Evaluation of iCoachTEAM | Final Report

June 2026

Prepared for:
Nicola Hodkowski
Director, Mathematics Education Research
Center for Learning Sciences Research
nhodkowski@digitalpromise.org

Prepared by:
SRI
Lauren Cassidy
Paul Burkander
Santiago Navia Jaramillo
Marta Mielicki

June 2026

Acknowledgments:

The authors thank Charles Harding, Leilani Lopez, Sophia Ouyang, Rebecca Schmidt, Erin Smith, and Kaily Yee for their contributions to the evaluation and to this report. The authors also thank Nicola Hodkowski and Viki Young at Digital Promise for their strong engagement with and feedback on the evaluation, as well as the iCoachTEAM Program Team members who spoke to us about the professional learning they provided. Most important, the authors are grateful to the staff in the partner districts and schools—district leaders, principals, coaches, and teachers—who gave generously of their time and insight in interviews, on surveys, and during classroom observations. Without them, this evaluation would not have been possible.

This report was prepared by SRI International with funds provided by Digital Promise under an Education Innovation and Research (EIR) Early-phase grant from the U.S. Department of Education (grant S411C200054). The views expressed in this report do not necessarily reflect the views or policies or the U.S. Department of Education.

Suggested citation:
Cassidy, L., Burkander, P., Navia Jaramillo, S., & Mielicki, M. (2026). *Evaluation of iCoachTEAM: Final report*. SRI.

Contents

List of Exhibits

Abstract

Nationally, students' math achievement has continued a downward trend since the COVID-19 pandemic. Digital Promise sought to address this challenge through Instructional Coaching for Tech-Enhanced Approaches to Mathematics (iCoachTEAM), a program that trained coaches to support middle school math teachers to use technology to promote student conceptual understanding and conduct formative assessment. Digital Promise received an Education Innovation and Research (EIR) Early-phase grant to pilot and test the efficacy of iCoachTEAM at improving student math achievement, with the 5-year grant period beginning in 2021 and extending through 2025. It partnered with SRI to examine the implementation and impact of the program.

This technical report describes a quasi-experimental study of iCoachTEAM, implemented in two pilot districts and four implementation districts. SRI matched schools within iCoachTEAM districts (treatment) to other schools in the same state implementing business-as-usual math coaching practices (comparison). The study compared math achievement on state standardized tests in grades 6–8 in the treatment and comparison schools after two years of iCoachTEAM implementation. SRI measured teacher practice outcomes using an observation tool developed by Digital Promise to capture technology use and the Danielson Framework for Teaching.

Despite medium to high implementation fidelity of the intervention, the study found no effects of iCoachTEAM on student or teacher outcomes. Under an intent-to-treat design, the study included all teachers in a study school, many of whom did not receive coaching, resulting in limited contrast between the treatment and comparison conditions. Moreover, the observation sample was too small to detect significant differences. However, the implementation study uncovered several implications for supporting teachers to move beyond the use of technology for procedural tasks and toward approaches that build conceptual understanding.

Executive Summary

Instructional Coaching for Tech-Enhanced Approaches to Mathematics (iCoachTEAM) is a school-based, one-to-one coaching intervention designed to improve middle school math instruction and increase math engagement and achievement of students whose test scores are a year or more behind grade level. Under a federal Education Innovation and Research (EIR) Early-phase grant, the iCoachTEAM project trained and supported instructional coaches to provide coaching to middle school math teachers. The coaching focused on using technology to support students' conceptual understanding and to conduct formative assessment. Digital Promise contracted with SRI to conduct an independent evaluation of the impact and implementation of iCoachTEAM.

Digital Promise piloted iCoachTEAM with five districts in the 2022–24 school years (Cohort 1) and used learnings from that two-year early implementation to inform programming for the next cohort. Digital Promise partnered with four districts in the 2023–24 school year (Cohort 2) to implement iCoachTEAM for two years. SRI's evaluation was initially meant to include only Cohort 2 schools, but the program team encountered recruitment challenges stemming from the COVID-19 pandemic. As a result, the number of schools implementing iCoachTEAM in 2023–25 was not large enough to detect meaningfully large treatment effects. Therefore, SRI included two of the five Cohort 1 districts in the impact analysis that were in the same state as one of the Cohort 2 districts. All Cohort 2 schools were included in both the impact and implementation studies.

SRI evaluated the impact of iCoachTEAM using a quasi-experimental design in which schools in iCoachTEAM districts (treatment) were matched to other schools in the same state that were implementing business-as-usual practices (comparison). The analysis sample for the impact study included four Cohort 1 schools that piloted iCoachTEAM in 2022–23 and 2023–24 and 12 Cohort 2 schools that implemented in in 2023–24 and 2024–25. The study compared math achievement on state standardized tests in grades 6–8 in the treatment and comparison schools. SRI measured teacher practice for coached Cohort 2 teachers only, using an observation tool designed by Digital Promise as well as the Framework for Teaching (adapted from Danielson, 2013; The Danielson Group, 2022). SRI observed teachers in spring 2023 (baseline) and again in spring 2025 (outcome).

To understand implementation in Cohort 2 schools, SRI administered teacher and coach surveys each spring, collected iCoachTEAM program data from Digital Promise at the end of each school year, and interviewed a sample of teachers and coaches in each district in 2024–25. Three districts implemented the program in all their middle schools, and the fourth, largest district implemented it in half of its middle schools; not all teachers in the participating schools received coaching due to capacity constraints. District- or school-based instructional coaches received training and ongoing mentoring to coach middle school math teachers (grades 6–8). Two to three coaches in each district, some of whom were also teachers, worked with select math teachers (except in one small district where all math teachers participated).

Over the course of the evaluation, implementation was mixed. District leaders engaged with the program, and coaches received and appreciated strong mentoring support from the iCoachTEAM Program Team. Coaches needed support to build coaching skills and math content knowledge before they could address technology use. Coaches were able to complete two coaching cycles with most of their teachers, and coaching focused more on math content, instructional strategies, and the use of technology over time. In the end, there was no significant difference in student's math scores between the treatment and comparison schools or in teachers' practice after two years of implementation.

Introduction

Improving students' math achievement is a national imperative. Since the COVID-19 pandemic, U.S. assessment data have shown the largest declines in math scores in decades for students in grades 4 and 8 (National Center for Education Statistics, 2022). Achievement gaps between student groups also persist, particularly for students in underresourced schools who experience more instructional interruption, less access to experienced math teachers, and fewer advanced math opportunities (Kuhfield et al., 2022). Improving math instruction, especially for students who test one year or more behind grade level, has important implications for increasing student achievement and reducing inequities.

Research suggests that improving math outcomes requires moving beyond procedural instruction and toward approaches that build conceptual understanding (Henningsen & Stein, 1997; Hill, 2001). Students develop deeper understanding when teachers use formative assessment practices that elicit and respond to student mathematical reasoning. Technology-supported instructional practices can further enhance learning when they provide immediate feedback and reinforce conceptual understanding. Studies have found small to moderate positive effects on student achievement under these conditions (Cheung & Slavin, 2013; Koedinger et al., 2010).

Instructional Coaching for Tech-Enhanced Approaches to Mathematics (iCoachTEAM) is designed to combine these components of high-quality math instruction. The program uses instructional coaching as the primary mechanism for helping teachers learn how to use technology strategically to conduct formative assessment and promote conceptual understanding. This approach is grounded in evidence from a meta-analysis of more than 80 studies demonstrating that instructional coaching can positively affect both teacher practice and student achievement (Kraft et al., 2018). Supporting middle school coaches to coach in math is particularly important, as coaches at this grade band tend to be generalists rather than math specialists (Campbell & Griffin, 2017). By training middle school coaches to coach in math, iCoachTEAM seeks to build the internal capacity of schools and districts to support and sustain effective math instruction after the project ends.

Under a federal Education Innovation and Research (EIR) Early-phase grant, Digital Promise piloted iCoachTEAM in five districts (Cohort 1) in the 2022–23 and 2023–24 school years and implemented it in four districts (Cohort 2) in the 2023–24 and 2024–25 school years. Digital Promise contracted with SRI to conduct an independent evaluation of the impact and implementation of iCoachTEAM.

This report explores how iCoachTEAM's comprehensive approach can impact student outcomes. It begins with an overview of the program, followed by descriptions and results of the impact and implementation studies. The findings may provide important evidence for policymakers and practitioners to consider as they look to scale effective approaches to improving math achievement.

