rounds completed each year was close to zero, as shown in Exhibit 11. As the ranges indicate, no teachers completed two rounds of coaching in either year. Coaches completed a full round of coaching with 13% of teachers in Year 1 (seven teachers) and 3% in Year 2 (one teacher). Based on these results, we decided to see how many teachers completed any activities associated with coaching, regardless of whether they completed a full round. About 60% of the teachers in Year 1 and Year 2 completed at least one coaching activity (34 teachers and 14 teachers, respectively) by participating in either a video debrief meeting or a reflection meeting with their PLI coach.

In general, the coaching that did occur did not closely follow the specific PLI guidelines for coaching outlined during PLI coach workshops and in the PLI guidebook. We reviewed the fidelity of PLI coaches' work products associated with individual coaching using a 15-item checklist focused on the documentation of the teacher and PLI coach video reflection process. The expected level of implementation was that PLI coaches would achieve a score of at least 14 on the 15-item video reflection checklist. Overall, 50% in Year 1 (six coaches) and 33% in Year 2 (four coaches) scored above a 7 on the video reflection checklist (not shown), and PLI coaches achieved widely different scores on the feedback they provided on teacher video submissions in different schools, as shown by the range in scores in Exhibit 11. Despite this variation, two thirds or more of teachers who participated in at least some individual coaching reported positive perceptions about PLI coach supportiveness.

Exhibit 11. Participation in Individual Instructional Coaching

Key Program Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: Individual Instructional Coaching		
Mean number of Professional Learning with Impact (PLI) coaching rounds that teachers completed	0.13 (0–1)	0.03 (0–1)
Mean number of items that PLI coaches achieved on a 15-item fidelity video reflection checklist	6.8 (0–12)	4.7 (0–14)
Percentage of items about PLI coach supportiveness rated <i>agree</i> or <i>strongly agree</i>	82.7%	69.1%

The low level of implementation of the coaching also was apparent when comparing the level of implementation with the preestablished threshold for adequate implementation fidelity, which was not met in either year. To meet the overall threshold for individual instructional coaching, 85% of schools would have had to meet a combination of the threshold for the PLI coaching rounds completed (an average of two rounds per teacher), the threshold for scoring on the 15-item fidelity video reflection checklist (a score of 11 or higher), and the threshold for teachers' perception of coach supportiveness (a score of 17 or higher out of 32 points). As shown in Exhibit 8, 8% of schools (one school with eight study teachers) in the three districts met the overall threshold for individual instructional coaching in Year 1. In Year 2, a higher percentage met the overall threshold than in Year 1—about 25% (three schools with nine study teachers).

Direct Support for TLTs. The fourth program component is direct support for TLTs. We examined two indicators, as shown in Exhibit 12. More than 90% of teachers (51 teachers)

attended the fall and winter team workshops, and the intended topics were fully covered during the workshops.

Exhibit 12. Participation in Direct Support for Teacher Learning Teams (TLTs)

Key Program Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: Direct Support for TLTs		
Percentage of teachers attending fall and winter team workshops	91.1%	N/A
Percentage of fall and winter workshops with topics covered as planned	100%	N/A

Despite these successes, implementation of the Direct Support for TLTs component did not meet the overall threshold for adequate fidelity. To meet it, all districts would have had to meet the threshold for attendance (95% of teachers) and the threshold for adequate topic coverage (all but one topic fully covered). Only one of the three districts met the overall threshold for direct support for TLTs, as shown in Exhibit 8.

Implementation Barriers and Facilitators for Program Components (RQ3)

To identify perceived barriers to program implementation related to the TLT cycles and individual coaching as well as facilitators of successful implementation, we used data from 12 PLI coach interviews in Year 1 (100%), 11 PLI coach interviews in Year 2 (92%), and open-ended items on the treatment teacher survey from 46 teachers in Year 1 (82%) and 41 teachers in Year 2 (76%) who responded to the survey and provided at least one open-ended response. Although these data do not allow us to quantify perceptions, they illustrate the types of barriers and facilitators that participants encountered.

Implementation Barriers for TLT Cycles and Individual Coaching. In both implementation years, PLI coaches reported a number of challenges to following the TLT cycle process. In particular, PLI coaches reported not having enough time to progress in the cycle. For example, in Year 1, PLI coaches cited teachers being out on leave for various reasons (e.g., COVID-19) as a challenge. As one PLI coach in Year 1 explained, the biggest challenge with implementing the TLT cycles is “trying to find the time to meet and to progress through the stages because everyone is here or not here at different times and trying to all get on the same page.” In Year 2, PLI coaches reported challenges stemming from competing priorities, such as learning a new English language arts (ELA) curriculum and “finding the time to dedicate outside of our other mile high [list of] things

that we have to do.” One PLI coach in Year 1 also shared that they did not have common planning time available during the school day that could be used for the TLT meetings.

Similarly, most PLI coaches in both implementation years reported lack of time as a major challenge to engaging in the one-on-one coaching rounds. In Years 1 and 2, PLI coaches shared that time constraints stemmed from a range of issues, such as mismatches between PLI coach and teacher schedules, PLI coaches or teachers having other duties, lacking a consistent time to meet, and needing to respect teachers’ time (e.g., not scheduling during planning time, before/after school, at lunch).

Data from an open-ended question on the teacher surveys paint a similar picture. We asked treatment teachers what prevents them from getting the most out of PLI. Similar to PLI coaches, teachers pointed to limited time and additional workload. One teacher specifically said, *“I would say the challenges are mostly time related. There is not much time in the day to get PLI stuff done.”* Some teachers who responded to this question reported that implementing PLI added to their already demanding workload and limited the time they had to fulfill other duties (e.g., planning and attending meetings).

Implementation Facilitators for TLT Cycles and Individual Coaching. For TLT cycle implementation, PLI coaches reported that the factors contributing to success in both Years 1 and 2 included setting a common goal for the TLT, using PLI as an opportunity for purposeful reflection, and applying the TLT cycle process as a way to stay focused and on track. In Year 1, a PLI coach explained that the biggest success factor for the TLT cycle was *“definitely, the team communication and collaboration, just setting aside a common goal for everyone and sharing what’s working, what’s not working, and how can we support each other. Then it trickles down into their individual and student goals. So just providing those opportunities as a team to talk about a common goal and how to accomplish it.”*

For one-on-one coaching implementation, the majority of PLI coaches in both years reported that the factors contributing to success included encouraging teachers to reflect on their practice and come to a new understanding or realization about their work with students. In Year 1, one PLI coach described how teacher reflection, an important aspect of PLI one-on-one coaching, contributed to successful implementation, noting, *“They are reflecting on their practice and they’re having realizations or coming to their realizations and it’s all through their own thinking. It’s not through me telling.”* In Year 2, PLI coaches discussed additional facilitators that supported successful implementation, including goal setting with teachers and focusing on the positives of the work.

We asked treatment teachers about what helps them to get the most out of the PLI program. Teachers reported that support from leadership—primarily their PLI coach—helped them to get

the most out of PLI. For example, one teacher wrote that their “*coach’s support, guidance, and mentorship*” was helpful in navigating PLI. Other teachers stated that having time set aside for work related to PLI helped them get the most out of the program, with one teacher mentioning the benefit of having “*more time to review data, create new ideas for engagement, and [talk through] ways to implement what we discussed.*”

Implementation Fidelity for the Strategy-to-Scale Components (RQ2)

PLI program specialists successfully implemented two of the four strategy-to-scale components—Guidance for District Management of PLI Implementation and PLI Program Protocols and Guidance Documents—and met component-level thresholds for adequate implementation fidelity. Aspects of the remaining two strategy-to-scale components—PLI Coach Workshops and Ongoing PLI Coach Support—did not consistently reach all PLI coaches and did not meet component-level thresholds for adequate implementation. Exhibit 13 shows whether each strategy-to-scale component (described in detail earlier in Exhibit 2) met the threshold for adequate implementation.

Implementation Fidelity for the Four Strategy-to-Scale Components—Guidance for District Management of PLI Implementation, PLI Program Protocols and Guidance Documents, PLI Coach Workshops, and Ongoing PLI Coach Support

PLI program specialists implemented each of the four strategy-to-scale program components.

Implementation of the first two strategy-to-scale components met the component-level thresholds for adequate implementation fidelity.

Implementation of the third and fourth components—PLI Coach Workshops and Ongoing PLI Coach Support—reached many but not all coaches, and therefore did not meet component-level thresholds for adequate implementation fidelity.

Exhibit 13. Scoring of Adequate Implementation of Each Key Strategy-to-Scale Component, by Year

Key Scaling Components	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Guidance for District Management of Professional Learning with Impact (PLI) Implementation	Fidelity Met 100% of districts (three of three) met threshold	N/A
PLI Program Protocols and Guidance Documents	Fidelity Met 100% of districts (three of three) met threshold	N/A

Key Scaling Components	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
PLI Coach Workshops	Fidelity Not Met 67% of districts (two of three) met threshold	N/A
Ongoing PLI Coach Support	Fidelity Not Met 83% of schools (10 of 12) met threshold	Fidelity Not Met 33% of schools (four of 12) met threshold

In this section, we describe the level of implementation for each strategy-to-scale component, as shown in Exhibits 14–17. For the two strategy-to-scale components that did not meet the component-level thresholds, we describe the thresholds in more detail.

Guidance for District Management of PLI Implementation. The first strategy-to-scale component is Guidance for District Management of PLI Implementation. We examined three indicators in Year 1, as shown in Exhibit 14. All three districts received written guidance about managing PLI implementation. All three districts also signed services agreements to commit to ensuring conditions for successful implementation. In addition, based on program specialists’ judgments, all PLI coaches met the coach qualification criteria: understanding of their schools’ curriculum, the ability to analyze student data, and the ability to help teachers identify instructional needs. This strategy-to-scale component reached all three districts and met the overall component-level threshold for adequate implementation fidelity, as shown in Exhibit 13.

Exhibit 14. Participation in Guidance for District Management of Professional Learning with Impact (PLI) Implementation

Strategy-to-Scale Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: Guidance for District Management of PLI Implementation		
Percentage of districts that received written guidance for successful implementation	100%	N/A
Percentage of districts that committed to ensuring conditions for successful implementation (i.e., a signed services agreement)	100%	N/A
Percentage of districts that hired/assigned qualified coaches	100%	N/A

PLI Program Protocols and Guidance Documents. The second strategy-to-scale component is PLI Program Protocols and Guidance Documents. We examined two indicators, as shown in Exhibit 15. All three districts received the PLI guidebook, the FfT, and access to the coaching platform, Whetstone. Training on these program protocols and guidance documents was fully and completely provided by the Danielson Group and Learning Forward. This strategy-to-scale component reached all three districts, as shown in the second panel of Exhibit 15. The component met the overall component-level threshold for adequate implementation fidelity, as shown in Exhibit 13.

Exhibit 15. Participation in Professional Learning with Impact (PLI) Program Protocols and Guidance Documents

Strategy-to-Scale Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: PLI Program Protocols and Guidance Documents		
Percentage of districts that received cross-site guidance materials	100%	N/A
Percentage of districts that received training for coaches on key PLI program materials	100%	N/A

PLI Coach Workshops. The third strategy-to-scale component is PLI Coach Workshops. We examined three indicators, as shown in Exhibit 16. All of the PLI coaches in the three districts attended the summer PLI coach workshop and refresher workshop. The AIR site coordinators reported that all of the intended topics were fully covered during the PLI coach workshops for the three districts. We also asked PLI coaches about their level of comfort with the content of the workshops. Almost all of the PLI coaches in the three districts reported comfort with the workshop content.

Exhibit 16. Participation in Professional Learning with Impact (PLI) Coach Workshops

Strategy-to-Scale Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: PLI Coach Workshops		
Percentage of coaches who attended coach workshops (summer coach workshop and refresher workshop)	100%	N/A

Strategy-to-Scale Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Percentage of coach workshops with topics fully covered as intended	100%	N/A
Percentage of coaches reporting comfort with the intended content	92.3%	N/A

Implementation of the PLI coach workshops’ strategy-to-scale component did not meet the overall threshold for adequate fidelity. All districts met the thresholds for attendance (92% of coaches) and adequate topic coverage (all but one topic fully covered). However, the threshold for coach comfort with workshop content (90% of respondents) was not met for each district. Less than 90% (87.5%) of coaches in one of the three districts agreed that they felt “well prepared to facilitate learning team cycles in PLI” and “well prepared to facilitate coaching rounds in PLI.” Two of the three districts met the overall threshold for the PLI coach workshops, as shown in Exhibit 13.

Ongoing PLI Coach Support. The fourth strategy-to-scale component is Ongoing PLI Coach Support. We examined four indicators, as shown in Exhibit 17.

The first program expectation for ongoing support for PLI coaches was that each coach would receive feedback from the program specialist on video recordings of both a TLT meeting and a coaching conversation. Very few coaches in the three districts submitted videos of both a TLT meeting and a coaching conversation for feedback from the program specialist. A modest share submitted one or the other: 33% and 17% of coaches in Year 1 and Year 2, respectively.

The second program expectation for ongoing support for PLI coaches is that the coach would meet one-on-one with their program specialist eight times in Year 1 and four times in Year 2. The mean number of such meetings in Year 1 was just a little less than expected, with a mean of 7.8 for the three districts. In Year 2, however, the mean number of one-on-one meetings was about a quarter of the number expected, with a mean of 0.8 one-on-one meetings. Overall, about two thirds of the coaches in the three districts met eight times one-on-one with their program specialist in Year 1, and very few coaches met four times with their program specialist in Year 2.

The third program expectation for ongoing support for PLI coaches is that the coach would participate in more than six virtual all-coach professional learning communities (PLCs) in both

years.³ Across the three districts, the mean number of virtual all-coach PLCs attended was less than expected in Year 1 and Year 2, with means of 4.3 and 5.0 meetings, respectively. No PLI coaches participated in more than six virtual all-coach PLCs in either year.

The final indicator for ongoing support for PLI coaches is about the quality of support that coaches indicated they received from the program specialist. All coaches in both years reported satisfaction with program specialist support.

Exhibit 17. Participation in Ongoing Professional Learning with Impact (PLI) Coach Support

Strategy-to-Scale Indicators	Implementation Sample	
	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: Ongoing PLI Coach Support		
Percentage of coaches reporting that the program specialist reviewed coaching conversation and Teacher Learning Team facilitation videos	8.3%	0%
Percentage of coaches who attended at least eight (Year 1) or four (Year 2) monthly one-on-one coaching call support sessions	66.7%	8.3%
Percentage of coaches who attended more than six monthly virtual all-coach professional learning community sessions each year	0%	0%
Percentage of coaches reporting satisfaction with program specialist support	100%	91.6%

Based on the data, implementation of the Ongoing PLI Coach Support strategy-to-scale component did not meet the overall threshold for adequate fidelity in Year 1 or 2. To meet the threshold for ongoing coach support, 85% of schools would have had to meet a combination of the threshold for program specialist review of TLT facilitation and coaching conversation videos (90% of coaches report that feedback is useful), the threshold for one-on-one coaching support calls attended (90% of coaches attend eight sessions in Year 1 and four sessions in Year 2), the threshold for monthly virtual all-coach PLCs attended (90% of coaches attend more than six PLCs per year), and the threshold for coach satisfaction with program specialist support (90% of

³ Although no PLI coaches participated in more than six virtual all-coach PLCs in either year, 71.4% of PLI coaches in all four districts and 83.3% of coaches in the three 2-year districts attended at least four virtual all-coach PLCs in Year 1. In Year 2, 78.6% of coaches in all four districts and 91.6% of coaches in the three 2-year districts attended at least four all-coach PLCs.

coaches report satisfaction). As shown in Exhibit 13, 83% of schools in the three districts met the overall threshold for ongoing coach support in Year 1. In Year 2, a much lower percentage of schools met the overall threshold than in Year 1.

Implementation Barriers and Facilitators for the Strategy-to-Scale Components (RQ3)

We used data from 12 PLI coach interviews in Year 1 (100%), 11 PLI coach interviews in Year 2 (92%), and four program specialist surveys in Years 1 and 2 (100% each year) to identify perceived barriers to implementing the strategy-to-scale components related to the PLI Coach Workshops and Ongoing PLI Coach Support as well as facilitators of successful implementation. Although these data do not allow us to quantify perceptions, they illustrate the types of barriers and facilitators that participants encountered.

Implementation Barriers for PLI Coach Workshops and Ongoing PLI Coach Support. In both Years 1 and 2, program specialists identified limited time and competing priorities facing coaches as barriers to implementation. In Year 2, one program specialist noted, *“[The coaches] challenges related to the other responsibilities like coaching three to four other schools, test coordinator, and designing professional development with other district coaches.”*

Although coaches reported feeling well supported for facilitating the TLT cycle process and one-on-one coaching in Year 1, coaches reported feeling less supported in Year 2. By design, in Year 2 of the PLI program, the individual one-on-one coach meetings with program specialists were modified to occasional meetings. One coach in Year 2 reported, *“And that [the monthly 1:1 with the program specialist] was my favorite part of my month...And that’s been taken away. And now I just feel lost.”*

Another barrier to implementing ongoing support for coaches was that only a few coaches submitted a video of a TLT meeting or coaching conversation reflection video to their program specialist. This may be due in part to confusion in Year 2 over whether there was an expectation that coaches would submit these videos to their program specialist. Coaches who did not submit these videos missed the opportunity to receive targeted feedback from the program specialist.

Implementation Facilitators for PLI Coach Workshops and Ongoing PLI Coach Support. In Years 1 and 2, program specialists reported that the main facilitators that supported coaches were the program specialist–individual coach calls (planned monthly in Year 1 and less frequently in Year 2) and the workshops. In Year 1, one program specialist explained, *“I think the one-on-one coaching sessions with specialists have helped keep them [the coaches] focused on PLI in a year where it might otherwise have fallen by the wayside. We tried to offer support that encouraged them to celebrate successes while acknowledging and validating the real challenges they were*

facing.” Similarly in Year 2, a program specialist reported, “I think that the one-on-one monthly coaching conversations between the [program specialist] and school coach were the most helpful. Also, the training they received was very helpful” in supporting coaches.

Although most coaches did not submit videos of a TLT meeting or coaching conversation to their program specialist for feedback, coaches who did submit videos found the feedback from their program specialist to be useful. Coaches who submitted videos and received feedback in Year 1 mentioned that it helped them to (1) prepare for coaching teachers, (2) identify where to use certain skills, and (3) understand how to use the KASAB [Knowledge, Attitudes, Skills, Aspirations, and Behaviors] model. One coach in Year 1 highlighted the usefulness of the feedback from the program specialist, saying, *“Because I felt like she had tagged multiple areas of the video that was submitted, and she was able to explain and tell me reflectively, exactly, what pieces of the process I was using, and what language I was using, she was able to pinpoint, make it clear to me that what I was doing was okay.”*

Implementation Variation Across Districts and Contextual Factors Affecting Implementation (RQ4)

The implementation findings presented above identify some indicators on which implementation varied across districts and within districts. In this section, we examine varied levels of buy-in and support for the program by school leaders as a possible explanation for the variation across study districts and schools. We chose to look closely at buy-in and support for the program by school leaders based on coach interview data and program specialist survey data in both implementation years that highlighted the importance of school leader support.

We do not have a direct measure of school leaders’ support for the program, but we did ask PLI coaches to rate principal support for a common weekly meeting time for teachers. There was a wide range of responses. Exhibit 18 lists districts from A to C in descending order based on the PLI coach-reported average rating for principal support in Year 2. The PLI coaches in District A all reported that principals ensured teachers have such time for meeting “to a great extent” in both implementation years, while the coaches in District C reported mixed support from principals in both years.

The data in the first two columns of Exhibit 18 suggest a positive association between principal support for meeting time and the number of learning team cycles completed. Although the analysis does not allow us to determine causality, the pattern appears to support our hypothesis that support by school leaders explains variation in the extent of implementation.

Principal support also may relate to the amount of one-on-one coaching. As described in the Implementation Fidelity for the Program Components section, across all three districts,

implementation of the one-on-one coaching rounds was lower than expected. The expected number of debrief meetings and reflection meetings per teacher was six per year.⁴ Similar to the variation in the number of learning team cycles completed, Exhibit 18 shows that in districts with higher levels of perceived principal support for teacher common meeting time, teachers, on average, completed more debrief and reflection meetings. Although teacher common meeting time is not required for implementing a coaching round, principal support for teacher common meeting time may be indicative of broader support for school engagement in the PLI program. Teachers in Districts A and B had on average about one debrief or reflection meeting in Year 2, compared with zero meetings for teachers in District C.

Exhibit 18. Variation by District in Year 2 Principal Support for Common Meeting Time, Average Number of Learning Team Cycles Completed, and Average Number of One-on-One Meetings

District	Year 2 (2022–23) in the Three 2-Year Districts		
	Extent to Which Principal Ensures That Teachers Have Common Meeting Time (4 is “to a great extent”): Coach Report	Average Number of Learning Team Cycles Completed by District	Average Number of One-on-One Debrief and Reflection Meetings Completed by District
District A	4.0	3.0	1.2
District B	2.9	1.9	1.2
District C	2.5	1.0	0.0

Source. Interviews with PLI coaches and coaching logs in spring 2023.

Note. We calculated the average rating reported by coaches in each district on a scale of 1 to 4 (not at all, to a limited extent, somewhat, and to a great extent) when describing the extent to which the principal ensures that teachers have a common weekly meeting time at least three times per month.

To further explore the hypothesis about the role of school leader support for the program, we also examined variation in implementation at the school level, specifically in Year 2. As shown in Exhibit 19, schools where PLI coaches perceived the highest level of principal support for teacher common meeting time generally completed more learning team cycles and one-on-one debrief or reflection meetings than schools with perceived lower principal support in Year 2.

⁴ As described earlier in the Implementation Fidelity for the Program Components section, the expected level of implementation for individual instructional coaching was that teachers would complete two rounds of coaching per year, and each round of coaching would include a planning meeting, two or more video debriefs, and a reflection meeting.

Exhibit 19. Variation by School in Year 2 Principal Support for Common Meeting Time, Average Number of Learning Team Cycles Completed, and Average Number of One-on-One Meetings

School	Year 2 (2022–23) in the Three 2-Year Districts		
	Extent to Which Principal Ensures That Teachers Have Common Meeting Time (4 is “to a great extent”): Coach Report	Average Number of Learning Team Cycles Completed by School	Average Number of One-on-One Debrief and Reflection Meetings Completed by School
School A	4	3	1.3
School B	4	3	1.0
School C	4	2	1.3
School D	4	2	3.3
School E	3	2	0.0
School F	3	2	0.0
School G	2	1	1.5
School H	2	1	0.0
School I	2	0	0.0
School J	1	4	1.0
School K	1	1	0.0
School L	*	1	0.0

Source. Interviews with PLI coaches and coaching logs in spring 2023.

Note. * means the school was not included in the analysis of coach interview data. We calculated the average rating reported by coaches in each school on a scale of 1 to 4 (not at all, to a limited extent, somewhat, and to a great extent) when describing the extent to which the principal ensures that teachers have a common weekly meeting time at least three times per month.

Impact Evaluation

In this section, we describe the design of the PLI impact evaluation and report findings on the impacts of PLI on teacher and student outcomes. These findings are associated with RQs 5–7 of the overall evaluation.

Design of the Impact Evaluation

The impact evaluation of the PLI program was based on a blocked-cluster RCT, in which participating schools were randomly assigned to the treatment and business-as-usual (BAU) control conditions within

target grades (Grades 4 and 5) and within districts. For two of the four participating districts, we further grouped schools into multiple random assignment blocks based on school characteristics that may be associated with student achievement, including (1) departmentalization status, (2) average proficiency rates across ELA and math at the target grade level, and (3) enrollment size and school demographic composition. Blocking on these school characteristics was likely to improve the statistical power of the impact estimates. Teachers in control schools were subject to their districts' BAU professional development requirements and opportunities; they did not have access to PLI activities or resources.

Sample

School sample. The impact evaluation included 32 schools total (17 treatment and 15 control), randomly assigned within nine blocks across four school districts and four states. All study schools were elementary schools serving high proportions of high-need students, each with three or more teachers in the target grade (Grade 4 or 5). We considered a school as serving a high portion of high-need students if (1) it was eligible for a Title 1 schoolwide program, or (2) at least 50% of students in the school were eligible for free or reduced-price lunch based on the 2019–20 Common Core of Data (CCD), the most recent CCD data available at the time of school recruitment. In addition, to be eligible for the study, a school must have already set aside time for teachers in the target grade to meet weekly as a team or be willing to do so.⁵ After recruiting the final sample of participating districts and schools, we conducted random assignment of schools within each district in the late spring and early summer of 2021, prior to the implementation of the summer PLI workshop in each district.

Impact of PLI on Three Types of Outcomes—Teachers' Quality of Classroom Practice, Teachers' Self-Efficacy, and Students' Standardized Test Scores in ELA and Math

We found no effect of PLI on teacher outcomes, including self-efficacy and classroom instruction.

PLI increased students' scores on standardized assessments in ELA, but had no effect on math scores. The effect on ELA was particularly strong for English learners, who were the only population to experience a significant increase in math scores.

⁵ We had originally also required that teachers in schools eligible for this study must not be departmentalized. However, we decided to relax this requirement because we found many elementary schools were shifting to a departmentalized structure.

Teacher sample. The sample of study teachers initially only included all teachers in the target grade in study schools at the time of random assignment who consented to participate in the study. However, because teacher rosters were not always finalized at the time of randomization, and grade-level assignments were often adjusted over the summer, some teachers joined the target grade at participating schools after randomization. Given that teachers’ decisions to join study schools were unlikely to be affected by the PLI program—particularly for teachers who joined study schools at the beginning of PY1—we assumed that teachers who joined before the baseline survey would not introduce a risk of bias to the RCT and decided to include those early joiners in the teacher impact sample. The teachers included in the study therefore include all teachers who consented to participate in the study at the time of random assignment who were still at study schools at the time of baseline data collection, plus joiners who joined study schools before the baseline survey and completed the baseline survey. We refer to this group of teachers as “the teacher sample.” The teacher sample included 155 teachers (87 treatment and 68 control).

Student sample. The sample for impact analyses of student outcomes measured at the end of each program year consisted of all students enrolled in the target grades at study schools at the time of the spring state assessments. According to the current What Works Clearinghouse (WWC, 2022) standards, the risk of bias due to student joiners is low for a school-level RCT of a teacher professional development program.

Measures and Data Sources

Teacher Outcome Measures

Quality of classroom practice. The primary teacher outcome for this impact evaluation is the quality of teachers’ classroom practice. For the baseline measure of teacher classroom practice, we recorded one lesson per teacher in the early fall of PY1 (2021).⁶ For teachers who taught both ELA and math, we randomly selected one subject to observe at baseline. For teachers who taught only math or only ELA, we observed their instruction in the subject taught. Teachers who did not teach ELA or math (about 9.7% of all study teachers) indicated teaching either science, social science, or both. For those teaching only one subject, we observed their instruction in that subject. For the remaining teachers—who taught both science and social science—we randomly selected a science or social studies lesson to observe at baseline.

For the outcome measure of teacher practice, we recorded classroom observations of two lessons in the spring of PY2 for each study teacher. For teachers who taught both ELA and

⁶ Although the PY1 classroom observation took place after random assignment, treatment teachers had not received any real treatment other than the 1-day preparatory workshop by that point in time, which by itself was unlikely to affect the quality of teachers’ classroom practice. Therefore, it is reasonable to consider the fall of PY1 observation a “baseline” measure of teacher practice.

math, one math lesson and one ELA lesson were observed. For teachers who taught both science and social science, one science lesson and one social science lesson were observed. For teachers who taught only one subject, we randomly selected two lessons to observe. The scheduling of all observation recordings depended on school schedules and the availability of data collectors (district employees) to minimize classroom disruptions, and every effort was made to ensure that the observation schedules at treatment and control schools overlapped. Although recordings were collected by district staff, observation recordings were coded by members of the research team.

The primary measure of teachers' classroom practice is the overall score from the Classroom Assessment Scoring System-Upper Elementary (CLASS-UE), an instrument that has rich evidence of reliability and validity (Goe et al., 2008). Certified CLASS observers at AIR coded the videos, blind to the study condition. To clearly distinguish the measure of classroom practice and the measure of student engagement, we created the CLASS-UE overall score based on the 10 dimensions within the Emotional Support, Classroom Organization, and Instructional Support domains, excluding the one dimension in the Student Engagement domain.

A subset (10%) of the lessons were each coded by seven independent coders to assess reliability. The inter-rater reliability (IRR) for the overall CLASS-UE score and the four domain scores estimated using Stata's "icc" command are presented in Exhibit 20; higher IRR values indicate lower measurement error. The IRRs for the CLASS-UE overall score and the Instructional Support domain score both meet WWC's (2022) threshold for acceptable IRR (at least 0.50) for outcome measures. However, the IRRs for the other three domain scores—Emotional Support, Classroom Organization, and Student Engagement—fall below the threshold of acceptable reliability. Findings from analyses of those domain scores, therefore, need to be interpreted with caution given the lack of sufficient reliability.

Exhibit 20. Inter-Rater Reliabilities of Measures of Classroom Practice

Classroom Practice Measure	Inter-Rater Reliability (IRR)
CLASS-UE Overall Score	0.511
CLASS-UE Domain Scores:	
Emotional Support	0.489
Classroom Organization	0.042
Instructional Support	0.513
Student Engagement	0.189

Note. IRR estimates are based on ratings from seven raters on six videos. CLASS-UE = Classroom Assessment Scoring System-Upper Elementary.

Survey-based teacher outcomes. In addition to classroom practice, we assessed the impact of PLI on a key intermediate teacher outcome as specified in PLI's theory of change: teachers' self-efficacy.⁷ For this outcome, we adapted the Teachers' Sense of Self-Efficacy Scale (short form) developed by Tschannen-Moran and Woolfolk-Hoy (2001).⁸ We gathered baseline data on teachers' self-efficacy with a short survey administered in the early fall of PY1 to all teachers in the RCT sample, and we included the self-efficacy measures in the spring teacher survey conducted in each program year. For the impact study, we designated the PY2 teachers' self-efficacy overall scale as the main measure of teachers' self-efficacy and considered the spring PY1 measure a supplemental outcome. We found the measure to be reliable, with a Cronbach's alpha of 0.92.

In addition to self-efficacy, surveys of teachers at baseline and each spring included questions that allowed us to measure supplemental teacher outcomes, such as teachers' feelings of collegial support, data usage, job satisfaction, and perception of performance. The survey items used to measure teacher self-efficacy and other teacher outcomes, as well as factor loadings and values of Cronbach's alpha, are presented in Appendix Exhibit E1. Appendix E also provides results of analyses that demonstrate the psychometric validity and invariance of the teacher survey outcome measures across treatment conditions (Appendix Exhibit E2) as well as correlations across teacher survey outcome measures (Appendix Exhibit E3).

Student Outcome Measures

Academic achievement. The key student outcome for the PLI program is academic achievement. Measures of student achievement were based on students' scores on standardized state assessments in mathematics and ELA collected from the study districts, standardized within subjects and grades based on districtwide means and standard deviations.⁹

Student engagement. We examined the CLASS-UE Student Engagement domain score based on the video-recorded lessons from the spring of PY2 as a supplemental student outcome measured at the classroom level. The CLASS-UE Student Engagement dimension score captures the degree to which all students in the class were focused on and participating in the learning activity presented or facilitated by the teacher.

⁷ The theory of change identifies some additional intermediate teacher outcomes—teacher knowledge and skill. We did not measure these outcomes because no reliable measures of knowledge and skill related to the FFT are available.

⁸ See <https://www.statisticssolutions.com/teachers-sense-of-efficacy-scale-tses/> for details about the measure.

⁹ Though some study teachers taught only math, only ELA, or neither, we focus on both of these outcomes because PLI was randomized at the school level and therefore is theorized to improve student performance in all subjects.

Covariates/Potential Moderators

To adjust for chance imbalance in certain sample characteristics and to improve the precision of impact estimates, we included baseline outcome measures as well as teacher or student background characteristics as covariates in the impact analyses as appropriate. Impact analyses of teacher outcomes included the following teacher characteristics as covariates: a baseline classroom observation score or survey measure (depending on which outcome is included in the model), years of teaching experience,¹⁰ and whether the teacher had a graduate degree based on the baseline teacher survey. Impact analyses of student outcomes included the following student-level covariates that were provided by participating school districts: gender, underrepresented minority status, eligibility for free or reduced-price lunch, English learner status, special education status, and baseline (prior-year) achievement scores. These covariates also were potential moderators in the differential impact analyses addressing RQ7.

Sample Sizes and Attrition

Sample sizes and attrition were determined for each outcome in each program year that the outcome was observed. Given that this study is a cluster/school-level RCT, for each outcome in a given program year, we first examined the school-level attrition rate for the outcome and then the individual-level (i.e., teacher or student level) attrition rate in nonattrititing schools.

Sample Size and Attrition for Teacher Outcomes

Because data on the teacher outcomes examined in this study came from two different sources—classroom observations and teacher surveys—we calculated attrition rates separately for teacher outcomes based on each data source. To determine school-level attrition rates for teacher outcomes, we considered a school attritted if (1) the school decided to discontinue participation in the study or (2) if none of the study teachers in the school had the relevant outcome data. The school-level attrition rate for each teacher outcome was calculated as the percentage of study schools in the original randomized sample that had attritted by the end of the given program year. We then calculated the teacher-level attrition rate for each outcome as the percentage of study teachers in nonattrititing schools who were excluded from the impact analysis due to missing outcome or baseline data.¹¹

Though we calculate attrition for classroom observations and teacher surveys separately, there was school-level attrition in each project year that affected both teacher outcomes. Specifically, at the beginning of PY1, three treatment schools (from two school districts) withdrew from the study. In addition, two randomization blocks with one school per condition had to be excluded

¹⁰ We identified three relevant ranges of years of experience based on the distribution of years of experience reported by teachers: 0–5 years, 6–10 years, and 11 or more years.

¹¹ We removed a small number of teachers for some outcomes from the analysis sample to achieve baseline equivalence. These removed teachers also are considered attritted and treated as such in our calculations of attrition in this section.

from the PY2 teacher impact analyses. This occurred because, in both randomization blocks, one of the two schools in the block no longer had any study teachers teaching the target grade in the spring of PY2. As a result, both blocks were left with only one nonattriting study school by the end of PY2 and had to be excluded from PY2 teacher impact analyses because treatment-control comparisons could not be made within these blocks.

Following, we discuss school-level and teacher-level attrition for the two types of teacher outcomes—classroom observation measures and teacher survey measures—separately. Although conceptualized as a student outcome, student engagement was measured based on observation data at the teacher level in this study. Therefore, we present information about sample size and attrition for student engagement together with the observation-based teacher outcomes. We summarize the sample sizes and attrition rates in Exhibit 21.

Classroom Observation Measures

In addition to the school-level attrition mentioned above, classroom observations were not collected for one remaining control school in PY2 because the remaining teacher in that school was out on leave. Therefore, attrition among schools for the classroom observation measures included five treatment schools and three control schools. As we illustrate in Exhibit 21, the resulting overall school-level attrition rate for teacher observation measures was 25%, and the differential attrition rate was 9.4%. This combination of overall and differential attrition rates exceeds the WWC thresholds for low attrition based on the optimistic and cautious attrition boundaries.

At the study schools that provided at least one spring classroom observation in PY2 (i.e., nonattriting schools), 120 teachers (64 treatment and 56 control) consented to participate and/or had baseline data. These 120 teachers make up the reference sample for our calculation of attrition. At the end of PY2, 66 teachers (34 treatment and 32 control) had classroom observations, leading to an overall individual-level attrition rate of 45% and a differential attrition rate of 4% for teacher observation outcomes. This combination of overall and differential attrition does not exceed the WWC optimistic boundary for high differential attrition. However, teacher observation measures are considered to have high attrition due to the high rate of cluster-level attrition.

Teacher Survey Measures

There was no additional school-level attrition for teacher survey measures beyond what was described above. Therefore, attrition among schools for the classroom observation measures included five treatment schools and two control schools. For teacher survey outcome measures in PY2, the resulting overall school-level attrition rate was 21.9%, and the differential attrition rate

was 16.1%. This combination of overall and differential attrition rates exceeds the WWC thresholds for low attrition based on the optimistic attrition boundary.

At the study schools where at least one teacher provided an endline survey (i.e., nonattrititing schools), 124 teachers (64 treatment and 60 control) were present in nonattrititing schools at baseline data collection. These 124 teachers make up the reference sample for our calculation of attrition. At the end of PY2, 60 teachers (30 treatment and 30 control) provided responses to the survey measure of self-efficacy, leading to an overall individual-level attrition rate of 51.6% and a differential attrition rate of 3.1% for our primary teacher survey outcome.¹² Though this combination of overall and differential teacher-level attrition does not exceed the WWC optimistic boundary for high differential attrition, teacher observation measures are considered to have high attrition due to the high rate of cluster-level attrition.

Exhibit 21. Sample Sizes and Attrition Rates at the School and Teacher Levels for Program Year 2 Teacher Outcomes

Program Year 2 Teacher Outcomes	Treatment		Control		Overall Attrition Rate	Differential Attrition Rate
	# Randomized	# Analytic Sample	# Randomized	# Analytic Sample		
Observation Measures: CLASS-UE Overall Score and Domain Scores						
Schools	17	12	15	12	25.0%	9.4%
Teachers	64	34	56	32	45.0%	4.0%
Survey Measures: Self-Efficacy						
Schools	17	12	15	13	21.9%	16.1%
Teachers	64	30	60	30	51.6%	3.1%

Note. One of the CLASS-UE domain scores (i.e., Student Engagement score) is considered a student outcome in this study even though it was based on teacher observation data. CLASS-UE = Classroom Assessment Scoring System-Upper Elementary.

Sample Size and Attrition for Student Achievement Outcomes

Because students’ achievement test scores were collected from administrative data systems that provided data for all study schools, student achievement outcomes did not have any school-level attrition. As a result, the analytical sample in each year includes all students in study schools who completed the standardized achievement test at the end of that school year. Following WWC guidance on the use of administrative data to measure student outcomes

¹² Attrition rates for supplemental survey outcomes are dependent on the number of teachers who responded to those survey scales. A full table of attrition rates for supplemental survey outcomes is presented in Appendix F.

(WWC, 2022, p. 51), missing outcome data on student achievement using administrative data are assumed to be low.

Analytic Methods

Intent-to-treat (ITT) analyses. Our main impact analyses were ITT analyses, which estimated the impacts of being randomly assigned to implement the PLI program on teacher and student outcomes. Because PLI was a 2-year program, our main impact analyses focused on outcomes at the end of PY2; impacts on teacher survey outcomes and student achievement at the end of PY1 were assessed as exploratory analyses. This section describes in detail the specifications of the analytic model for each type of outcome.

Impacts on teacher outcomes (RQ5). RQ5 addresses the impacts of PLI on two primary teacher outcomes: the quality of classroom practice and teacher self-efficacy. To assess the impact of PLI on the quality of teachers' classroom practice as measured with classroom observations conducted at the end of PY2, we used the following three-level model:

$$\text{Level 1 (lesson level): } Y_{ijk} = \pi_{0jk} + \pi_{1jk}MATH_{ijk} + \pi_{2jk}OTHER_{ijk} + \varepsilon_{ijk} \quad (1)$$

where

Y_{ijk} is the CLASS-UE overall score for lesson i ($i = 1$ or 2) taught by teacher j in school k ;

$MATH_{ijk}$ is a dummy indicator for the subject area of lesson i taught by teacher j in school k , coded 1 if the lesson is a math lesson and 0 otherwise, grand-mean centered;

$OTHER_{ijk}$ is a dummy indicator for the subject area of lesson i taught by teacher j in school k , coded 1 if the lesson is a science or social studies lesson and 0 otherwise, grand-mean centered;

π_{0jk} is the average CLASS-UE overall score across lessons for teacher j in school k ;

π_{1jk} is the difference in the CLASS-UE overall score between the mathematics lesson and the ELA lesson observed for teacher j in school k ;

π_{2jk} is the difference in the CLASS-UE overall score between the other (science or social studies) lesson and the ELA lesson observed for teacher j in school k ; and

ε_{ijk} is a random error associated with lesson i taught by teacher j in school k .

$$\text{Level 2 (teacher level): } \pi_{0jk} = \beta_{00k} + \beta_{01k}W_{jk} + r_{0jk} \quad (2)$$

$$\pi_{1jk} = \beta_{10k} \quad (3)$$

$$\pi_{2jk} = \beta_{20k} \quad (4)$$

where

W_{jk} is a vector of background characteristics for teacher j in school k (e.g., years of teaching experience, whether the teacher had a graduate degree, and baseline CLASS-UE overall score), grand-mean centered;

β_{00k} is the average CLASS-UE overall score across all study teachers in school k ;

β_{01k} is the relationship between teachers' background characteristics and their average CLASS-UE overall scores in school k ;

β_{10k} is the average difference in the CLASS-UE overall score between the mathematics lesson and the ELA lesson observed across all study teachers in school k ;

β_{20k} is the average difference in the CLASS-UE overall score between the other (science or social studies) lesson and the ELA lesson observed across all study teachers in school k ; and

r_{0jk} is a random error associated with teacher j in school k .

$$\text{Level 3 (school level): } \beta_{00k} = \sum_{m=1}^7 \gamma_{000m}B_{mk} + \gamma_{001}PLI_k + u_{00k} \quad (5)$$

$$\beta_{01k} = \gamma_{010} \quad (6)$$

$$\beta_{10k} = \gamma_{100} \quad (7)$$

$$\beta_{20k} = \gamma_{200} \quad (8)$$

where

B_{mk} , $m = 1 \sim 7$, is a set of dummy indicators for the seven random assignment blocks in the study sample;

PLI_k is a dummy indicator for treatment status (coded 1 if school k is in the treatment group and 0 otherwise);

γ_{000m} , $m = 1 \sim 7$, is a set of block fixed effects, representing the average CLASS-UE overall score for teachers in the control schools in each block;

γ_{001} is the pooled within-block difference between treatment and control schools in teachers' CLASS-UE overall scores;

γ_{010} is the average relationship between teachers' background characteristics and their CLASS-UE overall scores across all schools;

γ_{100} is the average difference in the CLASS-UE overall score between the mathematics lessons and the ELA lessons observed across all schools;

γ_{200} is the average difference in the CLASS-UE overall score between the other (science or social studies) lessons and the ELA lessons observed across all schools; and

u_{00k} is a random error associated with school k .