Overview of iCoachTEAM

iCoachTEAM was a two-year program that provided professional learning and supports for instructional coaches to help middle school math teachers use technology to promote students' conceptual understanding and to conduct formative assessment. Coaches engaged in coaching cycles with teachers, which consisted of goal setting, observations, and reflections on their practice. The iCoachTEAM Program Team piloted the program first with Cohort 1 districts and used the pilot learnings to inform program implementation for Cohort 2 districts. Cohort 1 piloted iCoachTEAM in the 2022–23 and 2023–24 school years, and Cohort 2 implemented the program in the 2023–24 and 2024–25 school years.

Program Description

iCoachTEAM was designed to build the capacity of coaches to support teachers to strategically use technology in their math instruction to improve student engagement and achievement in math (see the logic model in Figure 1).

Program Inputs

The iCoachTEAM Program Team developed professional learning and resources for coaches to learn how to support teachers in using technology to promote conceptual understanding and conduct formative assessment. The program team also worked with district partners to ensure adequate leader engagement to support coaching in their districts and schools. Leader-specific sessions at annual summer and winter institutes and quarterly meetings provided time for vison and goal setting, planning, and reviewing continuous improvement data.

Program Activities

The iCoachTEAM program's core activities were designed to change coach practice, which was hypothesized to change math instruction and ultimately improve student achievement. The core activities included (1) professional development for instructional coaches who may or may not have math backgrounds, (2) coaching of middle school math teachers, and (3) organizational support structures that set the conditions for coaching to happen.

Instructional Coach Development. Each year, the iCoachTEAM Program Team, which consisted of math educators, delivered an in-person, five-day summer institute; a virtual, two-day winter institute; and virtual monthly professional development. A coach mentor, who had previous experience as a math/science teacher and instructional coach, provided at least monthly virtual one-to-one mentoring to coaches.

The professional learning focused on developing coaches' math content knowledge for teaching and coaching in the middle school grades, building their fundamental coaching skills and practices, and increasing their understanding of how to use technology in math instruction for the iCoachTEAM program's stated purposes. Coaches had access to tools, resources, and strategies aligned with common teacher challenges.

Coaching. Middle school math teachers in grades 6–8 received one-to-one coaching from coaches who participated in the professional development. As part of the intervention, coaches were expected to provide each of their coached teachers with a full dosage of coaching—two completed coaching cycles per year that lasted 6–8 weeks each. For each cycle, coaches and teachers partnered to identify student learning goals and strategies for using technology-enhanced math instruction, observed and debriefed teachers' instruction, and helped teachers reflect on their progress in implementing the strategies over time.

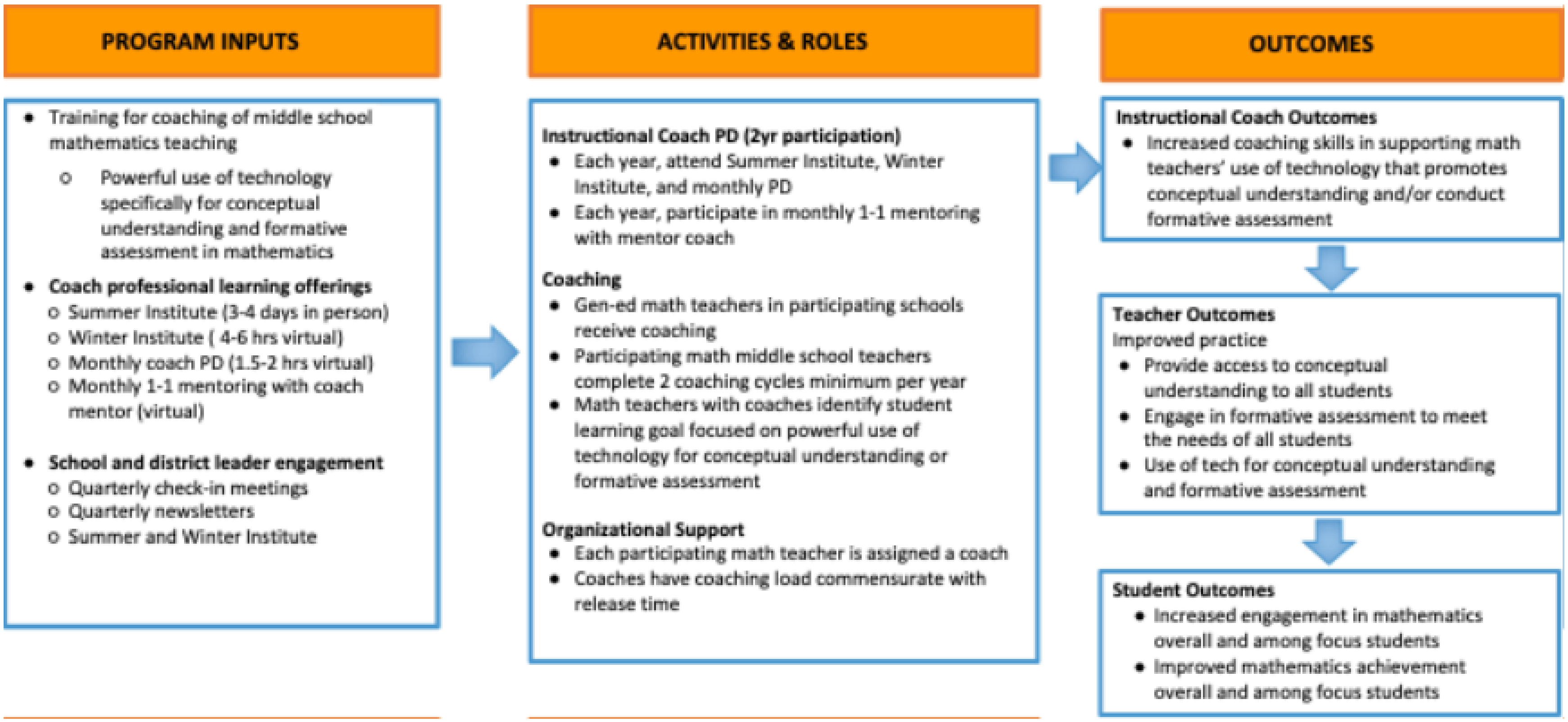
Coaches and teachers worked together during the school day. They met during planning time or before or after school to set learning goals and to debrief and reflect on their practice following classroom observations. In general, full-time coaches were able to meet with their teachers more consistently than coaches who were also teachers.

Organizational Support Structures. The original intent of iCoachTEAM was that all middle school math teachers would have an assigned coach and coaches would have release time commensurate with their coaching responsibilities. However, most districts only had 2–3 coaches who coached 2–3 teachers each. All teachers received coaching in only one district—a small district that had two coaches and six middle school math teachers. Ideally, schools and districts would also give coaches sufficient time in their schedules to engage in coaching. However, coaches in two districts were also teachers, which inhibited their time and flexibility to provide coaching.

Outcomes

These activities were intended to increase coaches' skills in supporting math teachers to use technology to promote conceptual understanding and conduct formative assessment, which would then improve teacher practice in these areas. Ultimately, stronger instruction would lead to higher student engagement in math and improved student achievement, especially for students who were one year or more behind grade level in math.

Figure 1. iCoachTEAM Logic Model



Evaluation of iCoachTEAM

The evaluation of iCoachTEAM included an impact study and an implementation study. For the main impact study, SRI used a quasi-experimental design to compare the math standardized test achievement of students in treatment schools with that of students in matched comparison schools. SRI also explored changes in instructional practice for coached teachers. For the implementation study, SRI used program, survey, and interview data to examine the extent to which districts implemented the program as intended and common facilitators and barriers to implementation.

Study Districts

The iCoachTEAM Program Team recruited districts that had a minimum of 50% of middle school students with math scores that were one year or more behind grade level. The original intent was for study districts to have preexisting coaching structures. However, relatively few districts met this criterion, and districts without preexisting coaching structures wanted to participate, in part to expand their math coaching capacity. The team therefore modified the eligibility criteria to include districts without preexisting coaching structures.

SRI's evaluation was initially designed to include only Cohort 2 districts that implemented the program in the 2023–24 and 2024–25 school years. Because of sample size challenges, two of the Cohort 1 districts that piloted in 2022–23 and 2023–24 were included in the impact study. These districts were located in the same state as one of the Cohort 2 districts and implemented the program for two years with no significant disruption to coaching. They could not be included in the implementation study because the external evaluation began after Cohort 1's initial implementation phase.

In the end, the evaluation included two districts in Cohort 1 and four districts in Cohort 2 (Table 1). These districts were located across the United States in suburban and rural areas. The evaluation included schools serving grades 6–8. Two Cohort 1 districts and all Cohort 2 districts participated in the impact study, and only Cohort 2 districts participated in the implementation study.