The estimate of primary interest from the Level 3 model is γ_{001} , which represents a precision-weighted overall ITT effect on teachers' CLASS-UE overall scores across all blocks.

To assess the ITT effects of PLI on teachers' self-efficacy and other survey-based teacher outcomes, we used a two-level model (teachers/schools) with a set of equations specified similarly to Equations (2), (4), and (5).

Impacts on student outcomes (RQ6). The ITT effects of PLI on student achievement were estimated separately for mathematics and ELA based on a two-level model specified as follows:¹³

$$\text{Level 1 (student level): } Y_{sk} = \pi_{0k} + \pi_{1k}X_{sk} + \varepsilon_{sk} \quad (9)$$

where

Y_{sk} is the test score in a given subject for student s in school k , standardized based on the district mean and standard deviation;

X_{sk} is a vector of student background characteristics (e.g., prior achievement in the same subject as the outcome, gender, race/ethnicity, eligibility for free or reduced-price lunch, English learner status, and special education status), grand-mean centered;

π_{0k} is the average test score across students in school k ;

π_{1k} is the relationship between student characteristics and achievement for students in school k ; and

ε_{sk} is a random error associated with student s in school k .

$$\text{Level 2 (school level): } \pi_{0k} = \sum_{m=1}^9 \beta_{00m}B_{mk} + \beta_{01}PLI_k + u_{0k} \quad (10)$$

$$\pi_{1k} = \beta_{10} \quad (11)$$

Where

β_{00m} , $m = 1 \sim 9$, is a set of block fixed effects, representing the average achievement of students in control schools in each block;

PLI_k is a dummy indicator for treatment status (coded 1 if school k is in the treatment group and 0 otherwise);

β_{01} is the pooled within-block difference between treatment and control schools in student achievement;

β_{10} is the average relationship between student characteristics and achievement across all students across all schools; and

u_{0k} is a random error associated with school k .

¹³Ideally, we would like to estimate the impact on student outcomes using a three-level model that also accounts for the nesting of students within teachers. However, the administrative data we obtained from study districts do not allow for clean matching of students to their teachers.

The estimate of primary interest from the above model is β_{01} , which represents a precision-weighted overall ITT effect on student achievement in a given subject.

The impact of PLI on student engagement, based on the CLASS-UE Student Engagement dimension score, was estimated using the same model used to assess the PLI impact on the quality of teachers' classroom practice, as measured by the CLASS-UE overall score described earlier.

Differential Impact Analyses (RQ7)

In addition to the main impact analyses, we conducted differential impact analyses to examine whether the impact of PLI on student achievement varied by student background characteristics, including prior achievement (RQ7). Each differential impact analysis was conducted by incorporating a cross-level interaction between treatment status and a potential moderator into the main student impact model described earlier: prior test scores in ELA, prior test scores in math, special education status, English learner status, low-income status, and underrepresented minority status.

Handling of Missing Data

We handled missing data according to acceptable methods as specified in the WWC Group Design Standards (Version 5.0). For both teacher and student outcome measures, we excluded cases with missing outcome data and imputed missing values on covariates—excluding baseline measures of the outcomes—using the dummy variable adjustment approach.

Findings

This section presents findings about the impacts of PLI on teacher and student outcomes in PY2 as well as findings about the differential impact of PLI on student achievement in PY2. Findings about the impacts of PLI in PY1 are presented in Appendices G and H.¹⁴

Impacts on Teacher Outcomes (RQ5)

Exhibit 22 presents the estimated impacts of PLI on teacher outcomes at the end of PY2. The first two panels of the exhibit show the estimated effect of PLI on the overall CLASS-UE score and domain subscores, as measured by classroom observations. We found no statistically significant differences in the quality of classroom practice between the two groups.¹⁵

¹⁴ 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 F.

¹⁵ Baseline equivalence analysis results presented in Exhibit F2 in Appendix F indicate that differences in baseline measures of the outcomes between the two study groups were all nonsignificant and in the “adjustable” range, according to the WWC (2022) standards, with effect sizes ranging from -0.17 to 0.23.

Exhibit 22. Impacts of Professional Learning with Impact on Teacher Outcomes in Year 2

Outcome Measure	Treatment Group		Control Group		Treatment – Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
CLASS-UE Overall Score	3.89	0.63	3.89	0.59	0.00	0.12	0.00	0.999
CLASS-UE Domain Scores								
Emotional Support	4.03	0.84	4.02	0.87	0.01	0.16	0.01	0.956
Classroom Organization	4.60	0.51	4.57	0.43	0.03	0.09	0.06	0.718
Instructional Support	3.39	0.78	3.42	0.73	-0.03	0.15	-0.04	0.826
Student Engagement	4.82	1.33	4.70	0.80	0.11	0.20	0.10	0.578
Teacher Survey Outcome								
Self-Efficacy	6.96	0.95	6.71	1.07	0.25	0.24	0.25	0.288

Note. The sample size for CLASS-UE = 34 teachers (63 observations) for the treatment group and 32 teachers (64 observations) for the control group. The sample size for the Student Engagement domain = 31 teachers (60 observations) for the treatment group and 27 teachers (57 observations) for the control group. The sample size for Teacher Self-Efficacy = 30 teachers for the treatment group and 30 teachers for the control group. The treatment group means are unadjusted means; the control group means are adjusted means computed based on the unadjusted treatment group means and the estimated group mean difference. Effect sizes were computed using Hedges' *g*, as recommended by the What Works Clearinghouse (2022). CLASS-UE = Classroom Assessment Scoring System-Upper Elementary; SD = standard deviation.

Exhibit 22 also indicates that PLI did not have a significant impact on our primary survey measure, teacher self-efficacy. Supplemental findings for additional survey-based teacher outcomes from PY1 and PY2 are presented in Appendix G. In almost every supplemental measure, we found no significant impact of PLI on survey outcomes.

Impacts on Student Outcomes (RQ6)

Exhibit 23 presents findings about the impacts of PLI on students' fourth-grade ELA and math test scores. Although PLI did not have a significant impact on students' fourth-grade math scores, it had a significant positive impact on students' fourth-grade ELA test scores, with an effect size of 0.09 standard deviations.¹⁶ Using the WWC's guidance on computing an improvement index, this effect size suggests that assigning an average student from a control school to a school that utilizes PLI will increase that student's percentile rank in fourth-grade ELA achievement by 3.6 percentiles (WWC, 2022).

¹⁶ Baseline equivalence analysis results presented in Exhibit F4 in Appendix F indicate that differences in prior achievement scores between the two study groups were all nonsignificant and in the "adjustable" range, according to the WWC (2022) standard, with effect sizes ranging from -0.002 to 0.036.

Exhibit 23. Impacts of Professional Learning with Impact on Fourth-Grade English Language Arts and Math Achievement in Year 2

Student Achievement Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
Fourth-Grade English Language Arts (ELA) Achievement	-0.07	0.97	-0.16	0.99	0.08	0.04	0.09	0.031
Fourth-Grade Math Achievement	-0.08	0.96	-0.13	0.96	0.05	0.05	0.05	0.319

Note. Sample size ELA achievement = 2,817 (1,492 treatment and 1,325 control). Sample size math achievement = 2,832 (1,496 treatment and 1,336 control). The treatment group means are unadjusted means; the control group means are adjusted means computed based on the unadjusted treatment group means and the estimated group mean difference. Effect sizes were computed using Hedges’ *g*, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Differential Impacts on Student Achievement (RQ7)

In addition to the main impacts of PLI on student achievement, we examined whether PLI impacts on student achievement varied by the following student baseline characteristics: prior achievement, eligibility for free or reduced-price lunch, special education status, English learner status, and underrepresented minority status (see Exhibit 24).

Exhibit 24. Differential Impacts of Professional Learning with Impact (PLI) on Fourth-Grade English Language Arts and Math Achievement in Year 2, by Student Characteristics

Student Characteristics	Impact Estimate for X = 0	Impact Estimate for X = 1	Difference in Impact	Standard Error	P-Value
Outcome: Fourth-Grade English Language Arts (ELA) Achievement					
Prior achievement	0.08	0.02	-0.05	0.03	0.055
Eligibility for free or reduced-price lunch	0.06	0.12	0.06	0.06	0.315
English learner status	0.06	0.32	0.26	0.09	0.006
Special education status	0.07	0.18	0.11	0.07	0.135
Underrepresented minority status	0.11	0.07	-0.04	0.05	0.466
Outcome: Fourth-Grade Math Achievement					
Prior achievement	0.05	0.08	0.02	0.03	0.404
Eligibility for free or reduced-price lunch	0.06	0.04	-0.02	0.06	0.730

Student Characteristics	Impact Estimate for X = 0	Impact Estimate for X = 1	Difference in Impact	Standard Error	P-Value
English learner status	0.03	0.22	0.19	0.09	0.031
Special education status	0.06	0.02	-0.04	0.07	0.577
Underrepresented minority status	0.07	0.04	-0.03	0.05	0.531

Note. Sample size ELA achievement = 2,817 (1,492 treatment and 1,325 control). Sample size math achievement = 2,832 (1,496 treatment and 1,336 control). 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, and non-White students; X = 0 otherwise. The listed effect of PLI when the trait is equal to 0 comes from the estimated coefficient on treatment status; the listed effect of PLI when the trait is equal to 1 is the sum of the estimated coefficient on treatment status and the estimated coefficient on the interaction between treatment status and the listed trait. The listed difference is the estimated coefficient on the interaction, and the *p*-value is the *p*-value on that interaction.

As shown in Exhibit 24, there were no statistically significant differential impacts of PLI on fourth-grade test scores by prior achievement, eligibility for free or reduced-price lunch, special education status, or underrepresented minority status. However, the impact of PLI on both fourth-grade ELA scores and math scores for English learners were 0.26 and 0.19 standard deviations larger, respectively, than the impacts for non-English learners. Both differential impacts by English learner status were statistically significant.¹⁷

¹⁷ Exhibit H1 in Appendix H presents the analysis of student outcomes for each subgroup—English learners and non-English learners—separately and includes estimated treatment effects and sample sizes for both subsamples.

Cost Study

Introduction to the Cost Study

This section describes the methods and results of an economic evaluation of the overall cost and cost-effectiveness of the PLI program. This evaluation relied on multiple sources of data, including teacher surveys and activity logs, in the Whetstone platform. These sources contained information about resources used to implement the PLI program, such as costs of materials, teacher time, coach time, and program specialist time. To estimate costs, we applied the ingredients approach (Levin et al., 2018), in which costs are estimated based on the personnel and nonpersonnel resources 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 (e.g., school staff spending time on the PLI program have given up their opportunity to spend time on another program or task). The goal of this evaluation is to help future adopters determine if PLI is an affordable program for their school or school district, what the necessary resources and costs are to implement the program, and how effective the program is in improving student outcomes in a school relative to the cost of the program.

This evaluation focuses on the assessment of the overall cost of PLI. The assessment of the overall cost is intended to estimate the full cost of implementing PLI in two distinct ways: (1) as a new, additional teacher professional development program or (2) as a replacement of an existing professional development program. If a district or school were to adopt PLI as a new, additional teacher professional development and learning program without changing any of its other programming, the cost estimates provided under this analysis describe the expected additional cost to the school or district to implement the program. In some cases, however, the adopting school or district may already have coaches or professional development activities and could repurpose those resources to pay for PLI instead. In these cases, the incremental cost of adopting PLI would more closely resemble the cost of implementing PLI as an augmentation of an existing professional development program.

Although it was our intention to also include an assessment of the cost-effectiveness of PLI, results from the impact evaluation and challenges with data meant that cost-effectiveness analyses of the teacher and student outcomes were either not possible or likely to be inaccurate. As a result, this analysis does not include a cost-effectiveness estimate. We discuss the reasons behind this decision in the Cost Effectiveness subsection.

PLI Resources and Program Components

Costs examined in this evaluation are separated into four components: teacher time, coach time, program specialist time, and materials. Exhibit 25 summarizes the program activities within each of the four components that are intended to contribute to the outcomes of the PLI program.

Exhibit 25. Professional Learning with Impact Program Components

Program Component	Description
Teacher Time	Teacher workshops All workshops were held during the first year of program implementation. Teachers were offered three workshops (summer 2021 [7 hours], fall 2021 [3.5 hours], and winter 2022 [6.5 hours]). At each workshop teachers worked together with coaches to learn about the Framework for Teaching Clusters, the Teacher Learning Team (TLT) cycle, and coaching round process.
	One-on-one meetings with coaches - Preparation: recording instruction, uploading video to Whetstone, commenting on video, responding to coach's comments, and any other preparation - Attendance: total time spent attending one-on-one meetings with coach
	TLT meetings - Preparation: any reported time spent preparing for meetings - Attendance: total time spent attending TLT meetings
Coach Time	Coach workshops All workshops were held during the first year of program implementation. Coaches participated in the teacher workshops as well as two additional workshops in summer 2021 (7 hours) and a fall refresher workshop in 2022 (3 hours). At each workshop coaches worked together with program specialists to build skills to facilitate the TLT cycle and coaching rounds.
	One-on-one meetings with teachers - Preparation: any reported time spent preparing for one-on-one meetings, such as reviewing videos of teachers' lessons - Attendance: total time spent attending one-on-one meetings with all assigned teachers (includes time spent waiting for a teacher who did not show up for meeting) - Logging: any reported time spent logging information about one-on-one meetings (occurs after meeting)
	TLT meetings - Preparation: any reported time spent preparing for meetings - Attendance: total time spent attending TLT meetings
	Time with program specialists - Total time spent attending monthly individual support calls with program specialists - Total time spent attending all-coach professional learning community (PLC) calls

Program Component	Description
Program Specialist Time	Time with coaches • Total time spent attending monthly individual support calls with coaches • Total time spent attending all-coach PLC calls
	Time spent on implementation and other program activities • Monthly implementation support: monitoring progress, providing feedback to partner leadership teams, and resolving team issues • PLI specialist calls: Calls held for program specialists on a regular basis by partner leadership to discuss upcoming activities, evaluate progress, resolve issues as a group, and so on • Data entry: entering coaching session data into Airtable
Materials	• Subscription to Whetstone online platform • Lesson books for all teachers and coaches

Data Sources

As part of data collection for both the implementation and impact evaluation, we gathered information to support an assessment of the costs associated with PLI. The primary ingredient measured in the cost analysis is time spent by teachers, coaches, and program specialists. We gathered the relevant data as follows:

- **Coach time preparing for and attending all meetings.** Activity data from the Whetstone platform provide information on the amount of time that coaches spent preparing for and attending one-on-one meetings with teachers, TLT meetings with teachers, one-on-one support meetings with program specialists, and monthly group support meetings with program specialists.
- **Teacher time attending meetings.** Whetstone also provides information on time spent by PLI teachers attending one-on-one meetings and TLTs.
- **Teacher time preparing for meetings.** Questions on teacher surveys allowed for the calculation of the average amount of time spent by PLI and BAU teachers on preparing for and attending one-on-one meetings and TLTs.
- **Teacher and program specialist time at workshops.** Attendance data collected during the program workshops allowed us to estimate the time that teachers and coaches spent attending and participating in PLI workshop trainings.
- **Additional program specialist time.** Time spent by program specialists with coaches came from activity logs from coaches using the Whetstone platform. Time spent by program specialists beyond their interactions with coaches came from monthly budgets constructed

by Learning Forward that estimated the number of hours required for program specialists for various activities in the coming month.¹⁸

- **Materials.** Members of the implementation team provided information on the cost of materials provided to teachers to implement PLI, including lesson books and subscriptions to the Whetstone platform.

We calculate the cost of time for coaches and teachers by multiplying the total number of hours spent by coaches and teachers by the national market rate for their hourly wages. We present more details about the calculation of these compensation rates in Appendix I.

Costs Per Teacher

We present all costs in terms of costs per teacher to make cost estimates more useful for schools and districts considering adopting the PLI program. Doing so allows adopting schools and districts to better estimate the total expected cost of adoption given the number of teachers in that school or district. This approach first excludes all teachers from schools and districts that attrited from the study due to a lack of PLI implementation. The PLI program was implemented over two academic years; four districts participated during the first year of implementation, but only three of these continued full implementation in the second year. As a result, the number of participating teachers in PY1 and PY2 differ, and information on time spent by teachers in PY2 is only available in three districts. We therefore calculate costs per teacher separately for each implementation year, including teachers from four districts in PY1 and teachers from the three remaining districts in PY2. The number of teachers used to determine costs per teacher in each implementation year is based on the number of teachers that (1) were assigned to participate in PLI from the RCT and (2) were still teaching at a study school in a target grade in the year the costs were measured.

Although the costs presented represent the average cost per teacher across all teachers assigned to the PLI program, the intensity of participation varies across teachers. Though some teachers report regularly meeting with coaches and attending nearly every TLT meeting, other teachers report little to no engagement with coaches and TLTs. As a result, estimates of teacher time—and in turn teacher cost—are near zero or equal to zero for some teachers. We include these low-intensity participants in our estimates of cost per teacher because doing so better represents a scenario in which teachers are given access to, but are not forced to engage with, the PLI program. One important caveat to this calculation is that participation intensity may have been affected by the COVID-19 pandemic, particularly in the first year of implementation. If a school or district explicitly incentivizes or requires participation or implements PLI in a

¹⁸ Although projected hours do not necessarily measure actual hours, we believe the two likely closely align because projections are for a short time period (the following 4 weeks) and can be informed by realized hours in the immediately preceding months.

different scenario that does not have the pressures of a pandemic, we expect the number of hours spent per teacher to increase relative to our estimates, and therefore increase the estimates of the cost per teacher.

Total Cost of PLI

In this section, we present the total cost of adopting PLI in two scenarios: (1) when a school or district does not have an existing professional development program and (2) when a school or district has an existing professional development program that includes coaches and PLCs and chooses to implement PLI instead of that program. In the first scenario, PLI can be seen as a stand-alone program. We consider this an estimate of the upper bound of the cost of adopting PLI as it represents the largest incremental cost of adopting PLI and our best estimate of the total cost of PLI. In the second scenario, PLI can be seen as a replacement for an existing program. We consider this an estimate of the lower bound of the cost of adopting PLI as it represents the incremental cost for a school or district that already relies on the coaching structure of PLI. We include this scenario as it best represents the contrast in costs between BAU and PLI schools in this study. In each scenario, we discuss the approach, limitations of the approach, and implications of the limitations for the estimate of the cost.

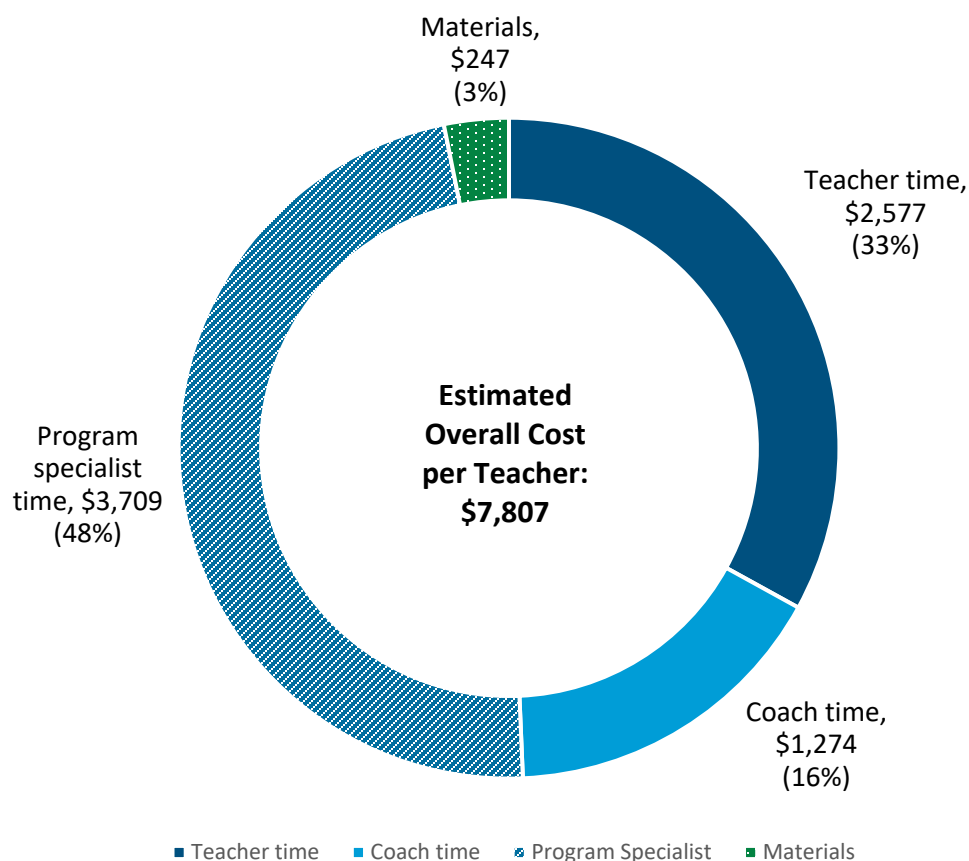
Total Cost of PLI as a Stand-Alone Program

To determine the cost of PLI as a stand-alone program, we multiplied the quantity of each ingredient listed in Exhibit 26 (i.e., time spent by staff and quantities of materials) by its price and summed the total costs. This approach accounts for all resources used to supply the primary ingredients to implement PLI and, in turn, represents the total cost of implementing PLI as a stand-alone program. Further details on how we determined prices are explained in Appendix I.