Table 1. Characteristics of Participating Districts

District	State	Region	Locale	Student enrollment (to nearest 100th)	Total schools serving grades 6-8
Cohort 2					
District A	IL	Midwest	Small, suburban	1,900 PK–8	2
District B	CA	West Coast	Midsize, suburban	11,100 K–12	3
District C	PA	East Coast	Small, suburban	1,800 K–12	1
District D	CA	West Coast	Small, rural	4,000 K–12	6
Cohort 1					
District E	PA	East Coast	Small, suburban	6,000 PK–12	2
District F	PA	East Coast	Small, rural	900, KG–12	2

Within iCoachTEAM districts, all schools serving any of grades 6–8 and all math teachers within those schools were offered the opportunity to participate in the coaching intervention, subject to coaching capacity. District or school leaders identified the coaches, who included full-time district-level coaches, assistant principals (in one district), and school-based coaches who also had teaching roles. How teachers were identified for coaching varied by district based on factors such as teachers' grade level, years of experience teaching, and availability, or if they volunteered to be coached. In the end, not all teachers within each school were coached because of coaching capacity, district decisions, and/or teacher choice.

Research Questions

The evaluation addressed research questions about the impact and implementation of iCoachTEAM. Because the goal of the intervention was to improve student math achievement after two years, the confirmatory impact research question tested whether iCoachTEAM increased the percent of students in intervention schools who were proficient or above proficient on the state standardized test in math after two years. Exploratory impact research questions examined effects on school-level level proficiency after one year, mediation effects, effects across proficiency rates, and changes in teacher practice after two years. Implementation research questions considered fidelity, facilitators, and barriers.

Impact Study

1. Does iCoachTEAM result in increased percentages of students who are proficient or above proficient based on state standardized tests in math in grades 6, 7, and 8, as

compared with the business-as-usual condition, after two years of coaching? (Confirmatory)

2. Does iCoachTEAM result in increased percentages of students who are proficient or above proficient based on state standardized tests in math in grades 6, 7, and 8, as compared with the business-as-usual condition, after one year of coaching? (Exploratory)

3. Does iCoachTEAM result in decreased percentages of students at the lowest proficiency level—that is, those who were not proficient, compared with those who were nearly proficient,[1] proficient, or above proficient—based on state standardized tests in math in grades 6, 7, and 8, as compared with the business-as-usual condition, after one and two years of coaching? (Exploratory)

4. At the school level, is there a relationship between the proportion of teachers in the school receiving iCoachTEAM coaching and changes in student proficiency rates in math in that school? (Exploratory)

5. Does teacher instructional practice improve after two years of iCoachTEAM coaching, as measured by the Math Observations of Technology for Conceptual Understanding (MOT-CU) and Formative Assessment (MOT-FA) rubrics and the Framework for Teaching observation rubric? (Exploratory)

Implementation Study

6. To what extent were the core components of iCoachTEAM implemented as intended?
7. What contextual factors impeded or enhanced implementation?

Data Sources

To answer the research questions for the impact and implementation studies, SRI collected publicly available extant data and conducted primary data collection in intervention districts.

Extant Data. SRI collected publicly available school-grade-level data from state standardized test scores from the 2021–22 school year through the 2024–25 school year (Table 2). SRI collected data for the California Assessment of Student Performance and Progress (CAASPP) Smarter Balanced Assessments, the Pennsylvania System of School Assessment (PSSA), and the Illinois Assessment of Readiness (IAR).

[1] Although students who are nearly proficient are not proficient, states in the study sample use “not proficient” and “nearly proficient” to distinguish the lowest two proficiency levels.

Table 2. Use of publicly available data by cohort and school year

School year	2021–22	2022–23	2023–24	2024–25
Cohort 1	Baseline/matching	Year 1 outcomes	Year 2 outcomes	
Cohort 2		Baseline/matching	Year 1 outcomes	Year 2 outcomes

Classroom Observations. SRI observed teachers in fall 2023, before coaching was received (baseline), and in spring 2025, after two years of implementation (outcome). Observers were trained to use multiple rubrics: the Math Observations of Technology for Conceptual Understanding (MOT-CU) and Formative Assessment (MOT-FA), validated instruments (Hodkowski et al., under review) designed by Digital Promise specifically to assess the use of technology in math instruction to promote conceptual understanding or conduct formative assessment; and the Framework for Teaching (Danielson, 2013; The Danielson Group, 2022).

iCoachTEAM Program Data. At the end of the 2023–24 and 2024–25 school years, SRI collected program data from the iCoachTEAM Program Team. These data included attendance data for professional development and coaching logs of coach-teacher interactions. The iCoachTEAM Program Team documented attendance at each of the professional learning offerings, including the summer and winter institutes, quarterly meetings, monthly professional development, and monthly mentoring sessions. In coaching logs, the coaches documented the focus of coach and teacher meetings and their progress through the coaching cycles. Coaches noted teachers' coaching goals, the strategies they would use to meet the goals, and how technology would be used.

Coach and Teacher Surveys. In spring 2024 and spring 2025, SRI administered a survey to participating coaches and teachers (see Table 3 for response rates). The surveys asked coaches about the focus and value of the iCoachTEAM training they received and about the coaches' confidence in their coaching of middle school math teachers. Both coaches and teachers responded to questions about the focus of coaching and the use of technology. The teacher survey asked teachers how they used technology in their math instruction to promote conceptual understanding and conduct formative assessment.

Table 3. Survey Response Rates

Survey	Spring 2024	Spring 2025
Coaches	11 (100%)	10 (91%)
Teachers	29 (91%)	22 (96%)

Interviews. In spring 2025, SRI conducted 38 interviews with district leaders, school principals, assistant principals, coaches, and teachers across nine schools in the four Cohort 2 iCoachTEAM districts. SRI also interviewed three iCoachTEAM Program Team members who conducted the professional learning and mentoring. SRI used semi-structured protocols with questions aligned to the research questions and major constructs of the logic model, tailored by role. Interview topics included districts' and schools' preexisting math curriculum, professional development, and expectations for technology use; frequency and focus of coaching; and perceived impacts of the program on teacher practice and student outcomes.

SRI conducted all evaluation activities independently. These activities include the collection of administrative and program data, observations, surveys, and interviews; identification of the comparison sample; and all descriptive, impact, and implementation fidelity analyses.

Impact Study

The impact study featured a quasi-experimental design in which schools within iCoachTEAM districts were matched to other schools in the same state that were not implementing iCoachTEAM (comparison schools). The originally planned study design was to conduct school-level random assignments within school districts, and to use student-level data to estimate both the intent-to-treat effect and the average treatment effect among the treated, that is, among students whose teachers received coaching. However, it was not feasible to recruit enough districts under this design, in part because districts wanted to retain discretion over which coaches participated in iCoachTEAM and which teachers received coaching, rather than randomly assigning schools. We therefore instead allowed recruited districts to implement iCoachTEAM districtwide and shifted to a quasi-experimental design in which we selected comparison schools from within the same states as districts implementing iCoachTEAM and used publicly available school-level data to estimate impacts. As a result, it was not possible to document the comparison condition or to estimate the average treatment effect among the treated. The study compared student math achievement on state standardized tests in grades 6–8 in treatment and comparison schools.

Impact Study Sample

The impact study included the two Cohort 1 schools that piloted iCoachTEAM in 2022–23 and 2023–24 and all Cohort 2 schools that implemented the intervention in 2023–24 and 2024–25. Because the impact sample is broader than the implementation sample—by including both cohorts—findings from the implementation study do not fully capture implementation among the impact sample.

For the student impact analysis, the sample includes all students in study schools who were assessed on state standardized math assessments. In intervention schools, all such students in grades 6–8 were included, regardless of whether their teacher received coaching. The analyses therefore capture the intent-to-treat effect.

SRI selected comparison schools from other districts within the same states as intervention districts, using coarsened exact matching within state. SRI matched comparison schools to intervention schools on prior math proficiency rates.[2] Selection of comparison schools was based on similar math proficiency rates in 2021–22 for Cohort 1 and in 2022–23 for Cohort 2. The comparison condition therefore represents business-as-usual practices among schools matched to intervention schools. Because the comparison schools were drawn from districts not implementing iCoachTEAM, there is minimal risk of contamination. SRI was not able to collect data on the math professional learning or coaching that may have taken place in the comparison schools.