Findings

Our estimate of the overall total cost of PLI per teacher is \$7,807 over 2 years. Exhibit 26 shows how the per-teacher cost of PLI is distributed across the four program components: coaches' time, teachers' time, program specialists' time, and materials.¹⁹

Exhibit 26. Upper-Bound Overall Cost Per Teacher of the 2-Year Professional Learning with Impact Program (2023 Dollars)



Note. Costs include the sum of total costs per teacher for both years of implementation. Percentages in parentheses indicate the percentage of the total accounted for by the indicated program component.

According to our analysis, 48% of the costs incurred in running the program come from program specialist time. This is a significant share, compared with the relatively smaller allocations of 33%, 16%, and 3% toward teacher time, coach time, and materials, respectively. Although program specialists spent the least amount of time on PLI activities (e.g., PY1: 906 hours for all program specialists, 977 hours for all coaches, and 1,764 hours for all teachers), the main factor that contributes to the high proportion of cost for their time is their salaries. As

¹⁹ The present values of all costs in Year 1 (2021–22) were calculated in 2023 dollars using a 3.0% discount rate.

shown in Exhibit I1. Compensation Rates for Personnel (in 2023 Dollars) in Appendix I, the average loaned compensation rate for program specialist is \$163.00 in PY1, which is approximately 33–40% higher than coaches and teachers (\$54.31 and \$66.14, respectively).

Exhibit 27 breaks down the overall per-teacher cost across the 2 years of the project. The total costs from PY1 and PY2 indicate that total per-teacher costs in PY2 were less than half of those from PY1. Although cost per teacher dropped in PY2 across all four components, costs due to teacher time exhibited the largest percentage drop (about 63%). This decrease in cost was primarily driven by a decrease in the percentage of teachers that engaged in one-on-one coaching (from about 85% of teachers in PY1 to about 46% of teachers in PY2). The decrease in engagement with one-on-one coaching by teachers also is reflected in the cost of coach time in PY2, which fell by 51% in comparison to PY1.

Exhibit 27. Overall Per-Teacher Cost of Professional Learning with Impact in Year 1 and Year 2 (in 2023 Dollars)

Program Component	Year 1 (2021–22) (N of teachers = 62)	Year 2 (2022–23) (N of teachers = 39)	Total
#1: Teacher Time	\$1,882	\$696	\$2,577
#2: Coach Time	\$856	\$418	\$1,274
#3: Program Specialist Time	\$2,390	\$1,319	\$3,709
#4: Materials	\$132	\$115	\$247
Total	\$5,259	\$2,547	\$7,807

Generalizability Considerations

The total costs presented above rely on a few assumptions that may not apply to every district or school that chooses to adopt PLI in the future. In this section, we outline the assumptions in our analysis that are unique to this particular implementation of PLI and may not be universally applicable.

Depending on the context in which PLI is adopted, the results presented in Exhibit 27 may either overestimate or underestimate the total cost of adoption. For example, Exhibit 27 may underestimate the cost of materials, as the presented costs do not include any costs associated with facilities or additional materials, such as computers and recording equipment. Although the costs presented in Exhibit 27 likely capture most of program specialists’ time, they do not account for additional time spent by program specialists not captured during data collection, including preparation for workshops, monthly individual support calls, and all-coach PLC calls. Moreover, the costs presented in Exhibit 27 are primarily determined by time spent by teachers, coaches, and program specialists. If the COVID-19 pandemic led teachers to spend

less time engaging with PLI activities—and in turn required coaches and specialists to spend less time on PLI activities—then our costs would likely underestimate the costs associated with adopting PLI.

If schools already have a professional development program, then teachers will already be spending time on professional development activities, and our estimate of the cost due to teacher time is an overestimate of the additional cost of adopting PLI. This is especially true if the schools already have a professional development program that utilizes coaches and administrative staff similar to the program specialists used in PLI. In such a case, we anticipate this overestimate to dominate the aforementioned underestimates, resulting in the costs from Exhibit 27 being an overestimate of the true costs. We present estimates of the total cost when a school or district has an existing professional development program in the next subsection.

Total Cost of PLI as a Replacement of an Existing Professional Development

While the previous section presented the estimated cost of adopting PLI for a school or district that does not already have a professional development program, some schools and districts may adopt PLI as a replacement of or an addition to an existing professional development program. In such a case, schools and districts may want to consider the incremental cost of adopting PLI as a replacement for an existing professional development program. To determine this incremental cost, we must measure the additional resources required to implement PLI beyond the resources required to implement the school or district's existing professional development. In practice, this can be measured as the difference between the cost of PLI and the cost of BAU. We view this incremental cost (i.e., the cost of adopting PLI as a replacement of an existing professional development program) as an estimate of the lower bound of the cost of adopting PLI.

Although the structure of an existing professional development program can take many forms, in this analysis, we focus on the professional development offered in BAU schools from the impact evaluation. Because BAU teachers at BAU schools also had the opportunity to participate in PLCs and one-on-one coaching sessions, determining the cost of BAU requires calculating the number of hours that teachers and coaches from BAU schools spent preparing for and participating in one-on-one coaching and PLC meetings (the BAU equivalent to TLT meetings). Our primary source of information on time spent on professional development in BAU schools is the teacher survey from PY2. In that survey, BAU and PLI teachers indicated the number of times they attended one-on-one and group meetings, the average amount of time spent in each type of meeting, and the average number of hours spent preparing for those meetings.

Despite the survey of BAU teachers in PY2, information about time spent by BAU teachers and coaches on professional development is limited. Because participants at BAU schools did not record activities in Whetstone, and BAU teachers were not asked about time usage in PY1, we

do not have information on BAU teachers in PY1 or BAU coaches in either year. Moreover, differences between attendance reported on surveys and recorded in Whetstone among PLI teachers lead us to believe that time reported on surveys is greatly overestimated. Although we believe the degree of overestimation is proportional for BAU and PLI teachers, this overestimation means we do not have a reliable source for the actual number of hours spent by BAU teachers on professional development in either project year.

To approximate the professional development costs for teachers and coaches in BAU schools, we scaled the time spent by PLI teachers and coaches recorded in Whetstone by the ratio of BAU teacher time to PLI teacher time (as reported in the PY2 teacher survey). Specifically, based on responses to the PY2 survey, the average amount of time spent by BAU teachers on professional development activities was about 85.8% of the time spent by PLI teachers on PLI activities (36.3 hours for BAU compared with 42.3 hours for PLI). The average amount of time spent by BAU teachers and coaches in PY2 was then calculated as the average amount of time recorded for PLI teachers and coaches in Whetstone in PY2 times the ratio (0.858). Because BAU teachers were not surveyed about time spent on professional development activities in PY1, we cannot apply the same methodology to determine BAU time in PY1. However, because professional development at BAU schools likely did not change between PY1 and PY2, we assume that the time spent by BAU teachers and coaches in PY1 is equivalent to the (estimated) time spent by BAU teachers and coaches in PY2. Our calculation for BAU teacher time does not account for time spent on workshops as there was no workshop equivalent in the BAU program relative to the workshops in the PLI program.²⁰

Although the BAU schools implemented a professional development program that utilized coaches and teacher time, we make an assumption that they did not use program specialists or PLI materials. We therefore assume the materials and time spent by program specialists to be add-on aspects of the PLI program and do not include estimates of equivalent costs for the BAU group. Thus, for this analysis, BAU costs are only the sum of costs associated with time spent by teachers and coaches in attending and preparing for PLCs and one-on-one coaching.

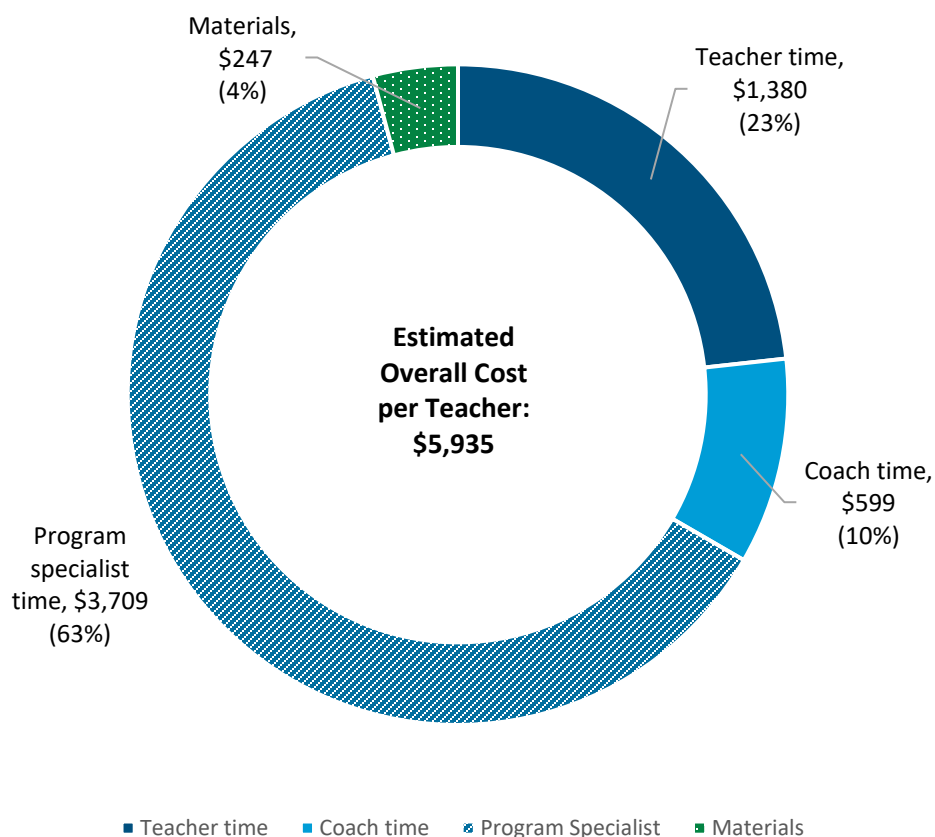
To determine total costs per teacher in the PLI and BAU conditions, we summed the costs of all components in PY1 and PY2 for both conditions and divided each by the number of teachers in each condition in each project year.

²⁰ If workshops were offered by a district, they were offered to all teachers in both the treatment and control schools. PLI workshops offered to teachers in PLI schools were offered in addition to the districtwide workshops offered to all teachers, and therefore can be considered additional to the districtwide workshops.

Findings

Our lower-bound estimate of the overall total cost of PLI per-teacher is \$5,935 over 2 years. Exhibit 28 shows how the per-teacher cost of PLI is distributed across the four program components: coaches' time, teachers' time, program specialists' time, and materials.²¹

Exhibit 28. Lower-Bound Overall Cost Per Teacher of the 2-Year Professional Learning with Impact Program (2023 Dollars)



Note. Costs include the sum of total costs per teacher for both years of implementation. Percentages in parentheses indicate the percentage of the total accounted for by the indicated program component.

According to our analysis, 63% of the costs incurred in running the program come from program specialist time. This is a significant share, compared with the relatively smaller allocations of 23%, 10%, and 4% toward teacher time, coach time, and materials, respectively.

Exhibit 29 breaks down the overall per-teacher cost across the 2 years of the project when PLI was an augmentation of an existing professional development program. As was found when

²¹ The present values of all costs in Year 1 (2021–22) were calculated in 2023 dollars using a 3.0% discount rate.

analyzing the cost of PLI as a stand-alone program, the total costs from PY1 and PY2 indicate that total per-teacher costs in PY2 were less than half of those from PY1. Although cost per teacher dropped in PY2 across all four components, costs due to teacher time exhibited the largest percentage drop (about 91%). As with the cost estimates based on PLI being an augmentation of an existing PLI program, the decrease in cost was primarily driven by a drop in the percentage of teachers that engaged in one-on-one coaching. The decrease in engagement with one-on-one coaching by teachers also is reflected in the cost of coach time in PY2, which fell by 88% in comparison to PY1.

Exhibit 29. Lower-Bound Overall Per-Teacher Cost of Professional Learning with Impact in Year 1 and Year 2 (in 2023 Dollars)

Program Component	Year 1 (2021–22) (N of teachers = 62)	Year 2 (2022–23) (N of teachers = 39)	Total
#1: Teacher Time	\$1,267	\$114	\$1,380
#2: Coach Time	\$535	\$64	\$599
#3: Program Specialist Time	\$2,390	\$1,319	\$3,709
#4: Materials	\$132	\$115	\$247
Total	\$4,324	\$1,612	\$5,935

Generalizability Considerations

The total costs presented above rely on a few assumptions that may not apply to every district or school that chooses to adopt PLI in the future. In this section, we outline the assumptions in our analysis that are unique to this particular implementation of PLI and may not be universally applicable.

The results from the analysis that views PLI as an augmentation to an existing professional development program are largely driven by two assumptions: that the existing professional development program does not already have program specialists (or an equivalent) and that the time spent by BAU teachers and coaches was constant across PY1 and PY2. Though not referred to as program specialists, districts usually have central office staff who provide some level of technical support to local school staff related to instruction. For example, a central office staff person may supervise instructional coaches. To the extent that this is true in a district, the results presented here overestimate the incremental cost of adopting PLI. In such a case, the costs due to program specialist time would be reduced. Our results also rely on the assumption that time spent among BAU teachers and coaches was constant across PY1 and PY2. If that assumption does not hold, our lower-bound estimate of the cost of PLI would either be an overestimate or underestimate depending on whether BAU time was higher or lower

than our estimate. Given the pattern of time spent by PLI teachers and coaches across the project years, we assume it is more likely that assuming constant time across project years for BAU teachers and coaches underestimates BAU time in PY1, though data limitations prevent us from evaluating that assumption. Taken together, we believe that if these two driving assumptions are incorrect, the lower-bound costs presented here would overestimate the lower-bound cost of PLI as an augmentation to an existing professional development program.

Cost-Effectiveness of PLI

Although this study intended to include a cost-effectiveness analysis, we concluded that a cost-effectiveness ratio would not be appropriate. As outlined in the Impact Evaluation section, we found no statistically significant effect of PLI on any of the primary teacher outcomes. We therefore cannot calculate cost-effectiveness for any of the primary teacher outcomes.

The impact evaluation did reveal an effect of PLI on student test scores. Specifically, being enrolled in a school assigned to PLI increases a student's ELA test scores by 0.08 standard deviations on average. The design of the impact evaluation meant that all students from treatment schools were included in the analysis, even if their teachers were excluded from the impact analysis because they joined the school after baseline data collection (in the fall of PY1). For these teachers, we do not have information on participation in PLI activities even though they had the opportunity to participate in all PLI activities and potentially benefited from the services provided within PLI. Although those teachers are likely contributing to the outcomes of students in PLI schools (and in turn the estimated impact of PLI on students), missing data on their participation means we are not able to assess their contribution to the overall cost of PLI. Because the sample of teachers used for the cost study does not align with the sample of students used in the impact evaluation, it would be misleading to calculate the cost-effectiveness of PLI for student outcomes. Therefore, we do not assess cost-effectiveness for the observed impact of PLI on student outcomes.

Discussion

The PLI program was designed to meet a need in the field for scalable, effective teacher professional learning programs, especially those that work in high-need schools. For scalability, the program taps local coaches who facilitate the two core activities for teachers: one-on-one coaching and TLTs. For effectiveness, the program incorporates a multidimensional rubric for describing the quality of classroom practice that has been shown in prior trials to be effective when used as a basis for giving teachers feedback.

This evaluation answers all of the research questions set out earlier, but does have some important limitations. We discuss these first before reviewing the primary findings about implementation and impact. We conclude by exploring the extent to which the evidence gathered sheds light on two important issues: the mechanisms of impacts and the scalability of the program.

Limitations

As discussed in the Evaluation Overview section, the study overlapped with the COVID-19 pandemic, leading to challenges that limited what we were able to learn.

First, during the 2020–21 school year, recruitment was difficult, resulting in a relatively small sample of participating schools and therefore less statistical power than intended. The minimum detectable effect sizes (MDEs) are relatively large, 0.64 for teacher self-efficacy and 0.48 for the overall CLASS-UE score. If the program had impacts (positive or negative) that were smaller in magnitude than the MDEs, it would be unlikely for the study to detect them.

The second limitation stems from the attrition of schools and teachers after randomization. Although some amount of teacher mobility is normal, the combination of typical mobility plus decisions by some of the remaining teachers to stop participating produced high overall attrition. This limitation does not affect the validity of inferences about impacts on student achievement, which are based on all students in the focal grade in all of the randomized schools. However for the findings about impacts on teacher outcomes, the attrition exceeds what is considered acceptable for drawing causal inferences with the highest confidence. The findings about impacts on teacher outcomes have the potential to meet WWC standards “with reservations.”

Findings About the Implementation of the PLI Program

The implementation data make clear that the PLI program was not implemented as designed. However, each of the planned professional learning activities occurred to some extent.

The best-implemented program components were the two delivered directly by the program providers. Specifically, a large majority of teachers attended and were engaged in the preparatory workshops. Teachers also attended the direct support for TLTs, which covered all of the intended topics. Together, these trainings spanned about 15 hours. (The preparatory workshops lasted 6 hours; the direct support for TLTs lasted 9 hours.)

The next best-implemented component was the TLTs. Although teams met about half as frequently as planned, they held many meetings. They met an average of 18 times across the 2 years. (Teachers attended an average of 11 meetings in PY1 and seven in PY2.) The coach did not always follow the collaborative inquiry model with high fidelity, but teams on average completed about three of the inquiry cycles (1.3 cycles in PY1 and 1.8 in PY2).

Last, the individual coaching component was barely implemented at all. About 60% of teachers engaged in at least one debrief meeting or reflection meeting, but it was uncommon to complete a full coaching round, following the PLI protocols for coaching. (The mean number of coaching rounds completed was 0.13 in PY1 and 0.03 in PY2.)

In sum, the implementation data show that teachers completed some of the PLI activities, most notably the workshops and TLT meetings. To the extent that there were statistically significant impacts, we attribute them to these limited activities rather than the full program that was intended.

Findings About Impact

The primary outcome measures examined in the impact analysis were teacher and student outcomes. For the primary teacher outcomes, the challenges with teacher mobility made it difficult to draw inferences with the highest level of confidence, as mentioned above. The impact analyses found no statistically significant impacts on either primary teacher outcome—self-efficacy or classroom practice.

For student achievement, we found PLI to have a positive effect. The impact on ELA achievement was positive and statistically significant (effect size [ES]= 0.09; $p=.031$). The impact on mathematics achievement was positive but not statistically significant (ES=0.05; $p=.319$).

There are many possible reasons why there was a clear impact on ELA achievement but not mathematics achievement. One is that teachers disproportionately chose to set goals related to ELA achievement. At the beginning of each TLT cycle, the teachers set a goal related to student achievement in an area of their own choice. In a supplemental analysis of implementation shown in Appendix B, we found that about half of goals set by TLTs were related to ELA, with only one of 11 schools (9%) explicitly including a goal related to mathematics. The remaining goals aimed to improve student performance broadly. The data did not allow us to directly

attribute the pattern of impact estimates to the differences in goals, but the distribution of goals is at least suggestive of a possible link.

Our final research question was about whether the impact of PLI on student achievement appears to be moderated by student background characteristics. English learner status—which is one of five background characteristics tested—had a statistically significant moderating effect. The fact that many TLTs set goals related to ELA could help explain this result, if, for example, the treatment teachers spent more class time on ELA than mathematics or focused their ELA-related efforts more on students who had the most room to improve in ELA achievement.

Through what mechanism did the program have an impact on student achievement?

The fact that there was an impact on student achievement raises the question of what mechanism caused the impact. The logic model for a teacher professional learning program usually posits that the program will improve teacher outcomes and that those improvements will cause impacts on student achievement. We cannot draw strong conclusions about impacts on teacher outcomes with the highest level of confidence, as discussed under the Limitations section above, but it is worth examining the evidence to develop hypotheses about the mechanisms at work.

Our analyses explored a number of teacher measures that could potentially be mechanisms for changes in student outcomes. As outlined in the Impact Evaluation section, we found no relationship between assignment to PLI and the two primary teacher outcome measures: self-efficacy and classroom practice. Surveys completed by teachers in spring PY1 and spring PY2 also included several supplemental measures that could potentially shed light on other mechanisms. These surveys asked about data use, job satisfaction, feelings of collegial support, perception of instructional performance, perceived success at responding to student needs, frequency of stress, and symptoms of depression (in addition to other measures). In nearly all cases, we found no significant difference between teachers in the treatment and control groups, indicating that these measures may not be salient mechanisms. Two exceptions from PY1 indicate that teachers assigned to PLI experienced more stress but higher levels of job satisfaction (Appendix G). However, given that these are supplemental measures with high attrition and the same findings do not appear in the PY2 survey, we hesitate to conclude that these measures can explain the effects of PLI on students' ELA scores. Overall, we find little evidence of a plausible mechanism for the effect of PLI on student achievement.