SRI conducted all analyses at the school level, with the analytic sample of students defined to include all those assessed on state standardized math assessments in grades 6–8. Table 4 lists the number of study districts, schools, and students.

Table 4. Number of Study Districts, Schools, and Students, by Cohort and Condition

Sample	Comparison	Treatment
Cohort 2		
Districts	566	4
Schools	928	12
Teachers, Year 1	.	64
Teachers, Year 2	.	68
Students, Year 1	358,972	5,857
Students, Year 2	355,417	5,889
Cohort 1		
Districts	132	2
Schools	312	4
Teachers, Year 1	.	16
Teachers, Year 2	.	.
Students, Year 1	84,692	1,662
Students, Year 2	82,894	1,646

Note. The same number of districts and schools were included in analyses for Year 1 and Year 2 outcomes. Because school-level administrative records do not inlcude counts of math teachers, the number of teachers in the comparison condition is not observed.

For the exploratory teacher analysis, SRI included teachers who received two years of coaching and were observed in fall 2023 (baseline) and spring 2025 (outcomes). The sample included 16 teachers across the four Cohort 2 districts (Table 5).

[2] Within state, it was not feasible to identify matches using other variables, such as student characteristics and the percent of students eligible for free and reduced price lunch. However, as shown in Table 8, across the analytic sample these characteristics were generally well balanced.

Table 5. Number of Teachers in the Observation Sample

Sample	Observed fall 2023	Observed spring 2025	Observed at both time points
Full sample	34	30	22
Coached teachers	26	23	16

Preregistration of the Study Design

The confirmatory research question, which addresses impacts on student achievement in math after two years of coaching, was preregistered with the Registry of Efficacy and Effectiveness Studies (REES). The matching plan included additional variables, such as the percentage of students eligible for free or reduced-price lunch, but it was not feasible to identify matches for each treatment school using the full set of preplanned covariates. The within-state matching therefore included grade levels served and percent proficient in math as the only matching variables. As reported later in Table 8, across states, the intervention and comparison schools are balanced on a broad range of measures.

Impact Study Design

SRI selected comparison schools from non–iCoachTEAM districts within the same states as intervention schools, using coarsened exact matching. Coarsened exact matching was implemented using the package “cem” for R, Version 1.1.31 (Iacus et al., 2009). All matching models included math proficiency in grades 6–8, and indicators for each of grades 6, 7, and 8 offered at the school. This ensured that, for example, a school serving only grades 7 and 8 could not be matched to a school serving grades 6–8. Cohort 1 schools, which began implementation in 2022–23, were matched on proficiency rates from 2021–22. Cohort 2 schools, which began implementation in 2023–24, were matched on proficiency rates from 2022–23.

SRI estimated student impacts (RQs 1–3) using school-level models of the following form:

$$y_{sdt} = \alpha + \tau T_{sd} + \gamma_1 * cohort\ 1 + \gamma_2 IL_{sd} + gamma_2 PA_{sd} + \beta X_{sd} + \rho_d + \epsilon_{sdt}$$

In this model, y_{sdt} is the outcome (either met/above proficiency, for RQs 1 and 2, or not proficient, for RQ 3) for school s in district d at time t; T_{sd} is an indicator equal to 1 if school s in district d implemented iCoachTEAM; $cohort\ 1$, IL_{sd}, and PA_{sd} are cohort and state indicators; and X_{sd} is a vector of school-level controls from the year before implementation, including percent proficient or above in math, percent eligible for free or reduced-price lunch, percent minority, school enrollment, and whether the school was rural, suburban, or a town (with urban as the omitted category). The parameter ρ_d is a district random effect to account for district-level selection into use of iCoachTEAM. To explore the relationship between the proportion of teachers receiving iCoachTEAM coaching and impacts on student proficiency (RQ 4), SRI added a term to the above model that was an interaction between T_{sd} and the proportion of math teachers in the school who received two coaching cycles.

To explore changes in teacher instruction (RQ 5), SRI conducted Stuart-Maxwell tests for each outcome measure from the MOT-CU and MOT-FA rubrics and the Framework for Teaching to test for homogeneity of marginal distributions to examine changes in teacher scores for coached teachers from baseline to outcome.

Measures

The study examined impacts on two school-level measures of student achievement and changes in 16 measures of teacher practice. SRI assessed impacts on student achievement in math using proficiency rates on state standardized math tests. For all schools, SRI combined proficiency rates across grades using weighted averages, in which the weights were the number of students in that grade, to form a measure of the share of students in grades 6–8 who were either proficient or above proficient in math. An additional measure looked at the percentage of students in the lowest proficiency level—those who were not proficient, compared with those who were nearly proficient, proficient, or above proficient.

SRI assessed teacher practice using multiple observation rubrics: the MOT-CU and MOT-FA rubrics created and validated by Digital Promise to gauge the use of technology to promote conceptual understanding and conduct formative assessment during math instruction (Hodkowski, 2026, under review), and the Framework for Teaching (Danielson, 2013; The Danielson Group, 2022).

With the Digital Promise rubrics, SRI examined the distribution of observation scores to test for change over time for each of four dimensions under the domains of conceptual understanding and formative assessment (Table 6). Each dimension was set on the 4-point scale of initial, emerging, developing, and extending levels of technology use. If no technology use was observed, the teacher received a score of "no technology."

Table 6. Domains and Dimensions of the Digital Promise Instrument

Domain	Dimension	Definition
Conceptual Understanding		Knowledge of the mathematical necessity of a particular mathematical relationship
	Depth of Mathematics	How teachers promote students' meaningful math understanding
	Productive Struggle	How students grapple with mathematics concepts
Formative Assessment		The ability to elicit and respond to students' mathematical thinking
	Eliciting Student Understanding	Ability to elicit student thinking (not that of the teachers); requires facilitating opportunities for students to share their reasoning so that conceptual progress toward the mathematical goal can be determined
	Responding to Student Understanding	Ability to respond to students' thinking and guide conceptual progress toward the mathematical goal; requires adjusting instruction based on the progress toward the goal that is inferred from the elicited reasoning

With the Danielson Framework for Teaching, SRI examined the distribution of observation scores to test for change over time for each of 14 elements (Table 7). These elements fell under the four components of the Classroom Environment domain: Communicating with Students, Using Questioning and Discussion Techniques, Engaging Students in Learning, and Using Assessment in Instruction. SRI scored each element on the Danielson Framework's 4-point scale (ranging from 1 = *unsatisfactory* to 4 = *distinguished*). See the Appendix for more detail on the Danielson Framework, observer training, and interrater reliability.

Table 7. Components and Elements of the Framework for Teaching's Classroom Environment Domain

Component	Element
Communicating with Students	Expectations for Learning
	Directions for Activities
	Explanations of Content
	Use of Academic Language
Using Questioning and Discussion Techniques	Quality of Questions/Prompts
	Discussion Techniques
	Student Participation
Engaging Students in Learning	Activities and Assignments
	Grouping of Students
	Instructional Materials and Resources
	Structure and Pacing
Using Assessment in Instruction	Clear Standards for Success
	Monitoring and Student Learning
	Feedback to Students

Note. Adapted from the *Danielson Framework for Teaching Evaluation Instrument* (2013) by Charlotte Danielson, with modifications incorporating langauge from the *Framework for Teaching* (2022) by the Danielson Group.

Data Analysis and Findings

This section describes the findings for both the confirmatory and exploratory student impact analyses and the exploratory teacher practice analysis.

Student Impact Analysis

Baseline Equivalence. Table 8 presents baseline means by condition and the effect size for the differences in means between conditions. Intervention and comparison schools were comparable with respect to prior proficiency in math, share of students eligible for free or reduced-price lunch, and school enrollment. For each of these, the effect size difference was less than 0.25, the threshold below which What Works Clearinghouse standards allow for regression adjustment to account for differences (What Works Clearinghouse, 2022). Intervention schools tended to have higher percentages of Hispanic students and fewer Black students, compared with comparison schools. Intervention schools were also in smaller school districts than comparison schools.

Table 8. Baseline Equivalence of Analysis Sample

Variable	Comparison mean (*SD*)	Treatment mean (*SD*)	Effect size difference
Percent proficient and above in math	27.9 (0.20)	26.3 (0.19)	−0.05
Percent eligible for free or reduced-price lunch	68.5 (0.22)	64.8 (0.23)	−0.10
Race or ethnicity			
Percent Asian	5.1 (0.05)	4.9 (0.05)	−0.02
Percent Black	10.9 (0.10)	8.1 (0.07)	−0.19
Percent Hispanic	42.5 (0.24)	53.8 (0.25)	0.27
Percent minority	53.4	61.9	0.21
Percent White	30.8 (0.21)	28.4 (0.20)	−0.07
Number of schools	1,240	16	

Note. Effect sizes for enrollment variables are calculated using Hedges' *g* and school-level standard deviations, and reported standard deviations for those variables are at the school level; because all other variables are binary at the student level, those effect sizes are calculating using the Cox Index.