There are a few explanations for why this study does not identify a clear mechanism for the impact of PLI on student achievement. It may be the case that the true mechanism is captured

by the teacher measures in this study, but the sample size limits the statistical power to find significant evidence of a mechanism. Because of the high attrition of schools and teachers, most PY2 teacher-level analyses included 70 or fewer teachers. Alternatively, it may be the case that the true mechanism is not captured by the included measures of classroom instruction and teacher sentiments. If so, open-ended discussions with teachers and students could help to uncover potential explanations for how the implementation of PLI at treatment schools led to improvements in student achievement.

Is the program scalable?

The implementation data raise some questions about the scalability of the PLI program. COVID-19 created conditions that made the program hard to implement, but some of the implementation challenges suggest that there may be ways to design the program to be more scalable.

First, does the PLI program design require too much time from teachers and coaches? At the outset of the project, districts made commitments about providing coaches time for their roles and teachers time to participate in the TLTs and coaching. However, the program was only partially implemented, and the qualitative data indicate that teachers and coaches did not always have the time necessary to implement the program.

These data suggest that one way to improve the scalability of the program could be to streamline it. For example, even if grade-level teams have a common meeting time each week, they may find it difficult to focus on PLI, as teams often need meeting time to attend to additional priorities. Also, the impact data show that there was a positive impact on ELA achievement even though TLTs met only about half as often as specified by the program. These data suggest that simply reducing the recommended frequency of the TLT meetings could make the program more scalable and have an impact on student achievement.

Another question raised by the implementation data is whether coaches need more support to implement the program with fidelity. The implementation data suggest that support from school leaders is necessary so that coaches have the time they need, which would enable them to engage in more TLTs and individual coaching. Setting aside concerns about dosage, coaches did not consistently follow the models for TLTs or individual coaching. Coaches may have deviated from the program specifications based on professional judgement. It also is possible that coaches did not receive enough ongoing training. From the implementation data, we know that districts selected qualified coaches, and they attended initial training for their roles. However, coaches did not receive all of the ongoing training that was offered, and that ongoing training may have been necessary to ensure fidelity.

References

- Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Sage Publishing.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long, (Eds.), *Testing structural equation models* (pp. 136–162). Sage Publishing.
- Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20(4), 872–882.
- Carmines, E. G., & McIver, J. D. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohinstdt & E. F. Borgatta (Eds.), *Social measurement: Current issues* (pp. 65–115). Sage Publishing.
- 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. [doi:10.1073/pnas.1315746110](https://doi.org/10.1073/pnas.1315746110)
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679. https://www.nber.org/system/files/working_papers/w19424/w19424.pdf
- Cowan, J., Goldhaber, D., & Theobald, R. (2017). *Teacher equity gaps in Massachusetts*. Massachusetts Department of Elementary and Secondary Education
- Garet, M. S., Wayne, A. J., Brown, S., Rickles, J., Song, M., & Manzeske, D. (2017). *The impact of providing performance feedback to teachers and principals* (NCEE 2018-4001). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/pubs/20184001/pdf/20184001.pdf>
- 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 at the American Institutes for Research. https://gtlcenter.org/sites/default/files/LF-2022-Standards-for-PL-Meta%20Analysis%20Report_Final.pdf

- Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, 11(2), 125–149. <https://eric.ed.gov/?id=EJ1096365>
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. National Comprehensive Center for Teacher Quality. <https://eric.ed.gov/?id=ED521228>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://eric.ed.gov/?id=EJ576570>
- 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 2017-4007). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://eric.ed.gov/?id=EJ1338645>
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (Research Paper. MET Project). Bill & Melinda Gates Foundation. <https://eric.ed.gov/?id=ED540960>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.
- Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, 54(1), 1–36. https://scholar.harvard.edu/files/mkraft/files/kraft_-_2019_-_teacher_effects_jhr.pdf
- Kraft, M. A., & Blazar, D. (2018). Taking teacher coaching to scale: Can personalized training become standard practice? *Education Next*, 18(4), 68–74. <https://eric.ed.gov/?id=EJ1191096>
- 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.
- Muthén, L. K., & Muthén, B. O. (2013). *Mplus user's guide* (7th ed.). Muthén & Muthén.
- National Center for Education Statistics. (2022a). *NAEP long-term trend assessment results: Reading and mathematics*. U.S. Department of Education. <https://www.nationsreportcard.gov/highlights/ltt/2022/>

- National Center for Education Statistics. (2022b). *NAEP report card: 2022 NAEP mathematics assessment*. U.S. Department of Education.
<https://www.nationsreportcard.gov/highlights/mathematics/2022/>
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. <https://www.nber.org/papers/w6691>
- Steinberg, M. P., & 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. S., & Tyler, J. H. (2011). *The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers* (Working Paper 16877). National Bureau of Economic Research. <https://www.nber.org/papers/w16877>
- U.S. Department of Education. (2014). *Fiscal year 2014 budget summary and background information*.
<https://www2.ed.gov/about/overview/budget/budget14/summary/14summary.pdf>
- Wayne, A. J., & Cogshall, J. G. (2022). How to ensure high-quality instructional coaching at scale. *Kappan*, 103(5), 42–46. <https://eric.ed.gov/?id=EJ1327376>
- 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.
<https://ies.ed.gov/ncee/wwc/Handbooks>
- Wheaton, B., Muthén, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology* (pp. 84–136). Jossey-Bass.
- Will, M. (2023, May 15). Teacher job satisfaction rebounds from last year's low. But there's still a ways to go. *Education Week*. <https://www.edweek.org/teaching-learning/teacher-job-satisfaction-rebounds-from-last-years-low-but-theres-still-a-ways-to-go/2023/05>

Appendix A. Additional Analytic Samples for the Implementation Analysis and Implementation Findings for All Four Analytic Samples

As described in the Analytic Samples for the Implementation Analysis section of the report, we defined primary and supplemental analytic samples for the implementation analysis to give readers information about the program activities associated with the reported impacts. The primary implementation samples include only the treatment teachers who were part of the analyses of impact on teacher outcomes, which were the teachers who remained in their original assignments across the 2 years. For Year 1 and Year 2, these samples span three districts, even though four districts participated in the study. This fourth study district experienced teacher mobility such that no study teachers remained in one of the two treatment schools in Year 2. With the loss of all study teachers in one of the treatment schools in Year 2, we did not have the teacher outcome data that would be needed to include this district in the Year 2 teacher impact findings.

To generate implementation findings that include teachers in the fourth district, we defined supplemental implementation samples. This appendix reports implementation findings for “all four districts” to supplement the findings for the “three 2-year districts” presented in the report. The findings based on these supplemental implementation samples are the most relevant to consider when interpreting the primary analyses of impact of student achievement because those analyses account for all students at the target grade levels, regardless of teacher mobility.

Exhibit A1 shows the count of districts, schools, teachers, Professional Learning with Impact (PLI) coaches, and program specialists for each of the four implementation analysis samples.

Exhibit A1. The Four Analytic Samples for Findings About Implementation

Participants	Implementation Sample			
	Year 1 (2021–22) in All Four Districts ²²	Year 1 (2021–22) in Three 2-Year Districts	Year 2 (2022–23) in All Four Districts ²³	Year 2 (2022–23) in Three 2-Year Districts
Districts	4	3	4	3
Schools	14	12	13	12
Teachers	63	56	41	38
Professional Learning with Impact Coaches	14	12	13	12
Program Specialists	5	5	5	5

Implementation Fidelity for the Program Components (RQ1) With All Four Districts

Exhibit A2 shows implementation of each of the four program components—Preparatory Teacher Workshop, Teacher Learning Team (TLT) Cycles, Individual Instructional Coaching, and Direct Support for TLTs—for all four districts in addition to the three 2-year districts in Years 1 and 2.

Exhibit A2. Participation in the Preparatory Teacher Workshop, Teacher Learning Team (TLTs) Cycles, Individual Instructional Coaching, and Direct Support for TLTs

Key Program Indicators	Implementation Sample			
	Year 1 (2021–22) in All Four Districts	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in All Four Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: Preparatory Teacher Workshop				
Percentage of teachers attending preparatory teacher workshop	85.7%	87.5%	N/A	N/A
Percentage of teacher workshops with topics covered as planned	100%	100%	N/A	N/A
Percentage of workshop respondents reporting a high level of intellectual engagement with the workshop content	80.8%	85.4%	N/A	N/A

²² In Year 1, three schools decided to discontinue their participation. These three schools are not included in the implementation analysis and therefore are not included in the count of 14 schools in the Year 1 implementation sample in all four districts.

²³ In Year 2, one school experienced high teacher mobility and no study teachers remained in the school in Year 2. This school is not included in the implementation analysis and therefore is not included in the count of 13 schools in the Year 2 implementation sample in all four districts.

Key Program Indicators	Implementation Sample			
	Year 1 (2021–22) in All Four Districts	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in All Four Districts	Year 2 (2022–23) in the Three 2-Year Districts
Percentage of workshop respondents reporting perceived value from engagement with Professional Learning with Impact (PLI)	92.3%	90.2%	N/A	N/A
Key Component: TLT Cycles				
Mean number of TLT meetings convened	12.15 (0–24)	11.67 (4–24)	10.69 (1–26)	9.42 (1–22)
Mean number of TLT meetings attended	10.68 (0–24)	10.73 (0–24)	8.66 (0–26)	7.29 (0–20)
Mean number of TLT cycles completed	1.14 (0–3)	1.33 (0–3)	1.64 (0–4)	1.83 (0–4)
Mean number of items that PLI coaches achieved on a 17-item fidelity stage protocol checklist	10.0 (0–16)	11.7 (0–16)	12.2 (0–17)	12.0 (0–17)
Percentage of items about team engagement rated <i>agree</i> or <i>strongly agree</i>	83.0%	80.8%	81.3%	80.7%
Key Component: Individual Instructional Coaching				
Mean number of PLI coaching rounds teachers completed	0.11 (0–1)	0.13 (0–1)	0.02 (0–1)	0.03 (0–1)
Mean number of items that PLI coaches achieved on a 15-item fidelity video reflection checklist	5.8 (0–12)	6.8 (0–12)	4.3 (0–14)	4.7 (0–14)
Percentage of items about PLI coach supportiveness rated <i>agree</i> or <i>strongly agree</i>	82.4%	82.7%	72.4%	69.1%
Key Component: Direct Support for TLTs				
Percentage of teachers attending fall and winter team workshops	84.1%	91.1%	N/A	N/A
Percentage of fall and winter workshops with topics covered as planned	100%	100%	N/A	N/A

Note. Some activities only occurred in Year 1 by design and are therefore only measured in Year 1. These are marked N/A in the exhibit.

Exhibit A3 shows scoring of adequate implementation of each of the four program components—Preparatory Teacher Workshop, TLT Cycles, Individual Instructional Coaching, and Direct Support for TLTs—for all four districts in addition to the three 2-year districts in Years 1 and 2.

Exhibit A3. Scoring of Adequate Implementation of Each Key Program Component, by Year

Key Program Components	Implementation Sample			
	Year 1 (2021–22) in All Four Districts	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in All Four Districts	Year 2 (2022–23) in the Three 2-Year Districts
Preparatory Teacher Workshop	Fidelity Not Met 25% of districts (1 of 4) met threshold	Fidelity Not Met 33% of districts (1 of 3) met threshold	N/A	N/A
Teacher Learning Team (TLT) Cycles	Fidelity Not Met 64% of schools (9 of 14) met threshold	Fidelity Not Met 75% of schools (9 of 12) met threshold	Fidelity Not Met 71% of schools (10 of 14) met threshold	Fidelity Not Met 83% of schools (10 of 12) met threshold
Individual Instructional Coaching	Fidelity Not Met 7% of schools (1 of 14) met threshold	Fidelity Not Met 8% of schools (1 of 12) met threshold	Fidelity Not Met 21% of schools (3 of 14) met threshold	Fidelity Not Met 25% of schools (3 of 12) met threshold
Direct Support for TLTs	Fidelity Not Met 25% of districts (1 of 4) met threshold	Fidelity Not Met 33% of districts (1 of 3) met threshold	N/A	N/A

Note. Some activities only occurred in Year 1 by design and are therefore only measured in Year 1. These are marked N/A in the exhibit.

Implementation Fidelity for the Strategy-to-Scale Components (RQ2) With All Four Districts

Exhibit A4 shows implementation of each of the four scaling components—Guidance for District Management of PLI Implementation, PLI Program Protocols and Guidance Documents, PLI Coach Workshops, and Ongoing PLI Coach Support—for all four districts in addition to the three 2-year districts in Years 1 and 2.

Exhibit A4. Districts' and Coaches' Participation in the Strategy-to-Scale Components

Strategy-to-Scale Indicators	Implementation Sample			
	Year 1 (2021–22) in All Four Districts	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in All Four Districts	Year 2 (2022–23) in the Three 2-Year Districts
Key Component: Guidance for District Management of Professional Learning with Impact (PLI) Implementation				
Percentage of districts that received written guidance for successful implementation	100%	100%	N/A	N/A
Percentage of districts that committed to ensuring conditions for successful implementation (i.e., signed services agreement)	100%	100%	N/A	N/A
Percentage of districts that hired/assigned qualified coaches	100%	100%	N/A	N/A
Key Component: PLI Program Protocols and Guidance Documents				
Percentage of districts that received cross-site guidance materials	100%	100%	N/A	N/A
Percentage of districts that received training for coaches on key PLI program materials	100%	100%	N/A	N/A
Key Component: PLI Coach Workshops				
Percentage of coaches who attended coach workshops (summer coach workshop and refresher workshop)	100%	100%	N/A	N/A
Percentage of coach workshops with topics fully covered as intended	87.5%	100%	N/A	N/A
Percentage of coaches reporting comfort with intended content	93.3%	92.3%	N/A	N/A
Key Component: Ongoing PLI Coach Support				
Percentage of coaches reporting that program specialist reviewed coaching conversation and teacher learning team facilitation videos	7.1%	8.3%	0%	0%

Strategy-to-Scale Indicators	Implementation Sample			
	Year 1 (2021–22) in All Four Districts	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in All Four Districts	Year 2 (2022–23) in the Three 2-Year Districts
Percentage of coaches who attended at least eight (Year 1) or four (Year 2) monthly one-on-one coaching call support sessions	51.7%	66.7%	7.1%	8.3%
Percentage of coaches who attended more than six monthly virtual all-coaches professional learning community sessions each year	0%	0%	0%	0%
Percentage of coaches reporting satisfaction with program specialist support	100%	100%	92.3%	91.6%

Note. Some activities only occurred in Year 1 by design and are therefore only measured in Year 1. These are marked N/A in the exhibit.

Exhibit A5 shows scoring of adequate implementation of each of the four scaling components—Guidance for District Management of PLI Implementation, PLI Program Protocols and Guidance Documents, PLI Coach Workshops, and Ongoing PLI Coach Support—for all four districts in addition to the three 2-year districts in Years 1 and 2.

Exhibit A5. Scoring of Adequate Implementation of Each Key Strategy-to-Scale Component, by Year

Key Scaling Components	Implementation Sample			
	Year 1 (2021–22) in All Four Districts	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in All Four Districts	Year 2 (2022–23) in the Three 2-Year Districts
Guidance for District Management of Professional Learning with Impact (PLI) Implementation	Fidelity Met 100% of districts (4 of 4) met threshold	Fidelity Met 100% of districts (3 of 3) met threshold	N/A	N/A
PLI Program Protocols and Guidance Documents	Fidelity Met 100% of districts (4 of 4) met threshold	Fidelity Met 100% of districts (3 of 3) met threshold	N/A	N/A

Key Scaling Components	Implementation Sample			
	Year 1 (2021–22) in All Four Districts	Year 1 (2021–22) in the Three 2-Year Districts	Year 2 (2022–23) in All Four Districts	Year 2 (2022–23) in the Three 2-Year Districts
PLI Coach Workshops	Fidelity Not Met 50% of districts (2 of 4) met threshold	Fidelity Not Met 67% of districts (2 of 3) met threshold	N/A	N/A
Ongoing PLI Coach Support	Fidelity Not Met 71% of schools (10 of 14) met threshold	Fidelity Not Met 83% of schools (10 of 12) met threshold	Fidelity Not Met 29% of schools (4 of 14) met threshold	Fidelity Not Met 33% of schools (4 of 12) met threshold

Note. Some activities only occurred in Year 1 by design and are therefore only measured in Year 1. These are marked N/A in the exhibit.

Appendix B. Additional Implementation Findings

Variation in School Teams' Goals for Students and Teachers

To further explore variation at the school level and provide context for the impact findings, we examined the teacher learning team student and teacher learning goals for each school for Year 2. Exhibit B1 shows about half of the schools created student and teacher learning goals that were not subject specific but aim to improve student performance more broadly. The remaining schools that had documented learning team goals created goals that included aspects of English language arts (ELA), such as a focus on higher-level discourse, language skills, and use of academic vocabulary. Only one school (School 2) also included a goal focused explicitly on math.

Exhibit B1. Variation by School in Year 2 Teacher Learning Team (TLT) Student and Teacher Learning Goals

School	TLT Student and Teacher Learning Goals
	Focus Area of Goal (General, English Language Arts [ELA], Math, and Science)
School 1	"If...students will be able to use an independently created graphic organizer to structure a summary of a teacher provided text..., then students will be able to write strong summaries."
School 2	"Math - If we practice multiplication facts 5 minutes daily, then by May students will know all of their facts 1–10. ELA - If we focus on modeling, practice and check for understanding with a gradual release of responsibility of MAIN IDEA AND DETAILS, students will show growth with their Benchmark Assessments...Science - If we provide more structure for the implementation of Group Roles...and practice more often, then students will be more successful with collaboration and knowing the roles within a group. By doing this, students will learn from each other and improve their language skills..."
School 3	"All students will receive small group instruction focused on their areas of most need."
School 4	"If we increase opportunities for discourse surrounding Depth of Knowledge level 3 questions in whole group instruction, then the data will support students' ability to reason, draw conclusions, explain and justify."
School 5	"The goal...is to facilitate a guided analysis of student assessment data using district assessments through student led conferences."
School 6	"We will explicitly instruct students and encourage students to use academic vocabulary in their 'talk' every day."
School 7	"If the team allows our students to 'teach the teacher' in order to show their understanding, then our students will be able to take accountability of their own learning."
School 8	"If we focus our learning around goal-setting and self-efficacy in the classroom, then we will see an increase in motivation and achievement."
School 9	"Use data to form small groups and use assessment and drill down measures to align small group instruction to the needs of their students."

School	TLT Student and Teacher Learning Goals
	Focus Area of Goal (General, English Language Arts [ELA], Math, and Science)
School 10	"75% of our...students will reach proficiency in ELA at end of grade test."
School 11	"Students will engage in higher level discussion without prompts."
School 12	n/a
School 13	n/a
School 14	n/a

Source. Professional Learning with Impact coaching logs in spring 2023.

Note. n/a means the school was not included in the analysis of the four districts because the school did not upload data to the coaching log platform. The schools in this exhibit are not listed in the same order as the schools presented in the school-level exhibit in the report to protect confidentiality (i.e., School A does not necessarily correspond to School 1).

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

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at School Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Implementation Component 1: Preparatory Teacher Workshop					
Proportion of summer team workshop sessions attended (measured by attendance records)	Teacher	Year 1 0 = Attended less than 3.5 full hours across both sessions or did not watch a makeup webinar recording 1 = Attended between 3.5 and 6 hours across both sessions and/or watched a makeup webinar recording 2 = Attended 6 or more hours across both sessions Adequate implementation = 2	N/A	Year 1 0 = < 95% of teachers with a score of 2 1 = ≥ 95% of teachers with a score of 2	N/A
Summer team workshop topic coverage (measured by the American Institutes for Research [AIR] observer's reports)	Workshop	Year 1 0 = Two or more topics not covered as planned, as reported by the AIR observer 1 = One topic not covered as planned 2 = All 13 topics covered "fully" or "partially" as planned (i.e., no significant deviations made) Adequate implementation = 1	N/A	Year 1 0 = one or more workshops with a score of 0 1 = all workshops with a score of 1 or more	N/A
Proportion of participants who demonstrate intellectual engagement with the content (measured by feedback survey)	Participant	Year 1 0 = Respondent does not "agree" or "strongly agree" with either statement 1 = Respondent "agrees" or "strongly agrees" with one of the two statements 2 = Respondent "agrees" or "strongly agrees" with both statements Adequate implementation = 2	N/A	Year 1 0 = < 95% of respondents with a score of 2 1 = ≥ 95% of respondents with a score of 2	N/A

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at School Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Proportion of participants who perceive value in their engagement with Professional Learning with Impact (PLI) (measured by feedback survey)	Participant	Year 1 0 = Respondent does not “agree” or “strongly agree” with either statement 1 = Respondent “agrees” or “strongly agrees” with one of the two statements 2 = Respondent “agrees” or “strongly agrees” with both statements Adequate implementation = 2	N/A	Year 1 0 = < 95% of respondents with a score of 2 1 = ≥ 95% of respondents with a score of 2	N/A
Key Program Component 1 Total Score: Preparatory Teacher Workshop	N/A	N/A	N/A	Sum of indicator scores (range 0–4): Adequate implementation = 3	Percentage of districts with a score of 3 Adequate sample score = > 85%
Key Implementation Component 2: Teacher Learning Team (TLT) Cycles					
Number of TLT meetings convened (measured by coach logs [online])	Coach	Years 1 and 2 0 = Team met for less than 90 minutes per month on average 1 = Team met between 90 and 135 minutes per month on average 2 = Team met more than 135 minutes for 8 full months each year Adequate implementation = 1	N/A	N/A	N/A
Number of TLT meetings attended (measured by coach logs [online])	Teacher	Years 1 and 2 0 = Attended less than 66% of TLT sessions 1 = Attended 66% to 90% of TLT sessions 2 = Attended more than 90% of TLT sessions Adequate implementation = 1	Years 1 and 2 0 = < 75% of teachers at the school/TLT received a score of 1 or 2 1 = ≥ 75% and < 94% of teachers at the school/TLT received a score of 1 or 2	N/A	N/A

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at School Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
			2 = ≥ 94% of teachers received a score of 2		
Number of TLT cycles completed (measured by coach stage protocols [online])	Coach	Years 1 and 2 0 = Completed less than 1 full cycle per year 1 = Completed ≥ 1 but less than 2 cycles per year 2 = Completed 2 full cycles per year Adequate implementation = 1	N/A	N/A	N/A
Fidelity of TLT work products (i.e., stage protocol text) (measured by fidelity scoring rubric applied to online record [stage protocol] for a randomly selected stage protocol for each coach)	Coach	Years 1 and 2 0 = Checklist score: ≤ 12 1 = Checklist score: 13–15 2 = Checklist score: > 16 Adequate implementation = 1	N/A	N/A	N/A
Teacher perceptions of team engagement (measured by spring teacher survey)	Teacher	Average score across all nine survey items	0 = mean score: < 19 1 = mean score: 19–26 2 = mean score: > 26 Adequate implementation = 1	N/A	N/A
Key Component 2 Total Score: TLT Cycles	N/A	N/A	Sum of five indicator scores (range 0–10): 0 = 0–4 1 = 5–8 2 = 9–10 Adequate implementation = 1	Percentage of coaches/schools with a score of 1 Adequate sample score = > 85%	N/A

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at School Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Implementation Component 3: Individual Instructional Coaching					
Number of PLI coaching rounds completed (measured by coach logs [online])	Teacher	Years 1 and 2 0 = Less than 1 full round completed per year 1 = 1 full round but less than 2 rounds completed per year 2 = 2 full rounds completed per year Adequate implementation = 1	Years 1 and 2 0 = school/TLT average of < 1 round completed 1 = school/TLT average of 1 full round completed and/or second round started but not completed 2 = school/TLT average of 2 rounds completed	N/A	N/A
Fidelity of coach's products (i.e., video reflection text) (measured by fidelity scoring rubric applied to online work product [video reflection text] for a randomly selected sample video reflection text for each coach)	Coach	Years 1 and 2 0 = Checklist score: ≤ 10 1 = Checklist score: 11–13 2 = Checklist score: > 13 Adequate implementation = 1	N/A	N/A	N/A
Teacher perceptions of coach supportiveness (measured by spring teacher survey)	Teacher	Years 1 and 2 Average score across all eight survey items	Years 1 and 2 0 = mean score: < 17 1 = mean score: 17–23 2 = mean score: > 23 Adequate implementation = 1	N/A	N/A

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at School Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Component 3 Total Score: Individual Instructional Coaching	N/A	N/A	Sum of three indicator scores (range 0–6): 0 = 0–2 1 = 3–4 2 = 5–6 Adequate implementation = 1	Percentage of coaches/schools with a score of 1 Adequate sample score = > 85%	N/A
Key Implementation Component 4: Direct Support for TLTs					
Proportion of fall and winter team workshops attended (measured by attendance records)	Teacher	Year 1 0 = Attended less than 4.5 full hours across both sessions or did not watch a makeup webinar recording 1 = Attended between 4.5 and 8 hours across both sessions and/or watched a makeup webinar recording 2 = Attended 8 or more hours across both sessions Adequate implementation = 2	N/A	Year 1 0 = < 95% of teachers with a score of 2 1 = ≥ 95% of teachers with a score of 2	N/A
Fall and winter team workshop topic coverage (measured by AIR observer's reports)	Workshop	Year 1 0 = Two or more topics not covered as planned, as reported by the AIR observer 1 = One topic not covered as planned 2 = All topics covered “fully” or “partially” as planned (i.e., no significant deviations made) Adequate implementation = 2	N/A	Year 1 0 = one or more workshops with a score of 0 1 = all workshops with a score of 1 or more	N/A
Key Component 4 Total Score: Direct Support for TLTs	N/A	N/A	N/A	Sum of two indicator scores (range 0–4): 0 = 0–1 1 = 2–4 Adequate implementation = 1	Percentage of districts with a score of 1 Adequate sample score = > 85%

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at School Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Strategy-to-Scale Component 1: Guidance for District Management of PLI Implementation					
Program provides written guidance for successful implementation to district partners (i.e., district booklet) (measured by recruitment/implementation records)	District	Year 1 0 = District officials did not receive booklet 1 = District officials received booklet Adequate implementation = 1	N/A	N/A	N/A
District commits to ensuring conditions for successful implementation (i.e., a signed services agreement) (measured by a signed services agreement)	District	Year 1 0 = District officials did not sign a services agreement with commitment language 1 = District officials signed a services agreement with commitment language Adequate implementation = 1	N/A	N/A	N/A
District hires/assigns qualified coaches (measured by a program specialist survey)	District	Year 1 0 = No coach in the district fully met requirements 1 = At least one coach in the district did not fully meet requirements 2 = All coaches in the district fully met requirements Adequate implementation = 1	N/A	N/A	N/A
Key Component 1 Total Score: Guidance for District Management of PLI Implementation	N/A	N/A	N/A	Year 1 Sum of indicator scores (range 0–4): Adequate implementation = 3	Percentage of districts with a score of 3 Adequate sample score = 100%

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at School Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Key Strategy-to-Scale Component 2: PLI Program Protocols and Guidance Documents					
Program provides cross-site guidance materials (measured by implementation records)	Workshop	Year 1 0 = Program specialists did not provide refined materials (PLI guidebook, Framework for Teaching [FFT], and Whetstone access) to all coaches 1 = Program specialists provided refined materials (PLI guidebook, FFT, and Whetstone access) to all coaches	N/A	N/A	N/A
Program specialists provide training to coaches on key PLI program materials (measured by AIR observer's reports)	Workshop	Year 1 0 = Program specialists did not cover use of all PLI guidance materials during workshops 1 = Program specialists covered basic use of all PLI guidance materials during workshops 2 = Program specialists covered use of PLI guidance materials during workshops fully and completely Adequate implementation = 1	N/A	N/A	N/A
Key Component 2 Total Score: PLI Program Protocols and Guidance Documents	N/A	N/A	N/A	Year 1 Sum of indicator scores (range 0–3): Adequate implementation = 2	Percentage of districts with a score of 2 Adequate sample score = 100%
Key Strategy-to-Scale Component 3: PLI Coach Workshops					
Proportion of coach workshops attended (summer coach workshop [Year 1] and refresher workshop [Year 2]) (measured by attendance records)	Coach	Year 1 0 = Attended fewer than 7 full hours across both sessions and/or did not watch a makeup webinar recording 1 = Attended 7–9 hours and/or watched a makeup webinar recording 2 = Attended more than 9 hours	Year 1 Adequate implementation = 92% or more of coaches received an attendance score of 1 or 2	N/A	N/A

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at School Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
Coach workshop topics covered as intended (measured by AIR observer's report)	Workshop	Year 1 0 = Two or more topics not covered as planned, as reported by the AIR observer 1 = One topic not covered as planned 2 = All topics covered "fully" or "partially" as planned (i.e., no significant deviations made) Adequate implementation = 1	N/A	Year 1 0 = one or more workshops with a score of 0 1 = all workshops with a score of 1 or more	N/A
Coach reported comfort with intended content (measured by workshop feedback surveys)	Coach	Year 1 0 = Respondent did not "agree" or "strongly agree" with either statement 1 = Respondent "agrees" or "strongly agrees" with one of the two statements 2 = Respondent "agrees" or "strongly agrees" with both statements Adequate implementation = 2	Year 1 0 = < 90% of respondents with a score of 2 1 = ≥ 90% of respondents with a score of 2	N/A	N/A
Key Component 3 Total Score: PLI Coach Workshops	N/A	N/A	N/A	Year 1 Sum of indicator scores (range 0–6): Adequate implementation = 3	Percentage of districts with a score of 3 Adequate sample score = 100%
Key Strategy-to-Scale Component 4: Ongoing PLI Coach Support					
Program specialist review of coaching conversation and TLT facilitation videos (measured by coach interview)	Coach	Years 1 and 2 0 = Reported a mean score of ≤ 3 1 = Reported a mean score of 4–7 2 = Reported a mean score of > 7 Adequate implementation = 1	Years 1 and 2 Adequate implementation = 90% of coaches reported a score of 1 or 2	N/A	N/A
Number of monthly one-on-one coaching call support sessions attended (by coach) (measured by program specialist logs)	Coach	Year 1 0 = Attended < 5 sessions 1 = Attended 5–7 sessions 2 = Attended > 7 sessions Adequate implementation = 2	Year 1 Adequate implementation = 90% of coaches attended > 7 sessions	N/A	N/A

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at School Level	Indicator Scoring at District Level	Indicator Scoring at Sample Level
		Year 2 0 = Attended < 2 sessions 1 = Attended 2–3 sessions 2 = Attended > 3 sessions Adequate implementation = 2	Year 2 Adequate implementation = 90% of coaches attended > 3 sessions		
Number of monthly virtual all-coaches professional learning communities attended (by coach) (measured by program specialist logs)	Coach	Year 1 0 = Attended < 5 sessions 1 = Attended 5–6 sessions 2 = Attended > 6 sessions Adequate implementation = 2 Year 2 0 = Attended < 4 sessions 1 = Attended 4–6 sessions 2 = Attended > 6 sessions Adequate implementation = 2	Years 1 and 2 Adequate implementation = 90% of coaches attended > 6 sessions	N/A	N/A
Coach satisfaction with program specialist support (measured by coach interview)	Coach	Years 1 and 2 0 = Reported a mean score of ≤ 3 1 = Reported a mean score of 4–7 2 = Reported a mean score of > 7 Adequate implementation = 1	Years 1 and 2 Adequate implementation = 90% of coaches reported a score of 1 or 2	N/A	N/A
Key Component 4 Total Score: Ongoing PLI Coach Support	N/A	N/A	Sum of 4 indicator scores (range 0–8): 0 = 0–2 1 = 3–5 2 = 6–8 Adequate implementation = 1	Percentage of coaches/schools with a score of 1 Adequate sample score = > 85%	N/A

Appendix D. Descriptive Information About the Sample of Schools Included in the Randomized Controlled Trial

Exhibit D1. Characteristics of Professional Learning with Impact (PLI) and Business-as-Usual (BAU) Schools

District	Randomization block	School	Grades	Target Grade	Departmental-ization	Enrollment	% Black	% Hispanic	% White	% FRPL	ELA Proficiency Rate	Math Proficiency Rate	Average Proficiency Rate
PLI Schools													
A	1	School 1	PK–5	4	Yes	787	37%	18%	32%	76%	21%	26%	23%
A	2	School 2	PK–5	4	Yes	993	29%	18%	37%	55%	28%	48%	38%
A	2	School 3	PK–5	4	Yes	818	22%	16%	49%	47%	33%	42%	38%
A	2	School 4	PK–5	4	Yes	800	28%	18%	42%	47%	42%	48%	45%
A	3	School 5	PK–5	4	Yes	932	29%	17%	43%	51%	41%	64%	52%
A	3	School 6	PK–5	4	Yes	506	0%	4%	93%	37%	59%	66%	63%
A	4	School 7 ^a	PK–5	4	Yes	745	35%	15%	36%	75%	56%	53%	55%
A	4	School 8 ^a	PK–5	4	Yes	625	36%	14%	41%	86%	31%	51%	41%
A	4	School 9	PK–5	4	Yes	550	37%	10%	42%	80%	37%	56%	47%
B	9	School 19 ^b	PK–6	5	No	554	34%	63%	1%	98%	3%	0%	2%
B	9	School 20 ^a	PK–8	5	No	663	36%	55%	6%	96%	2%	7%	5%
B	10	School 21 ^b	PK–6	4	No	313	43%	14%	36%	60%	30%	37%	34%
C	11	School 24	PK–5	4	Yes	491	13%	48%	33%	99%	60%	70%	65%
C	11	School 25	PK–5	4	Yes	391	38%	21%	34%	99%	62%	61%	61%
C	11	School 26	3–5	4	Yes	432	22%	56%	19%	99%	56%	57%	57%
D	12	School 29	PK–6	4	Yes	572	1%	62%	33%	66%	31%	29%	30%

District	Randomization block	School	Grades	Target Grade	Departmental-ization	Enrollment	% Black	% Hispanic	% White	% FRPL	ELA Proficiency Rate	Math Proficiency Rate	Average Proficiency Rate
D	13	School 30	PK-5	4	No	738	2%	56%	40%	56%	34%	27%	31%
Average Among PLI Schools						642	26%	30%	36%	72%	37%	44%	40%
BAU Schools													
A	1	School 10	3-5	4	Yes	345	37%	12%	36%	70%	14%	15%	15%
A	2	School 11	PK-5	4	Yes	607	24%	13%	54%	61%	35%	38%	37%
A	2	School 12	PK-5	4	Yes	501	31%	11%	46%	60%	48%	37%	42%
A	2	School 13	PK-5	4	Yes	1100	25%	16%	49%	53%	38%	53%	46%
A	3	School 14	PK-5	4	Yes	952	26%	15%	45%	46%	45%	55%	50%
A	3	School 15	PK-5	4	Yes	631	5%	5%	84%	49%	47%	58%	52%
A	3	School 16	PK-5	4	Yes	757	8%	10%	74%	46%	40%	66%	53%
A	4	School 17	PK-5	4	Yes	861	28%	15%	44%	71%	46%	57%	51%
A	4	School 18	PK-5	4	Yes	849	37%	21%	27%	75%	40%	56%	48%
B	10	School 22 ^b	K-8	4	No	652	32%	62%	5%	93%	15%	7%	11%
B	9	School 23 ^b	PK-6	5	No	330	51%	36%	12%	89%	14%	18%	16%
C	11	School 27	PK-5	4	Yes	503	13%	40%	41%	99%	64%	70%	67%
C	11	School 28	PK-5	4	Yes	476	18%	43%	32%	99%	52%	61%	57%
D	12	School 31	PK-5	4	Yes	805	1%	66%	30%	55%	42%	38%	40%

District	Randomization block	School	Grades	Target Grade	Departmental-ization	Enrollment	% Black	% Hispanic	% White	% FRPL	ELA Proficiency Rate	Math Proficiency Rate	Average Proficiency Rate
D	13	School 32	PK-5	4	No	590	1%	49%	46%	59%	34%	31%	33%
Average Among BAU Schools						664	23%	28%	41%	68%	38%	44%	41%

^a School attrited from the study in Year 2.

^b School attrited from the study in Year 1.

Note. FRPL = free or reduced-price lunch; ELA = English language arts.

Appendix E. Assessment of the Measurement Properties of Teacher Survey Scales

The use of composite scores and comparisons of constructs necessitates that the measurement of teacher constructs is validated and determined invariant across groups (in this case, treatment and control conditions). The reliability and validity of the teacher outcome measures were evaluated by subjecting them to a multi-group confirmatory factor analysis (MGCFA) to determine the fit and appropriateness of the measure for the populations using recommended guidelines by Byrne (2008), Kline (2011), and Chen (2007). MGCFA stems from a class of structural equation modeling (SEM) techniques that facilitate the assessment of construct validity as well as configural and measurement invariance (i.e., how well measurement models generalize across groups of individuals; Brown, 2006).

We assessed the fit of the 11 teacher survey measures (teacher self-efficacy, student-centered beliefs about teaching, situated teacher self-efficacy for instructional improvement, collegial support, insights about how to teach, confidence in effectiveness of normal teaching practices, efforts to teach better, perceived success at responding to student needs, perception of instructional performance, data use, and job satisfaction) using Mplus 8.0 (Muthén & Muthén, 2017) separately for teachers in treatment and control schools. Full information maximum likelihood estimation was used to address missing data by estimating means and intercepts. We evaluated models using four recommended fit indices (Browne & Cudeck, 1993; Hu & Bentler, 1999) as follows:

- **Root mean square error of approximation (RMSEA)**, where values $< .08$ indicate acceptable fit and values $< .05$ indicate close fit
- **Confirmatory fit index (CFI) and Tucker-Lewis Index (TLI)**, where estimates $> .90$ indicate acceptable fit and values $> .95$ indicate close fit
- **Standardized root mean square residual (SRMR)**, where values $< .08$ are considered acceptable

We evaluated measurement invariance using MGCFA to test for configural invariance. Configural invariance was evaluated in Model 1 by testing for cross-group invariance. All parameters were estimated freely in Model 1 to determine whether the same number of factors and their respective loading patterns would hold across groups. Establishing configural invariance ensured that each of the survey items was associated with the same underlying factor across treatment and control conditions (Chen, 2007). Model 1 was used as the baseline model against which all subsequent models were compared (Byrne, 2008). The specification of

additional cross-group equality constraints facilitated the evaluation of measurement equivalence in subsequent estimations of factorial and intercept invariance (Byrne, 2008). Model 2 tested for factorial invariance (i.e., whether or not the set of indicators assess the same construct across groups) by constraining the unstandardized factor loadings to be equal across groups (Kline, 2011). Factor loadings and intercepts were constrained to be equal in Model 3 (scalar invariance).

The evaluation of model invariance and parsimony was guided by the recommendations of Chen (2007). Goodness-of-fit indices are reported, and nested chi-square tests facilitated the determination of improvement in model fit. The chi-square statistic (χ^2) is restrictive in two ways: (a) it assesses exact fit (rather than reasonable good fit), and (b) it is likely to be significant for sample sizes larger than 75 to 200 cases (Bollen & Long, 1993). A chi-square to degrees-of-freedom (χ^2/df) ratio of less than 3 is an indicator of reasonable fit and an alternative fit index to consider in addition to χ^2 (Carmines & McIver, 1981; Kline, 2011; Wheaton et al., 1977). Nested tests of chi-square values and other fit indices were used to determine whether model fit was significantly reduced with the introduction of constraints. We also reported relative change in CFI, TLI, RMSEA, and SRMR as indicators of measurement of noninvariance (i.e., improvement in fit with each added constraint across the three models evaluated) as they are less sensitive to sample size. Models were considered invariant and parsimonious when changes in CFI and TLI did not exceed 0.01, changes in RMSEA did not exceed .015, and changes in SRMR did not exceed .03 (Chen, 2007).

In Exhibit E1, we list the survey items associated with teacher outcome measures, standardized factor loadings, and values of Cronbach's alpha, a measure of internal consistency. The teacher survey outcome measures meet What Works Clearinghouse standards for internal consistency, with values of Cronbach's alpha ranging from .63 to .93. Exhibit E2 presents the MGCFA results along with model comparisons. Results of these analyses demonstrate that configural invariance was established within this sample of teachers; therefore, we can assume that the same constructs are measured among teachers in treatment and control schools.