Program Effects. Table 9 shows impact model estimates for the percentage of students proficient or above after one and two years of implementation (RQs 1 and 2) and for the percentage of students at the lowest proficiency level after one and two years of implementation (RQ 3). In both years, iCoachTEAM had a small, positive impact (less than 1 percentage point) on the percentage of students proficient or above in math; in both years, the impact is not statistically significant.[3] In Year 1, iCoachTEAM had a small, statistically insignificant, negative effect on the share of students in the lowest proficiency level.[4] In Year 2, Illinois stopped separately reporting the percentage of students not proficient in math and instead reported the

[3] Based on the realized standard errors, the study was powered to detect a change in proficiency of 5.3 percentage points or larger with 80% power.
[4] Across districts, there were two or three proficiency levels below proficient—for example, not met, partially met, and nearly met proficiency.

combined percentage of students not proficient and nearly proficient. Therefore, District A was excluded from the Year 2 analysis of impacts on the percentage of students at the lowest proficiency level. Among the other districts, the percentage of students at the lowest proficiency level in math increased by 0.82 percentage points, although this change was not significant.

Table 9. Year 1 and Year 2 Impact Estimates for Students Proficient or Above and Not Proficient

Variable	% proficient or above in math (*SE*)		% in lowest proficiency in math (*SE*)	
	Year 1	Year 2	Year 1	Year 2
(Intercept)	5.26 (1.73)	4.72 (1.65)	76.09 (1.8)	75.94 (2.38)
iCoachTEAM	0.88 (2.28)	0.09 (2.01)	−0.23 (1.9)	0.82 (2.58)
Cohort 1	6.10*** (1.46)	7.96*** (1.41)	−15.35*** (1.6)	−18.19*** (1.92)
Illinois	0.38 (1.1)	46.97*** (1.06)	7.87*** (1.17)	omitted
Pennsylvania [a]	−2.68* (1.15)	−4.16*** (1.09)	6.33*** (1.16)	7.68*** (1.44)
Percent proficient and above, baseline	1.00*** (0.04)	1.03*** (0.04)	−1.41*** (0.05)	−1.45*** (0.06)
Percent minority, baseline	−0.01 (0.01)	0.00 (0.01)	0.05*** (0.01)	0.04* (0.01)
Percent eligible for free or reduced-price lunch, baseline	−0.03** (0.01)	−0.03** (0.01)	0.07*** (0.01)	0.07*** (0.02)
Grade 6–8 enrollment, baseline	0.00 (0.00)	0.00 (0.00)	0.00*** (0.00)	0.00* (0.00)
Rural	0.61 (0.81)	0.46 (0.76)	−4.77*** (0.73)	−3.84*** (1.02)
Suburban	0.31 (0.69)	0.54 (0.64)	−3.04*** (0.59)	−3.02*** (0.85)
Town	0.40 (0.93)	−0.12 (0.86)	−3.51*** (0.85)	−2.99* (1.24)
Number of districts	701	701	701	486
Number of schools	1,256	1,256	1,256	912

[a] Only Pennsylvania had iCoachTEAM districts in Cohort 1. It also had one iCoachTEAM district in Cohort 2.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 10 shows results addressing RQ 4 on whether there is a relationship between the proportion of teachers in the school receiving iCoachTEAM coaching and the effect of iCoachTEAM on school-level math proficiency. To address this question, SRI created an interaction between the treatment indicator for being a school implementing iCoachTEAM and a continuous measure of the proportion of math teachers in that school who received two full cycles of iCoachTEAM coaching in year 1 (the recommended dosage).[5] As found in the implementation study (described next), capacity constraints and teachers' dispositions to receive coaching limited how many teachers could reasonably be coached. In five schools, none of the teachers received two full coaching cycles in year 1, and in two schools, all teachers received the full dosage in year 1. In the remaining nine schools, the percentage of teachers who received two coaching cycles in year 1 ranged from 17% to 57%.

SRI did not find a statistically significant association between the proportion of teachers receiving two full coaching cycles in year 1 and impacts on percentage of students proficient in math.[6] For the one-year impact model, the coefficient on iCoachTEAM indicates that schools in which no math teachers received two full coaching cycles saw a small but statistically insignificant gain of 0.8 percentage points in the percentage of students proficient or above in math, compared with schools that did not participate in iCoachTEAM. The coefficient on the interaction term between being an iCoachTEAM school and the proportion of teachers receiving the full dosage (iCoachTEAM*2 cycles) indicates that an increase from no teachers to all teachers receiving two full cycles was associated with a small and statistically insignificant gain of 0.2 percentage points in the percentage of students proficient in math. For the two-year impact model, the coefficient on iCoachTEAM indicates a statistically insignificant decline of −1.3 percentage points in the percentage of students proficient or above in math among schools with no teachers who received two full coaching cycles in year 1. Increasing the proportion of teachers receiving two full coaching cycles in year 1 from none to all was associated with a statistically insignificant increase in effectiveness of 3.6 percentage points.

Table 10. Year 1 and Year 2 Association Between Implementation and Impacts

Variable	% proficient or above in math (*SE*)	
	Year 1	Year 2
(Intercept)	5.26 (1.73)	4.73 (1.65)
iCoachTEAM	0.80 (3.00)	−1.29 (2.77)
iCoachTEAM*2 cycles	0.19 (5.07)	3.64 (5.02)
Cohort 1	6.10*** (1.46)	7.97*** (1.41)

[5] We use the year 1 percentage of teachers receiving two coaching cycles for both the year 1 and year 2 analyses because there was insufficient variation in the year 2 measure (with nearly all schools having 100% of teachers receiving two coaching cycles) and because as a lagging outcome, student achievement in year 2 may be more influenced by coaching received in year 1.

[6] Teachers receiving two full cycles is included as a proportion ranging from 0-1, rather than as a percentage, because a 1 percentage point increase in the share of teachers receiving two full cycles is not a meaningful change.

Variable	% proficient or above in math (*SE*)	
	Year 1	Year 2
District A	0.38 (1.10)	46.96*** (1.06)
District C – Cohort 2	−2.68* (1.15)	−4.16*** (1.09)
Percent proficient and above, baseline	1.00*** (0.04)	1.03*** (0.04)
Percent minority, baseline	−0.01 (0.01)	0.00 (0.01)
Percent eligible for free or reduced-price lunch, baseline	−0.03** (0.01)	−0.03** (0.01)
Grade 6–8 enrollment, baseline	0.00 (0.00)	0.00 (0.00)
Rural	0.61 (0.81)	0.47 (0.76)
Suburban	0.31 (0.69)	0.53 (0.64)
Town	0.40 (0.93)	−0.11 (0.86)
Number of districts	701	701
Number of schools	1,256	1,256

* $p < .05$. ** $p < .01$. *** $p < .001$.

Teacher Practice Analysis

To address exploratory RQ 5, SRI examined whether teacher practice improved after two years of iCoachTEAM coaching by comparing classroom observation scores from fall 2023 (baseline) and spring 2025 (outcomes). Distributions of scores for each element of the Digital Promise rubrics and the Framework for Teaching at baseline and outcome are reported in Tables 11 and 12.

Table 11. Distribution of Scores for the Digital Promise Rubric at Baseline and Outcome

Rubric element	Time point	No technology [a]	Initial	Emerging	Developing	Extending	Stuart-Maxwell test
Conceptual Understanding: Depth of Mathematics/ Cognitive Demand	Baseline	1	6	8	1	0	$\chi^2(4) = 1.4$, $p = .844$
	Outcome	0	5	8	3	0	
Conceptual Understanding: Student Productive Struggle	Baseline	1	6	9	0	0	$\chi^2(4) = 4.18$, $p = .383$
	Outcome	0	5	7	4	0	
Formative Assessment: Eliciting Students' Mathematical Reasoning	Baseline	1	11	3	1	0	$\chi^2(4) = 2.09$, $p = .719$
	Outcome	0	9	5	2	0	
Formative Assessment: Responding to Elicited Student Reasoning	Baseline	1	9	3	3	0	$\chi^2(4) = 4.80$, $p = .308$
	Outcome	0	6	5	4	1	

[a] If no technology use was observed during the lesson, the teacher received a “no technology” score.