Exhibit E1. Survey Items Associated With Teacher Survey Outcome Measures, Their Corresponding Factor Loadings, and Values of Cronbach's Alpha

Teacher Outcome Measure	Question Number	Question Stem	Item	Factor Loadings (β)		Cronbach's Alpha (α)
				BAU	PLI	
Teacher self-efficacy	Q2	<i>How much can you do to...</i>	...get through to your most difficult students?	0.81	0.57	0.92
			...keep students on task on difficult assignments?	0.77	0.79	
			...motivate students who show low interest in schoolwork?	0.78	0.73	
			...control disruptive behavior in the classroom?	0.73	0.81	
Student-centered beliefs about teaching	Q3	<i>To what extent do you agree or disagree with the following statements?</i>	Instruction should be built around problems or questions with clear, correct answers.	0.64	0.77	0.65
			Effective/good teachers demonstrate the correct way to solve a problem.	0.60	0.53	
			My primary role as a teacher is to help students learn the course content.	0.63	0.56	
Situating teacher self-efficacy for instructional improvement	Q4	<i>To what extent do you feel supported and able to ...</i>	...identify and implement new strategies that better meet your students' learning needs?	0.82	0.86	0.91
			...find and implement practical solutions to your day-to-day teaching challenges?	0.86	0.90	
			...reflect on your teaching practice to make effective adjustments?	0.82	0.90	
			...learn new strategies to increase student engagement?	0.78	0.88	

Teacher Outcome Measure	Question Number	Question Stem	Item	Factor Loadings (β)		Cronbach's Alpha (α)
				BAU	PLI	
Collegial support	Q5	<i>Considering only your colleagues at your school who teach at your same grade level, indicate the extent to which you agree or disagree with the following statements.</i>	My colleagues trust my judgment.	0.82	0.35	0.93
			I have colleagues to whom I can turn to for advice or support.	0.78	0.74	
			My colleagues respect teachers who are expert at their craft.	0.84	0.84	
			My colleagues support each other in their efforts to improve teaching.	0.85	0.91	
			My colleagues genuinely care about each other.	0.95	0.87	
			My colleagues are competent professionals.	0.94	0.96	
			My colleagues are working hard to achieve what's best for all students we teach.	0.96	0.87	
			When working with my colleagues, it's okay to ask for help or admit you don't know something.	0.82	0.81	
Insights about how to teach	Q6	<i>In the 2022–23 school year ...</i>	...I became aware of important aspects of my teaching that I had not noticed before.	0.67	0.92	0.87
			...I stopped using some practices that I realized are less effective.	0.56	0.85	
			...I had many “a-ha” moments about my teaching.	0.82	0.91	
			...I realized I was undervaluing important aspects of teaching.	0.73	0.60	
			...New instructional practices that I tried shifted my beliefs about teaching and learning.	0.67	0.88	

Teacher Outcome Measure	Question Number	Question Stem	Item	Factor Loadings (β)		Cronbach's Alpha (α)
				BAU	PLI	
Confidence in effectiveness of normal teaching practices	Q7	<i>In the 2022–23 school year ...</i>	...I formed a greater appreciation for some teaching practices I commonly use.	0.52	0.51	0.80
			...My classroom experiences reconfirmed the importance of some of my ideas about effective teaching.	0.61	0.82	
			...I had some experiences that reminded me to use some of my existing practices more often.	0.92	0.90	
			...I learned that some of my normal, day-to-day teaching practices are more important than I had previously thought.	0.72	0.84	
Efforts to teach better	Q8	<i>Compared with a typical year, in the 2022–23 school year...</i>	...I tried harder to improve the quality of my teaching.	0.19 (ns)	0.88	0.84
			...I experimented more with new teaching practices.	0.62	0.82	
			...I spent more time trying to make significant improvements in how I teach.	0.87	0.78	
			...I was more intentional about incorporating best practices into my teaching.	0.40	0.99	
Perceived success at responding to student needs	Q9	<i>In the 2022–23 school year, more than in a typical year...</i>	...My instruction has met my students' needs.	0.81	0.79	0.78
			...I have found effective ways to engage my students.	0.67	0.60	
			...I have known what to do when I saw the need for an improvement in my classroom.	0.54	0.81	
			...I have been the teacher I wanted to be for my students.	0.71	0.57	

Teacher Outcome Measure	Question Number	Question Stem	Item	Factor Loadings (β)		Cronbach's Alpha (α)
				BAU	PLI	
Perception of instructional performance	Q10	<i>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.</i>	Classroom organization (e.g., routines for distributing materials, behavior management, instructional formats such as group work)	0.65	0.55	0.88
			Content-specific teaching techniques (e.g., using manipulatives to teach ratios or graphic organizers to teach writing)	0.65	0.73	
			Emotional support for students (e.g., demonstrating warmth or sensitivity to student concerns)	0.59	0.73	
			Instructional support (e.g., how to use questioning and discussion techniques, how to assess and support student understanding)	0.81	0.81	
			Lesson planning (e.g., aligning lesson plans to curriculum standards)	0.77	0.69	
			Student engagement (e.g., ensuring classroom participation, promoting active academic engagement with content)	0.78	0.97	
			Inclusive instruction (e.g., ensuring that all students feel like they belong, connecting content to students' personal lives, using cultural referents appropriately)	0.72	0.70	
Data use	Q11	<i>During the 2022–23 school year, how often did you use student learning data in your classroom teaching practice?</i>	I use student learning data to set student learning goals.	0.43	0.72	0.63
			I use student learning data to discuss student needs with my colleagues.	0.40	0.81	
			I track my students' levels of participation or engagement to understand how they are responding to changes I make to my instruction.	0.67	0.64	

Teacher Outcome Measure	Question Number	Question Stem	Item	Factor Loadings (β)		Cronbach's Alpha (α)
				BAU	PLI	
Job satisfaction	Q15	<i>How often do you feel the following way about being a classroom teacher?</i>	I am content with my profession as a teacher.	0.85	0.82	0.92
			I am satisfied with being a teacher at this school.	0.77	0.77	
			I find my work full of meaning and purpose.	0.79	0.70	
			I am enthusiastic about my job.	0.91	0.83	
			My work inspires me.	0.86	0.85	
			I am proud of the work I do.	0.80	0.67	
			I feel hopeful I can continue to be a classroom teacher until I retire.	0.79	0.84	

Note. Standardized factor loadings are reported. All reported factor loadings are significant at $p \leq .05$ or higher. Response options for Q2 range from 1 (nothing) to 9 (a great deal). Response options for Q3, Q5, Q6, Q7, Q8, and Q9 range from 1 (strongly disagree) to 4 (strongly agree). Response options for Q4 range from 1 (not at all) to 4 (to a great extent). Response options for Q10 range from 1 (I did not improve) to 4 (I improved a large amount). Response options for Q11 range from 1 (never) to 5 (about weekly or more often). Response options for Q15 range from 1 (never or almost never) to 4 (very often). BAU = business as usual; PLI = Professional Learning with Impact; ns = not statistically significant.

Exhibit E2. Multigroup Confirmatory Factor Analysis Nested Model Comparisons for All Teacher Survey Outcome Measures

	Q2		Q3		Q4		Q5		Q6		Q7		Q8		Q9		Q10		Q11		Q15	
	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
χ^2 (df)	18.03 (8)	18.74 (10)	3.53 (3)	3.54 (4)	10.69 (7)	11.31 (10)	103.33 (48)	111.78 (54)	22.04 (15)	37.26 (18)	16.44 (8)	16.97 (10)	15.58 (7)	22.36 (10)	8.28 (7)	10.49 (10)	38.05 (36)	47.32 (41)	2.59 (2)	2.93 (3)	46.05 (35)	49.89 (41)
χ^2/df	2.25	1.87	1.18	0.89	1.53	1.13	2.15	2.07	1.47	2.07	2.06	1.70	2.23	2.24	1.18	1.05	1.06	1.15	0.86	0.98	1.32	1.22
CFI	0.94	0.95	0.98	1.00	0.99	0.99	0.93	0.93	0.97	0.93	0.95	0.96	0.96	0.94	0.98	0.99	0.99	0.98	0.99	1.00	0.98	0.98
TLI	0.91	0.93	0.97	1.00	0.98	0.99	0.92	0.93	0.96	0.92	0.93	0.95	0.93	0.92	0.98	0.99	0.99	0.98	0.96	1.00	0.97	0.99
RMSEA	0.15	0.13	0.05	0.00	0.09	0.05	0.15	0.14	0.09	0.14	0.14	0.12	0.15	0.16	0.06	0.03	0.03	0.05	0.07	0.00	0.08	0.07
SRMR	0.09	0.13	0.09	0.09	0.03	0.06	0.11	0.15	0.06	0.20	0.07	0.07	0.07	0.15	0.08	0.12	0.12	0.13	0.07	0.09	0.05	0.12
Model Comparisons																						
Δ df		2		1		3		6		3		0.53		10		3		10		2		6
$\Delta \chi^2$		0.71		0.71		0.62		8.45		15.22		2		6.78		2.21		9.46		1.32		3.84
Δ CFI		0.01		0.02		0.00		0.00		-0.04		0.01		-0.02		0.01		-0.01		0.01		0.00
Δ TLI		0.02		0.03		0.01		0.01		-0.04		0.02		-0.01		0.01		-0.01		0.04		0.02
Δ RMSEA		0.00		-0.05		-0.04		-0.01		0.05		-0.02		0.01		-0.03		0.02		-0.07		-0.01
Δ SRMR		0.03		0.00		0.03		0.04		0.14		0.00		0.08		0.04		0.01		0.02		0.07
		vs. M1		vs. M1		vs. M1		vs. M1		vs. M1		vs. M1		vs. M1		vs. M1		vs. M1		vs. M1		vs. M1

Note. χ^2 = chi-square statistic, Δ = delta or change, df = degrees of freedom, CFI = comparative fit index; TLI = Tucker Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root mean square residual; M1 = Model 1.

Model 1: All parameters are estimated freely across groups (unconstrained).

Model 2: Factor loadings are constrained to be equal across groups.

Exhibit E3. Correlation Matrix of Scale Means

	Outcome	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14a	Q14b	Q15
Q2	Teacher efficacy	1														
Q3	Student-centered beliefs about teaching	0.15	1													
Q4	Situated teacher self-efficacy for instructional improvement	0.37***	0.22*	1												
Q5	Collegial support	0.39***	0.13	0.29***	1											
Q6	Insights about how to teach	0.17	0.10	-0.01	0.26**	1										
Q7	Confidence in effectiveness of normal teaching practices	0.24**	0.04	0.24**	0.21*	0.42***	1									
Q8	Efforts to teach better	0.31***	0.07	0.02	0.39***	0.58***	0.25**	1								
Q9	Perceived success at responding to student needs	0.32***	0.02	0.30***	0.18	0.00	0.15	0.20*	1							
Q10	Efforts to teach better	0.23*	0.05	0.17	0.16	0.38***	0.41***	0.33***	0.44***	1						
Q11	Data use	0.26**	0.09	0.29***	0.16	0.08	0.09	-0.01	0.22*	0.25**	1					
Q12	Stress frequency	-0.05	-0.07	-0.23*	-0.13	0.12	0.16	0.17	-0.08	0.09	0.08	1				
Q13	Coping with stress	0.11	0.07	0.35***	0.21*	-0.02	0.03	0.01	0.37***	0.18	0.04	-0.60***	1			
Q14a	Little interest	-0.02	-0.11	-0.27**	-0.11	0.00	-0.01	0.02	-0.23*	-0.17	-0.07	0.43***	-0.61***	1		
Q14b	Feelings of depression	0.03	-0.15	-0.27**	-0.15	0.04	0.01	0.03	-0.16	-0.09	0.00	0.40***	-0.52***	0.81***	1	
Q15	Job satisfaction	0.31***	-0.02	0.42***	0.46***	0.12	0.15	0.12	0.36***	0.26**	0.17	-0.41***	0.53***	-0.49***	-0.37***	1

Note. Pearson correlation coefficients. * $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix F. Attrition and Baseline Equivalence for Outcome Measures in Program Year 2 and Program Year 1

Exhibit F1. Attrition for Teacher Survey Outcome Measures in Program Year 2 and Program Year 1

Teacher Outcomes	Treatment		Control		Overall Attrition	Differential Attrition
	# Randomized	# Analytic Sample	# Randomized	# Analytic Sample		
Year 2 Survey Measures: School-Level Attrition						
Schools	17	12	15	13	21.9%	16.1%
Year 2 Survey Measures: Teacher-Level Attrition						
Student-centered beliefs about teaching	64	33	60	30	49.2%	1.6%
Situated teacher self-efficacy for instructional improvement	64	33	60	30	49.2%	1.6%
Collegial support	64	29	60	30	52.4%	4.7%
Insights about how to teach	64	37	60	34	42.7%	1.1%
Confidence in effectiveness of some of my normal teaching practices	64	37	60	34	42.7%	1.1%
Efforts to teach better	64	37	60	34	42.7%	1.1%
Perceived success at responding to student needs	64	37	60	34	42.7%	1.1%
Perception of instructional performance	64	33	60	30	49.2%	1.6%
Data use	64	26	60	28	56.5%	6.0%
Stress frequency	64	36	60	36	41.9%	3.8%
Coping with stress	64	36	60	34	43.5%	0.4%
Symptoms of depression	64	36	60	34	43.5%	0.4%
Job satisfaction	64	31	60	28	52.4%	1.8%
Year 1 Survey Measures: School-Level Attrition						
Schools	17	14	15	15	9.4%	17.6%
Year 1 Survey Measures: Teacher-Level Attrition						
Self-efficacy	71	56	68	46	26.6%	11.2%
Perception of instructional performance	71	56	68	46	26.6%	11.2%

Teacher Outcomes	Treatment		Control		Overall Attrition	Differential Attrition
	# Randomized	# Analytic Sample	# Randomized	# Analytic Sample		
Situated teacher self-efficacy for instructional improvement	71	56	68	46	26.6%	11.2%
Student-centered beliefs about teaching	71	56	68	46	26.6%	11.2%
Data use	71	56	68	46	26.6%	11.2%
Collegial support	71	56	68	46	26.6%	11.2%
Stress frequency	71	56	68	46	26.6%	11.2%
Coping with stress	71	56	68	46	26.6%	11.2%
Little interest	71	56	68	46	26.6%	11.2%
Feelings of depression	71	56	68	46	26.6%	11.2%
Job satisfaction	71	56	68	46	26.6%	11.2%

Baseline Equivalence of Analytical Samples in Impact Analysis

Exhibit F2. Baseline Observation Scores and Survey Measures of Teachers Included in the Analyses of Year 2 Impact on Teacher Outcomes, by Study Group

Baseline Measure	Treatment Group		Control Group		Treatment – Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
CLASS-UE Overall Score	3.89	0.55	3.99	0.59	-0.10	0.13	-0.17	0.476
CLASS-UE Domain Scores:								
Emotional Support	4.05	0.87	4.20	0.80	-0.15	0.18	-0.18	0.407
Classroom Organization	4.66	0.37	4.56	0.55	0.10	0.12	0.20	0.412
Instructional Support	3.34	0.75	3.51	0.80	-0.18	0.19	-0.23	0.348
Student Engagement	4.81	0.87	4.99	0.96	-0.19	0.25	-0.20	0.444
Teacher Survey Outcome								
Self-efficacy	3.14	0.49	3.07	0.55	0.07	0.13	0.13	0.600

Note. Sample size for “Master’s degree or higher” and “Years of teaching experience”= 66 teachers. Sample size for CLASS-UE scores = 66 teachers (34 treatment and 32 control) and 127 observations (63 treatment and 64 control). Sample size for Student Engagement domain = 58 teachers (31 treatment and 27 control) and 117 observations (60 treatment and 57 control). Sample size for Teacher Self-Efficacy= 30 teachers for the treatment group and 30 teachers for the control group. The baseline equivalence analyses were based on a two-level 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. The Professional Learning with Impact (PLI) group means are unadjusted means; the business-as-usual (BAU) group means are adjusted means computed based on the unadjusted PLI group means and the estimated group mean difference. Standard deviations presented are the pooled standard deviations of the outcome for the PLI and BAU groups. Effect sizes were computed using the Cox index for dichotomous measures and Hedges’ g for continuous measures, as recommended by the What Works Clearinghouse (2022). CLASS UE = Classroom Assessment Scoring System-Upper Elementary; SD = standard deviation.

Exhibit F3. Baseline Equivalence of Survey Outcomes for Teachers Included in the Analyses of Year 2 Impact on Supplemental Survey Outcomes, by Study Group

Baseline Measure of Supplemental Outcome	Treatment Group		Control Group		Treatment – Control Difference	Standard Error	Effect Size	P-Value	Obs.
	Mean	SD	Mean	SD					
Student-centered beliefs about teaching	2.98	0.25	2.99	0.33	0.00	0.07	-0.02	0.945	63
Situated teacher self-efficacy for instructional improvement	3.09	0.66	2.99	0.62	0.10	0.15	0.15	0.532	63
Collegial support	3.56	0.38	3.49	0.39	0.07	0.10	0.19	0.470	59
Perception of instructional performance	3.06	0.55	2.99	0.39	0.08	0.11	0.16	0.492	63
Data use	4.46	0.48	4.37	0.46	0.09	0.13	0.19	0.484	54
Job satisfaction	2.83	0.60	2.97	0.63	-0.14	0.16	-0.22	0.397	59

Note. See Exhibit F1 for sample sizes by treatment condition. All *p*-values on baseline equivalence tests come from the estimated coefficient on treatment status in two-level models similar to the main impact model. The Professional Learning with Impact (PLI) group means are unadjusted means; the business-as-usual (BAU) group means are adjusted means computed based on the unadjusted PLI group means and the estimated group mean difference. Standard deviations presented are the pooled standard deviations of the outcome for the PLI and BAU groups. Effect sizes were computed using the Cox index for dichotomous measures and Hedges' *g* for continuous measures, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Exhibit F4. Baseline Equivalence of Student Characteristics for English Language Arts and Math Analytical Samples from the Year 2 Achievement Impact Analysis

Baseline Characteristic or Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
Sample for English Language Arts (ELA) Achievement Analysis								
Female	47.4%	0.499	50.8%	0.500	-3.4%	0.077	-0.082	0.079
Black	28.0%	0.449	27.1%	0.437	0.9%	0.222	0.027	0.838
Hispanic	31.6%	0.465	24.5%	0.432	7.0%	0.139	0.212	0.012
Other race	4.0%	0.195	4.7%	0.208	-0.7%	0.193	-0.105	0.368
Underrepresented minority	61.8%	0.486	56.2%	0.499	5.6%	0.197	0.140	0.240
Special education student	14.4%	0.351	14.8%	0.358	-0.4%	0.109	-0.019	0.775
English learner	10.3%	0.303	11.1%	0.273	-0.9%	0.212	-0.056	0.660
Low socioeconomic status ^a	42.4%	0.494	47.2%	0.487	-4.8%	0.131	-0.118	0.136

Baseline Characteristic or Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
Prior ELA test score	-0.107	0.980	-0.105	0.970	-0.002	0.049	-0.002	0.972
Prior math test score	-0.078	0.916	-0.110	0.992	0.032	0.081	0.033	0.694
Sample for Math Achievement Analysis								
Female	47.5%	0.500	50.6%	0.500	-3.1%	0.077	-0.076	0.102
Black	28.1%	0.450	27.2%	0.437	0.9%	0.222	0.028	0.837
Hispanic	31.6%	0.465	24.5%	0.433	7.0%	0.140	0.212	0.012
Other race	3.9%	0.195	4.6%	0.207	-0.7%	0.193	-0.102	0.382
Underrepresented minority	61.9%	0.486	56.4%	0.499	5.5%	0.198	0.138	0.249
Special education student	14.4%	0.352	14.9%	0.359	-0.5%	0.108	-0.023	0.722
English learner	10.2%	0.302	11.1%	0.273	-1.0%	0.212	-0.061	0.634
Low socioeconomic status ^a	42.6%	0.495	47.4%	0.488	-4.7%	0.131	-0.116	0.145
Prior ELA test score	-0.106	0.980	-0.105	0.971	-0.002	0.050	-0.002	0.971
Prior math test score	-0.077	0.917	-0.112	0.991	0.035	0.081	0.036	0.668

^a In one district, all students were considered low socioeconomic status. As a result, baseline equivalence tests on low socioeconomic status excluded this district due to collinearity.

Note. Sample size ELA achievement = 2,817 (1,492 treatment and 1,325 control). Sample size math achievement = 2,832 (1,496 treatment and 1,336 control). All *p*-values on baseline equivalence tests come from the estimated coefficient on treatment status in two-level models similar to the main impact model. The Professional Learning with Impact (PLI) group means are unadjusted means; the business-as-usual (BAU) group means are adjusted means computed based on the unadjusted PLI group means and the estimated group mean difference. Standard deviations presented are the pooled standard deviations of the outcome for the PLI and BAU groups. Effect sizes were computed using the Cox index for dichotomous measures and Hedges' *g* for continuous measures, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Exhibit F5. Baseline Equivalence of Student Characteristics for English Language Arts and Math Analytical Samples from the Year 2 Achievement Impact Analysis, Using All Four Districts

Baseline Characteristic or Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
Sample for English Language Arts (ELA) Achievement Analysis								
Female	47.3%	0.499	50.9%	0.500	-3.5%	0.074	-0.086	0.055
Black	28.5%	0.452	26.7%	0.441	1.8%	0.237	0.054	0.708
Hispanic	32.7%	0.469	26.2%	0.443	6.5%	0.206	0.191	0.125

Baseline Characteristic or Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
Other race	4.0%	0.197	4.6%	0.204	-0.6%	0.185	-0.083	0.456
Underrepresented minority	63.5%	0.481	58.3%	0.497	5.2%	0.270	0.133	0.415
Special education student	14.7%	0.354	15.5%	0.365	-0.9%	0.103	-0.041	0.516
English learner	11.8%	0.323	12.4%	0.289	-0.6%	0.303	-0.033	0.859
Low socioeconomic status ^a	46.2%	0.499	51.3%	0.495	-5.2%	0.234	-0.125	0.376
Prior ELA test score	-0.085	0.990	-0.069	0.970	-0.015	0.073	-0.016	0.831
Prior math test score	-0.049	0.940	-0.058	1.005	0.008	0.091	0.009	0.927
Sample for Math Achievement Analysis								
Female	47.5%	0.500	50.6%	0.500	-3.2%	0.074	-0.077	0.087
Black	28.5%	0.451	26.8%	0.441	1.6%	0.239	0.050	0.730
Hispanic	33.1%	0.471	26.2%	0.444	6.9%	0.210	0.201	0.114
Other race	4.0%	0.196	4.6%	0.204	-0.6%	0.185	-0.085	0.449
Underrepresented minority	63.9%	0.480	58.4%	0.497	5.5%	0.272	0.140	0.394
Special education student	15.0%	0.357	15.7%	0.365	-0.7%	0.103	-0.034	0.583
English learner	12.4%	0.329	12.8%	0.289	-0.5%	0.308	-0.025	0.893
Low socioeconomic status ^a	46.6%	0.499	52.2%	0.495	-5.6%	0.226	-0.135	0.323
Prior ELA test score	-0.088	0.991	-0.067	0.970	-0.020	0.073	-0.021	0.779
Prior math test score	-0.050	0.940	-0.052	1.003	0.003	0.092	0.003	0.978

^a In one district, all students were considered low socioeconomic status. As a result, baseline equivalence tests on low socioeconomic status excluded this district due to collinearity.