Table 12. Distribution of Scores for the Framework for Teaching at Baseline and Outcome

Rubric element	Time point	Unsatisfactory	Basic	Proficient	Distinguished	Stuart-Maxwell test
Communicating with Students: Expectations for Learning	Baseline	0	8	8	0	$\chi^2(3) = 4.00$, $p = .262$
	Outcome	1	5	7	3	
Communicating with Students: Directions for Activities	Baseline	0	9	5	2	$\chi^2(3) = 1.29$, $p = .733$
	Outcome	0	6	8	2	
Communicating with Students: Explanations of Content	Baseline	1	5	9	1	$\chi^2(3) = 0.22$, $p = .974$
	Outcome	1	6	8	1	
Communicating with Students: Use of Academic Language	Baseline	0	2	12	2	$\chi^2(3) = 2.00$, $p = .572$
	Outcome	0	3	13	0	
Using Questioning and Discussion Techniques: Quality of Questions/Prompts	Baseline	4	8	4	0	$\chi^2(3) = 5.50$, $p = .139$
	Outcome	1	12	2	1	
Using Questioning and Discussion Techniques: Discussion Techniques	Baseline	0	6	8	2	$\chi^2(3) = 0.40$, $p = .940$
	Outcome	0	5	8	3	
Using Questioning and Discussion Techniques: Student Participation	Baseline	1	7	6	2	$\chi^2(3) = 0.33$, $p = .954$
	Outcome	1	8	5	2	
Engaging Students in Learning: Activities and Assignments	Baseline	0	2	12	2	$\chi^2(3) = 3.18$, $p = .365$
	Outcome	0	3	9	4	
Engaging Students in Learning: Grouping of Students	Baseline	0	1	8	7	$\chi^2(3) = 2.00$, $p = .572$
	Outcome	0	4	7	5	
Engaging Students in Learning: Instructional Materials and Resources	Baseline	1	7	6	2	$\chi^2(3) = 5.50$, $p = .139$
	Outcome	0	5	10	1	
Engaging Students in Learning: Structure and Pacing	Baseline	1	9	6	0	$\chi^2(3) = 1.15$, $p = .764$
	Outcome	1	7	7	1	
Using Assessment in Instruction: Clear Standards for Success	Baseline	0	6	10	0	$\chi^2(3) = 3.18$, $p = .365$
	Outcome	0	6	7	3	
Using Assessment in Instruction: Monitoring and Student Learning	Baseline	1	7	6	2	$\chi^2(3) = 1.35$, $p = .718$
	Outcome	0	7	8	1	
Using Assessment in Instruction: Feedback to Students	Baseline	1	7	8	0	$\chi^2(3) = 2.143$, $p = .543$
	Outcome	0	6	9	1	

On the Digital Promise rubric, across both Conceptual Understanding and Formative Assessment, average scores reflected some growth from baseline to outcome (Table 13). None of the differences was statistically significant.

Table 13. Descriptives from Digital Promise Rubric Observations for Teachers Coached for Two Years (n = 16)

Component	Element	Time	*M* Score	*SD* Score
Conceptual Understanding	Depth of Mathematics/Cognitive Demand	Baseline	1.56	0.73
		Outcome	1.88	0.72
	Student Productive Struggle	Baseline	1.50	0.63
		Outcome	1.94	0.77
Formative Assessment	Eliciting Students' Mathematical Reasoning	Baseline	1.25	0.68
		Outcome	1.56	0.73
	Responding to Elicited Student Reasoning	Baseline	1.50	0.89
		Outcome	2.00	0.97

There were modest, positive differences in teacher scores on the Framework for Teaching from baseline to outcome. Consistent with the goals of iCoachTEAM, teachers showed the most growth in using assessment in instruction (Table 14). None of the differences was statistically significant.

Table 14. Descriptives from Framework for Teaching Observations for Teachers Coached for Two Years (n = 16)

Component	Time	*M* Score	*SD* Score
Communicating with Students	Baseline	2.67	0.64
	Outcome	2.72	0.68
Engaging Students in Learning	Baseline	2.31	0.66
	Outcome	2.34	0.74
Using Assessment in Instruction	Baseline	2.54	0.65
	Outcome	2.71	0.65
Using Questioning and Discussion Techniques	Baseline	2.10	0.81
	Outcome	2.19	0.76

Implementation Study

The implementation study measured implementation fidelity in the Cohort 2 treatment schools and examined factors that facilitated or inhibited implementation. Implementation fidelity was informed by iCoachTEAM program data and coach and teacher surveys. Interview data provided context on the facilitators and barriers to implementation. This section describes SRI's approach first and then presents the implementation study findings.

Fidelity Measures

To measure implementation fidelity, SRI worked with the iCoachTEAM Program Team to specify indicators for the three key program components: school and district leader engagement, coach development, and provision of high-quality math coaching. For each indicator, SRI established the appropriate unit of measurement, defined the indicator scoring at the unit level (e.g., teacher), and defined the indicator scoring at the district level. For each component, SRI and the iCoachTeam Program Team established district-level adequacy thresholds for implementation. For the sample-level component score, SRI and the iCoachTeam Program Team set the threshold at 100% of all study districts meeting the district-level threshold for high fidelity (Table 15).

Table 15. Fidelity of Implementation Scoring System for Each Key Component in the Logic Model

Indicator	Unit of measurement	Indicator scoring at unit level	Indicator scoring at district level	Data source
Key Component 1. School and District Leader Engagement				
1.1: District leader participation in summer institute	District	NA	1 (High) = At least one district leader attends **summer** institute (one day) 0 (Low) = No district leader attends **summer** institute (one day)	iCoachTEAM attendance data
1.2: District leader participation in winter institute	District	NA	1 (High) = At least one district leader attends **winter** institute (one day) 0 (Low) = No district leader attends **winter** institute (one day)	iCoachTEAM attendance data
1.3: District leaders attend quarterly meetings with Digital Promise	District	NA	2 (High) = A district leader is present at **all** meetings 1 (Medium) A district leader is present at **3 of 4** meetings 0 (Low) = A district leader present at **2 or fewer** meetings	iCoachTEAM attendance data
District-level component score		*High = High on **all** 3 indicators* *Medium = High on 1.1 and 1.2 AND Medium on 1.3* *Low = Not High or Medium*		
Sample-level component score		*High = **100%** of districts are **high*** *Medium = Not High or Low* *Low = More than 33% of districts are low*		
Key Component 2: Coach Development				
2.1: Coaches participate in summer institute	Coach	1 (High) = Coach is present for **all** days 0 (Low) = Coach is **not** present for all days	2 (High): ≥ 75% of the coaches meet the threshold 1 (Medium): 50% to 74% of the coaches meet the threshold 0 (Low): < 50% of the coaches meet the threshold	iCoachTEAM attendance data

Indicator	Unit of measurement	Indicator scoring at unit level	Indicator scoring at district level	Data source
2.2: Coaches participate in winter institute	Coach	1 (High) = Coach is present for **all** days 0 (Low) = Coach is **not** present for all days	2 (High): ≥ 75% of the coaches meet the threshold 1 (Medium): 50% to 74% of the coaches meet the threshold 0 (Low): < 50% of the coaches meet the threshold	iCoachTEAM attendance data
2.3: Coaches receive monthly professional development (PD)	Coach	1 (High) = Coach receives PD at least **6 times** 0 (Low) = Coach receives PD **less** than 6 times	2 (High): ≥ 75% of the coaches meet the threshold 1 (Medium): 50% to 74% of the coaches meet the threshold 0 (Low): < 50% of the coaches meet the threshold	iCoachTEAM attendance data
2.4: Coaches receive monthly mentoring	Coach	**Year 1:** 1 (High) = Coach receives mentoring at **least 7 times** **Year 2:** 1 (High) = Coach attends 1:1 sessions with a mentor **3 times** per coaching cycle 0 (Low) = Coach receives mentoring **less** than 7 times or attends **fewer** than 3 1:1 sessions with a mentor per coaching cycle	2 (High): ≥ 75% of the coaches meet the threshold 1 (Medium): 50% to 74% of the coaches meet the threshold 0 (Low): < 50% of the coaches meet the threshold	iCoachTEAM attendance data
2.5: Coaches value the mentoring support they receive	Coach	1 (High) = coach reports support was "**quite valuable**" (4) or higher (5-point scale) 0 (Low) = coach reports support was "**somewhat valuable**" (3) or lower	2 (High): ≥ 75% of the coaches meet the threshold 1 (Medium): 50% to 74% of the coaches meet the threshold 0 (Low): < 50% of the coaches meet the threshold	Coach survey
District-level component score	*High = High on 2.1, 2.3, and 2.4 AND Medium or higher on 2.2 and 2.6* *Medium = Not High or Low* *Low = Low on 2.1, 2.3, AND 2.4*			
Sample-level component score	*High = At least 75% of districts are High and none are Low* *Medium = Not High or Low* *Low = More than 50% of districts are Low*			

Indicator	Unit of measurement	Indicator scoring at unit level	Indicator scoring at district level	Data source
Component 3: Provision of High-Quality Math Coaching				
3.0: Middle school math teachers in the school are coached	School	1 (High) = 70% or more of gen ed math teachers receive any coaching 0 (Low) = Not High	2 (High): ≥ 75% of schools meet the threshold 1 (Medium): 50% to 74% of the schools meet the threshold 0 (Low): < 50% of schools meet the threshold	Coaching logs
3.1: Teachers participate in coaching cycles	School	1 (High) = ≥ 70% of gen ed math teachers participated in ≥ **2** coaching cycles 0 (Low) = < 70% of gen ed math teachers participated in ≥ **2** coaching cycles	2 (High): ≥ 75% of schools meet the threshold 1 (Medium): 50% to 74% of the schools meet the threshold 0 (Low): < 50% of schools meet the threshold	Coaching logs
3.2: Coaches help teachers promote conceptual understanding or formative assessment in mathematics	School	1 (High) = ≥ 70% of gen ed math teachers select "**agree**" or "**strongly agree**" (on a 4-point Likert scale) 0 (Low) = < 70% of gen ed math teachers select "**agree**" or "**strongly agree**"	2 (High): ≥ 75% of schools meet the threshold 1 (Medium): 50% to 74% of the schools meet the threshold 0 (Low): < 50% of schools meet the threshold	Teacher survey
3.3: Coaching has an instructional focus	Coach	1 (High) = Scale average of coach responses is ≥ **3.5** ("moderate extent," across 5 survey items) 0 (Low) = Scale average of coach responses is < **3.5**	2 (High): ≥ 75% of coaches meet the threshold 1 (Medium): 50% to 74% of the coaches meet the threshold 0 (Low): < 50% of coaches meet the threshold	Coach survey
3.4: Coaching focuses supporting teachers to use technology	Coach	1 (High) = Scale average of coach responses is ≥ **3.5** ("moderate extent," across 5 survey items) 0 (Low) = Scale average of coach responses is < **3.5**	2 (High): ≥ 75% of coaches meet the threshold 1 (Medium): 50% to 74% of the coaches meet the threshold 0 (Low): < 50% of coaches meet the threshold	Coach survey
District-level component score	*High = High on 3.1, 3.2, and 3.4 AND Medium or higher on 3.0 and 3.3* *Medium = Not High or Low* *Low = Low on 3.1, 3.2, AND 3.4*			

Indicator	Unit of measurement	Indicator scoring at unit level	Indicator scoring at district level	Data source
Sample-level component score		*High = At least 75% districts are High and none Low* *Medium = Not High or Low* *Low = More than 50% of districts are Low*		

Fidelity Findings

SRI measured implementation fidelity across all participating teachers, coaches, and schools within the four Cohort 2 districts (Table 16). School-level results were rolled up to the district level. For indicators based on the coach and teacher surveys, SRI measured fidelity of implementation across the teachers and coaches who responded to the survey each year.

Table 16. Number of Teachers, Coaches, and Schools in Which Implementation Was Measured

District	School year	Number of middle school math teachers	Number of coached middle school math teachers	Number of coaches	Number of schools
District A	2023–24	11	7	2	2
	2024–25	9	6	2	2
District B	2023–24	14	9	3	3
	2024–25	19	5	3	3
District C	2023–24	7	5	3	1
	2024–25	6	6	3	1
District D	2023–24	32	9	3	6
	2024–25	34	7	3	6
Sample	2023–24	64	30	11	12
	2024–25	68	24	11	12

Table 17 shows the implementation level for each component within each district for each year of the implementation study. Overall, implementation was mixed across the two years of the initiative: While the School and District Leader Engagement component was consistently high, the other two components remained at mid-level fidelity in most districts. Interview data help contextualize some of the implementation successes and challenges across the components.

Table 17. Implementation Fidelity Summary

Year	Key component	Number of indicators	District A	District B	District C	District D	Full sample
2023–24	Key Component 1: School and District Leader Engagement	3	High	High	High	High	High
	Key Component 2: Coach Development	5	High	Medium	High	Medium	Medium
	Component 3: Provision of High-Quality Math Coaching	5	Medium	Medium	Medium	Medium	Medium
2024–25	Key Component 1: School and District Leader Engagement	3	High	High	High	High	High
	Key Component 2: Coach Development	5	Medium	Medium	Medium	Medium	Medium
	Component 3: Provision of High-Quality Math Coaching	5	Medium	Medium	High	Medium	Medium

District and school leader engagement was high in both years of implementation.

Strong alignment between districts' priorities and the initiative's goals facilitated strong district leader engagement. Across the districts, leaders consistently identified the need to improve students' math outcomes. One district leader explicitly noted that increasing conceptual understanding in math was a district priority, and another recognized it as an important part of math development. One district leader said the iCoachTEAM program coincided with the district's restructuring of its coaching positions and met its specific need to develop middle school math teachers, who are often multiple-subject teachers. "We knew that 6–8 was an area of need for our teaching staff," the district leader said, "especially because they are all multiple subject, and you know, when you don't have that math background coming to the table in middle school, that can be a rough transition."

District leaders' buy-in to the program was evidenced by their consistent attendance at the summer and winter institutes. In several cases, more than one leader from a district attended each of the sessions. Similarly, district and school leaders from all the districts attended the four quarterly meetings with the iCoachTEAM Program Team in both years.

Coaches participated in and valued mentoring more than other professional learning opportunities.

The mentorship provided to the coaches as part of iCoachTEAM was implemented with high fidelity: All but one coach across all the districts and years received the expected frequency of monthly mentoring. Additionally, all coaches agreed that the mentorship they received was "very valuable" and appreciated the mentor's responsiveness and flexibility, answering emails and calls at any time. Coaches described the mentor as helping them learn coaching practices, establish goals with teachers, discuss instructional strategies, and communicate effectively with teachers, especially those who were more resistant. One coach said the mentor "was a great sounding board around getting coached. With data, or with brainstorming around helping a teacher who was stuck and how to lean on a strategy."

By contrast, coach attendance at the other professional development opportunities was less consistent, especially if they were offered offsite. Although all coaches participated in the two-day winter institute in both years, several missed the multiday summer institute, particularly in the second year.

The indicator with the greatest area for improvement was attendance at the monthly professional development: Only two districts met fidelity on this indicator in Year 1, and only one district met fidelity in Year 2. Many coaches missed several sessions because of scheduling conflicts or lack of time, which hindered their ability to participate in the ongoing supports. This was particularly true for coaches who were also teachers or assistant principals and had competing responsibilities. One coach described the difficulty of attending meetings in the middle of the school day that require them to take a period off from teaching: "We get all the [release] time we want, but you can see how that can hurt with the continuity of your teaching."

While the number of middle school math teachers who were coached was limited, those who did receive coaching participated consistently in coaching cycles, especially in Year 2.