Note. Sample size ELA achievement = 3,058 (1,635 treatment and 1,423 control). Sample size math achievement = 3,081 (1,651 treatment and 1,430 control). All *p*-values on baseline equivalence tests come from the estimated coefficient on treatment status in two-level models similar to the main impact model. The Professional Learning with Impact (PLI) group means are unadjusted means; the business-as-usual (BAU) group means are adjusted means computed based on the unadjusted PLI group means and the estimated group mean difference. Standard deviations presented are the pooled standard deviations of the outcome for the PLI and BAU groups. Effect sizes were computed using the Cox index for dichotomous measures and Hedges' *g* for continuous measures, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Exhibit F6. Baseline Equivalence of Student Characteristics for English Language Arts and Math Analytical Samples from the Year 1 Achievement Impact Analysis

Baseline Characteristic or Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
Sample for English Language Arts (ELA) Achievement Analysis								
Female	47.8%	0.500	47.8%	0.500	-0.1%	0.075	-0.001	0.976
Black	30.0%	0.458	27.8%	0.436	2.2%	0.206	0.065	0.603
Hispanic	30.2%	0.459	28.7%	0.451	1.4%	0.128	0.041	0.594
Other race	4.4%	0.206	4.2%	0.201	0.2%	0.214	0.032	0.802
Underrepresented minority	62.5%	0.484	59.2%	0.496	3.3%	0.199	0.084	0.486
Special education student	13.1%	0.338	13.4%	0.339	-0.2%	0.111	-0.013	0.849
English learner	9.2%	0.289	10.7%	0.294	-1.5%	0.190	-0.099	0.388
Low socioeconomic status ^a	46.0%	0.499	46.3%	0.494	-0.3%	0.134	-0.007	0.936
Prior ELA test score	-0.024	0.957	-0.065	0.897	0.041	0.044	0.044	0.348
Prior math test score	-0.019	0.959	-0.082	0.910	0.063	0.060	0.067	0.296
Sample for Math Achievement Analysis								
Female	47.9%	0.500	48.0%	0.500	0.0%	0.075	0.000	0.996
Black	29.8%	0.458	27.8%	0.436	2.0%	0.207	0.060	0.633
Hispanic	30.0%	0.459	29.0%	0.452	1.0%	0.127	0.030	0.697
Other race	4.4%	0.206	4.2%	0.201	0.2%	0.215	0.033	0.802
Underrepresented minority	62.3%	0.485	59.5%	0.496	2.8%	0.199	0.071	0.553
Special education student	13.2%	0.339	13.2%	0.337	0.1%	0.113	0.004	0.957
English learner	9.3%	0.290	10.7%	0.295	-1.4%	0.189	-0.097	0.397
Low socioeconomic status ^a	46.1%	0.499	46.2%	0.494	-0.1%	0.134	-0.003	0.970
Prior ELA test score	-0.023	0.955	-0.065	0.896	0.041	0.043	0.045	0.340
Prior math test score	-0.018	0.958	-0.082	0.909	0.064	0.060	0.069	0.283

^a In one district, all students were considered low socioeconomic status. As a result, baseline equivalence tests on low socioeconomic status excluded this district due to collinearity.

Note. Sample size ELA achievement = 2,928 (1,532 treatment and 1,396 control). Sample size math achievement = 2,936 (1,535 treatment and 1,401 control). All *p*-values on baseline equivalence tests come from the estimated coefficient on treatment status in two-level models similar to the main impact model. The Professional Learning with Impact (PLI) group means are unadjusted means; the business-as-usual (BAU) group means are adjusted means computed based on the unadjusted PLI group means and the estimated group mean difference. Standard deviations presented are the pooled standard deviations of the outcome for the PLI and BAU groups. Effect sizes were computed using the Cox index for dichotomous measures and Hedges' *g* for continuous measures, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Appendix G. Impact of Professional Learning with Impact on Supplemental Outcomes From the Teacher Survey in Program Year 2 and Program Year 1

Exhibit G1. The Effect of Professional Learning with Impact (PLI) on Supplemental Outcomes From the Teacher Survey in Program Year 2

Teacher Survey Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value	Obs.
	Mean	SD	Mean	SD					
Student-centered beliefs about teaching ^a	3.02	0.40	2.97	0.50	0.05	0.11	0.11	0.642	63
Situated teacher self-efficacy for instructional improvement ^a	3.23	0.64	3.22	0.65	0.01	0.15	0.01	0.956	63
Collegial support ^a	3.32	0.53	3.47	0.49	-0.14	0.13	-0.28	0.276	59
Insights about how to teach	2.89	0.62	2.86	0.48	0.02	0.15	0.04	0.876	71
Confidence in effectiveness of some of my normal teaching practices	3.15	0.38	3.24	0.45	-0.09	0.11	-0.21	0.418	71
Efforts to teach better	3.40	0.75	3.47	0.43	-0.07	0.14	-0.11	0.625	71
Perceived success at responding to student needs	3.51	0.46	3.34	0.38	0.17	0.09	0.40	0.059	71
Perception of instructional performance ^a	3.04	0.47	2.95	0.54	0.08	0.12	0.17	0.487	63
Data use ^a	4.53	0.42	4.32	0.51	0.21	0.13	0.44	0.114	54
Stress frequency	4.36	0.76	4.14	0.74	0.22	0.18	0.29	0.223	70
Coping with stress	2.75	1.00	2.99	1.14	-0.24	0.31	-0.22	0.435	70
Symptoms of depression	2.13	1.01	1.88	0.97	0.24	0.24	0.24	0.321	70
Job satisfaction ^a	2.71	0.66	2.71	0.79	0.00	0.19	0.00	0.992	59

^a Denotes an outcome model that included a baseline measure of the outcome.

Note. See Exhibit F1 for sample sizes by treatment condition. All *p*-values come from the estimated coefficient on treatment status in two-level models that nest teachers within schools. The PLI group means are unadjusted means; the business-as-usual group means are adjusted means computed based on the unadjusted PLI group means and the estimated group mean difference. Effect sizes were computed using Hedges' *g*, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Exhibit G2. The Effect of Professional Learning with Impact (PLI) on Supplemental Outcomes From the Teacher Survey in Program Year 1

Teacher Survey Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
Self-efficacy ^a	3.32	0.31	3.33	0.33	-0.01	0.06	-0.05	0.799
Perception of instructional performance ^a	3.33	0.47	3.27	0.48	0.06	0.07	0.13	0.368
Situated teacher self-efficacy for instructional improvement ^a	3.18	0.65	3.31	0.63	-0.13	0.11	-0.20	0.220
Student-centered beliefs about teaching ^a	3.01	0.38	2.91	0.34	0.10	0.07	0.28	0.127
Data use ^a	4.53	0.54	4.57	0.52	-0.04	0.08	-0.08	0.625
Collegial support ^a	3.52	0.52	3.63	0.44	-0.10	0.07	-0.21	0.169
Stress frequency	4.18	0.72	3.90	0.76	0.28	0.14	0.38	0.044
Coping with stress	3.00	0.93	3.28	1.15	-0.28	0.18	-0.27	0.132
Symptoms of depression	2.08	0.93	1.86	0.87	0.23	0.18	0.25	0.201
Job satisfaction ^a	2.84	0.59	3.07	0.63	-0.23	0.09	-0.38	0.013

^a Denotes an outcome model that included a baseline measure of the outcome.

Note. Sample size = 102 teachers (56 treatment and 46 control). All *p*-values come from the estimated coefficient on treatment status in two-level models that nest teachers within schools and include baseline measures of the outcome. The PLI group means are unadjusted means; the business-as-usual group means are adjusted means computed based on the unadjusted PLI group means and the estimated group mean difference. Effect sizes were computed using Hedges' *g*, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Appendix H. Impact of Professional Learning with Impact (PLI) on Supplemental Student Outcomes

Impact of PLI on Student Outcomes in Program Year 2, by English Learner Status

Exhibit H1. Impacts of Professional Learning with Impact on Fourth-Grade English Language Arts and Math Achievement in Year 2, by English Learner Status

Student Characteristics	Treatment Group			Control Group			Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	N	Mean	SD	N				
English Learners										
Fourth-grade English language arts (ELA) achievement	-0.28	0.89	153	-0.56	0.86	107	0.28	0.09	0.32	0.001
Fourth-grade math achievement	-0.21	0.83	152	-0.41	0.86	108	0.21	0.09	0.25	0.015
Non-English Learners										
Fourth-grade ELA achievement	-0.05	0.98	1339	-0.11	0.99	1218	0.06	0.04	0.06	0.110
Fourth-grade math achievement	-0.07	0.97	1344	-0.10	0.96	1228	0.03	0.05	0.04	0.504

Note. The treatment group means are unadjusted means; the control group means are adjusted means computed based on the unadjusted treatment group means and the estimated group mean difference. Effect sizes were computed using Hedges' *g*, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Impact of PLI on Student Outcomes in Program Year 2, Including All Four Districts

Exhibit H2. Impacts of Professional Learning with Impact on Fourth-Grade English Language Arts and Math Achievement in Year 2

Student Achievement Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
Fourth-grade English language arts (ELA) achievement	-0.04	1.00	-0.14	0.99	0.10	0.04	0.10	0.016
Fourth-grade math achievement	-0.05	0.99	-0.09	0.97	0.04	0.06	0.04	0.557

Note. Sample size ELA achievement = 3,058 (1,635 treatment and 1,423 control). Sample size math achievement = 3,081 (1,651 treatment and 1,430 control). The treatment group means are unadjusted means; the control group means are adjusted means computed based on the unadjusted treatment group means and the estimated group mean difference. Effect sizes were computed using Hedges' *g*, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Exhibit H3. Differential Impacts of Professional Learning with Impact (PLI) on Fourth-Grade English Language Arts and Math Achievement in Year 2, by Student Characteristics

Student Characteristics	Impact Estimate for X = 0	Impact Estimate for X = 1	Difference in Impact	Standard Error	P-Value
Outcome: Fourth-Grade English Language Arts (ELA) Achievement					
Prior achievement	0.099	0.063	-0.036	0.028	0.197
Eligibility for free or reduced-price lunch	0.079	0.125	0.046	0.057	0.423
English learner status	0.075	0.321	0.245	0.087	0.005
Special education status	0.082	0.200	0.118	0.070	0.090
Underrepresented minority status	0.120	0.088	-0.033	0.054	0.543
Outcome: Fourth-Grade Math Achievement					
Prior achievement	0.036	0.066	0.030	0.027	0.001
Eligibility for free or reduced-price lunch	0.054	0.015	-0.039	0.057	0.920
English learner status	0.015	0.202	0.188	0.084	0.025
Special education status	0.040	0.008	-0.032	0.067	0.630
Underrepresented minority status	0.043	0.030	-0.014	0.052	0.351

Note. Sample size ELA achievement = 3,058 (1,635 treatment and 1,423 control). Sample size math achievement = 3,081 (1,651 treatment and 1,430 control). 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 the district mean. For analyses of differential impact by other student characteristics, X = 1 for female students, students eligible for free or reduced-price lunch, English learners, special education students, and non-White students; X = 0 otherwise. The listed effect of PLI when the trait is equal to 0 comes from the estimated coefficient on treatment status; the listed effect of PLI when the trait is equal to 1 is the sum of the estimated coefficient on treatment status and the estimated coefficient on the interaction between treatment status and the listed trait. The listed difference is the estimated coefficient on the interaction, and the *p*-value is the *p*-value on that interaction.

Impact of PLI on Student Outcomes in Program Year 1

Exhibit H4. Impacts of Professional Learning with Impact on Fourth-Grade English Language Arts and Math Achievement in Year 1

Student Achievement Measure	Treatment Group		Control Group		Treatment–Control Difference	Standard Error	Effect Size	P-Value
	Mean	SD	Mean	SD				
Fourth-grade English language arts (ELA) achievement	-0.04	0.98	-0.04	0.96	0.00	0.03	0.00	0.95
Fourth-grade math achievement	-0.04	0.96	-0.04	0.95	0.00	0.04	0.00	0.95

Note. Sample size ELA achievement = 2,928 (1,532 treatment and 1,396 control). Sample size math achievement = 2,936 (1,535 treatment and 1,401 control). The treatment group means are unadjusted means; the control group means are adjusted means computed based on the unadjusted treatment group means and the estimated group mean difference. Effect sizes were computed using Hedges' *g*, as recommended by the What Works Clearinghouse (2022). SD = standard deviation.

Exhibit H5. Differential Impacts of Professional Learning with Impact (PLI) on Fourth-Grade English Language Arts and Math Achievement in Year 1, by Student Characteristics

Student Characteristics	Impact Estimate for X = 0	Impact Estimate for X = 1	Difference in Impact	Standard Error	P-Value
Outcome: Fourth-Grade English Language Arts (ELA) Achievement					
Prior achievement	-0.001	-0.068	-0.067	0.025	0.008
Eligibility for free or reduced-price lunch	0.000	0.004	0.004	0.044	0.928
English learner status	0.008	-0.056	-0.063	0.073	0.384
Special education status	0.016	-0.094	-0.110	0.062	0.075
Underrepresented minority status	0.055	-0.035	-0.091	0.044	0.037
Outcome: Fourth-Grade Math Achievement					
Prior achievement	-0.001	-0.087	-0.086	0.025	0.001
Eligibility for free or reduced-price lunch	0.005	0.000	-0.005	0.046	0.920
English learner status	0.007	-0.042	-0.049	0.073	0.501
Special education status	0.013	-0.064	-0.077	0.061	0.211
Underrepresented minority status	0.027	-0.014	-0.041	0.044	0.351

Note. Sample size ELA achievement = 2,928 (1,532 treatment and 1,396 control). Sample size math achievement = 2,936 (1,535 treatment and 1,401 control). 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 the district mean. For analyses of differential impact by other student characteristics, X = 1 for female students, students eligible for free or reduced-price lunch, English learners, special education students, and non-White students; X = 0 otherwise. The listed effect of PLI when the trait is equal to 0 comes from the estimated coefficient on treatment status; the listed effect of PLI when the trait is equal to 1 is the sum of the estimated coefficient on treatment status and the estimated coefficient on the interaction between treatment status and the listed trait. The listed difference is the estimated coefficient on the interaction, and the *p*-value is the *p*-value on that interaction.

Appendix I. Additional Information About the Cost Study

Valuing Professional Impact with Learning (PLI) Teacher, Coach, and Program Specialist Time

Teacher's Salary. We estimated the value of time invested by teachers in PLI programming by calculating a national weighted average salary based on 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.²⁴

$$\text{Weighted Average} = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i}$$

To arrive at this salary estimate, we used the weighted average equation above, where i represents the number of years of teaching experience, n represents the total number of teachers, w represents the salary weights, and X represents NCES salary. We assigned salary weights (w_i) based on the number of years of teaching experience reported by PLI teachers in the teacher survey. For example, five PLI teachers had 5 years of teaching experience (out of a sample size of 83 teachers), thus we assigned a weight of 0.06. Next, we multiplied the assigned weights by the corresponding NCES national average salaries for the same qualifications (i.e., degree) and years of experience as PLI teachers to get the proportion of weighted salary for a given number of years of teaching experience ($w_i X_i$). For example, three out of the five PLI teachers with 5 years of experience held a bachelor's degree, which corresponds to a \$51,230 national average salary; the other two teachers held a master's degree, which corresponds to a \$60,050 national average salary. We took the average of these salaries and multiplied by the assigned weight ($w_{i=5} = 0.06$) to get the proportion of weighted salary for the PLI teachers with 5 years of teaching experience ($w_{i=5} X_{i=5} = \$3,298$). We then repeated this process for all years of teaching experience reported by PLI teachers, summed all of the weighted salary proportions, and divided by the sum of the weights to obtain the annual national weighted average salary for all PLI teachers (\$64,668).

²⁴ 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

Coach’s and Program Specialist’s Salaries. We estimated coach and program specialist salaries based on actual participants’ salaries. We averaged 19 coach salaries across four school districts to get an overall average salary for all PLI coaches. Similarly, we averaged program specialists’ actual loaded hourly salary rates to get an overall average hourly compensation rate for all PLI program specialists.²⁵

Adjusting Salaries to Account for Compensation and Inflation

After estimating annual salaries for teachers, coaches, and program specialists, we applied a 35% benefit rate to the teacher and coach salary estimates to account for overall compensation (pay and benefits).²⁶ We assumed that the benefit rate for PLI coaches and teachers is equal to the average rate for instructional staff in the U.S. Bureau of Labor Statistics, Employer Costs for Employee Compensation data.²⁷ Next, we adjusted the teacher, coach, and program specialist 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) (1.058% inflation rate).²⁸ We then converted coach and teacher annual national overall compensation to hourly rates by dividing by 1,440 hours worked per year.^{29,30,31,32} Exhibit I1 reports the hourly compensation rates used for program staff in Year 1 and Year 2.

Exhibit I1. Compensation Rates for Personnel (in 2023 Dollars)

Position	Year 1 Hourly Rate	Year 2 Hourly Rate
Teacher	\$66.14	\$64.21
Coach	\$54.31	\$52.73
Program Specialist	\$163.40	\$150.00

²⁵ Program specialist hourly salary rates came from Learning Forward and Danielson Group invoices.

²⁶ Benefit rate source: <https://www.bls.gov/opub/btn/volume-8/teachers-job-requirements.htm>

²⁷ Benefit rate source: <https://www.bls.gov/opub/btn/volume-8/teachers-job-requirements.htm>

²⁸ U.S. Bureau of Labor Statistics Consumer Price Index for All Urban Consumers (CPI-U), 2022–23. <https://www.bls.gov/cpi/>

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 midlevel representation of the number of hours worked in a K–12 school year and is based on national survey data and an 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 (one standard school year), or 8 hours per day at 180 days per year, which equals 1,440 annual work hours.

Cost Analysis Calculations

Teacher, coach, and program specialist (i.e., program staff) estimates of time allocation and the numbers of staff involved in each program year (i.e., Year 1 and Year 2) were multiplied by the compensation rates to estimate the overall cost of program staff time for both years of the program. The total cost of program staff time for each program activity (outlined in Exhibit 25. PLI Program Components) was summed to get a total cost for the “Teacher Time,” “Coach Time,” and “Program Specialist Time” program components. Then, we divided the cost for each program component (i.e., Teacher Time, Coach Time, Program Specialist Time, Program Fees) by the number of teachers served in the respective program year to estimate the year-specific program component costs on a per-teacher basis. The 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 3.0% discount rate so that the costs reflect those at the time of program completion.³³ The annual per-teacher costs in Year 1 and Year 2 were then aggregated to generate the overall per-teacher cost associated with teacher time, coach time, program specialist time, and program fees necessary for program implementation. These calculations are expressed in the equations that follow, where i represents a given PLI program activity; *Coach Time* and *Teacher Time* are the total amount of hours staff spent on program activity i ; *Coach Compensation* and *Teacher Compensation* are the fixed, overall staff compensation (pay and benefits); and *Number of Teachers* is the total number of treatment teachers in a given year (Year 1 or 2) of the program.

*Total Per Teacher Cost*_{Y1} =

$$\frac{\left[\left(\sum_{i=1}^I \text{Coach Time}_{i,Y1} \times \overline{\text{Coach Compensation}} \right) + \left(\sum_{i=1}^I \text{Teacher Time}_{i,Y1} \times \overline{\text{Teacher Compensation}} \right) + \left(\sum_{i=1}^I \text{Program Specialist Time}_{i,Y1} \times \overline{\text{Program Specialist Compensation}} \right) + (\text{Program Fees}_{Y1}) \right]}{\text{Number of Teachers}_{Y1}}$$

*Total Per Teacher Cost*_{Y2} =

$$\frac{\left[\left(\sum_{i=1}^I \text{Coach Time}_{i,Y2} \times \overline{\text{Coach Compensation}} \right) + \left(\sum_{i=1}^I \text{Teacher Time}_{i,Y2} \times \overline{\text{Teacher Compensation}} \right) + \left(\sum_{i=1}^I \text{Program Specialist Time}_{i,Y2} \times \overline{\text{Program Specialist Compensation}} \right) + (\text{Program Fees}_{Y2}) \right]}{\text{Number of Teachers}_{Y2}}$$

Overall Per Teacher Cost = *Per Teacher Cost*_{Y1} + *Per Teacher Cost*_{Y2}

³³ See 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>

About the American Institutes for Research®

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit institution 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](https://www.air.org).



AIR® Headquarters

1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | [AIR.ORG](https://www.air.org)



Notice of Trademark: "American Institutes for Research" and "AIR" are registered trademarks. All other brand, product, or company names are trademarks or registered trademarks of their respective owners.

Teacher Professional Learning in Grade-Level Teams: The Implementation and Impact of a Program Delivered During the COVID-19 Pandemic: Final Evaluation Report. September 2024. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International ([CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)) License.

The contents of this report were developed under a grant from U.S. Department of Education, Office of Elementary and Secondary Education, Award No. U411B180032; However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

24987_07/24