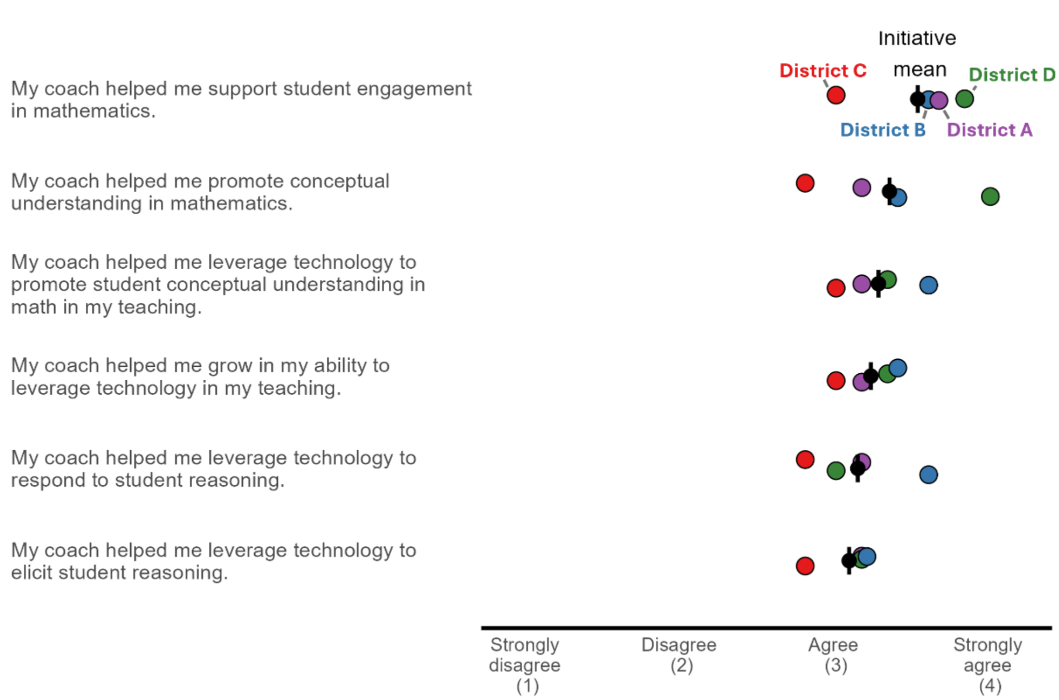
iCoachTEAM initially intended to provide coaching to all middle school math teachers. However, coach capacity constraints and teachers' dispositions to receive coaching limited how many teachers could be coached. As a result, districts fell short on the indicator that captures the extent to which iCoachTEAM's support was delivered across all eligible teachers. Only one of the four districts met fidelity in both years. In Year 1, 30 of the 64 middle school math teachers across all participating districts (44%) received some coaching; in Year 2, 24 of 68 (35%) received some coaching.

However, of the teachers who were coached, the proportion who received two coaching cycles increased from Year 1 (80%) to Year 2 (92%). Whereas six of the 30 coached teachers in Year 1 did not participate in two full coaching cycles and did not meet fidelity, only two of the 24 teachers in Year 2 missed the mark. This was a function, in part, of coaches becoming more comfortable with the coaching process and being more prepared. For example, one coach described how much time they spent before each coaching session in Year 1 and how they became more comfortable letting teachers drive the session in Year 2:

> ***I can compare my year one to where ... I was staying so late that I needed to have this lesson and creating all this stuff for [the teachers] without them there. It was my lesson, my agenda, because I thought I had to have all these answers versus now ... they are coming up with their challenge.***

Additionally, teachers saw the value of the coaching and were likely more receptive to engaging in coaching cycles in Year 2. Almost all teachers in Year 1 and Year 2 “agreed” or “strongly agreed” that coaches helped them promote conceptual understanding or formative assessment in math (Figure 2).

Figure 2. Teacher Reports on the Usefulness of Coaching, by District, 2024–25

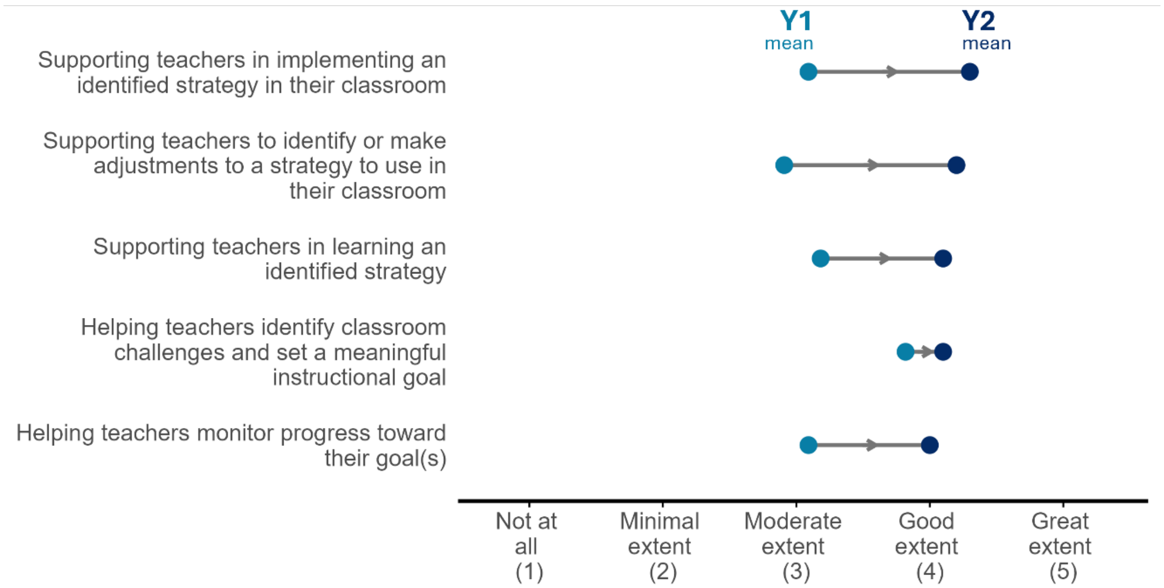


Note. Data are from the teacher survey, Question 19 (Items b, c, d, e, f, and g). Respondent $n = 22$.

Coaching was more focused on instruction and technology year over year.

Coaches’ perceptions of how much the coaching they provided focused on instruction increased across districts from Year 1 to Year 2 (Figure 3). In Year 1, five of the 11 coach respondents (45%) felt that coaching focused on instructional practice to a moderate extent; by Year 2, nine of the 10 coach respondents (90%) felt this way. Whereas only two districts met fidelity on this indicator in Year 1, three met fidelity by Year 2. By Year 2, coaches, none of whom were math content specialists at the outset, were more comfortable supporting teachers with math content.

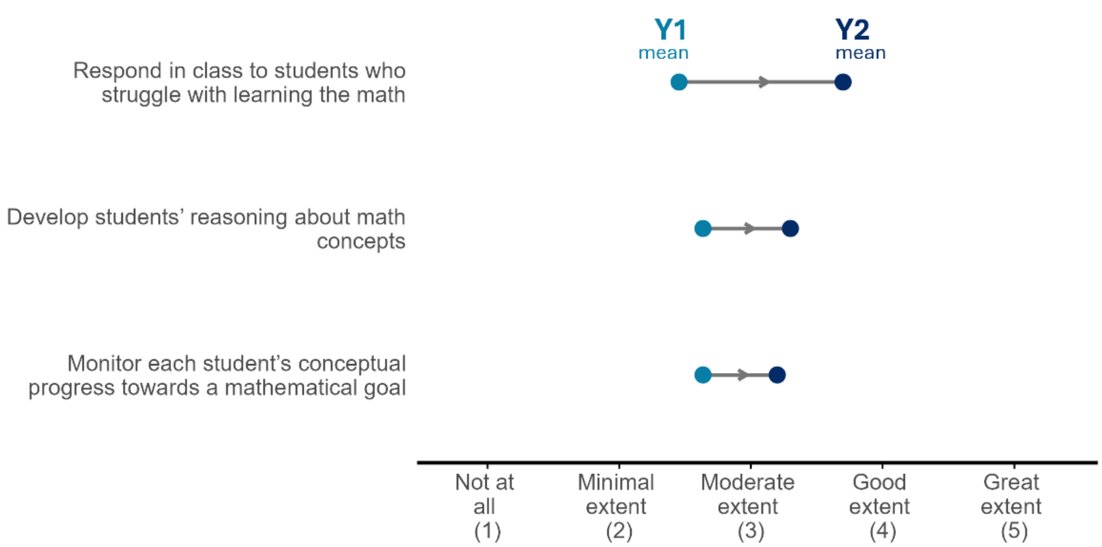
Figure 3. Coach Reports on the Focus of Coaching, 2023–24 and 2024–25



Note. Data are from the coach survey, Question 17 (Items a, b, c, d, and e). Respondent *n* = 11 in Year 1, 10 in Year 2.

Coaches also reported an increase from Year 1 to Year 2 in the extent to which they supported teachers to use technology in their instruction (Figure 4). Consistent with the goals of iCoachTEAM, coaches reported a greater focus on helping teachers use technology to develop students' reasoning and monitor their conceptual progress toward a math goal in Year 2 than they did in Year 1.

Figure 4. Coach Reports of Supporting Teachers to Use Technology for Various Purposes, 2023–24 and 2024–25



Note. Data are from the coach survey, Question 18 (Items a, b, and c). Respondent *n* = 11 in Year 1, 10 in Year 2.

Interviews indicated that few teachers integrated technology in new or different ways that would substantially support conceptual understanding.

Despite an increased focus on technology integration during coaching in Year 2, teachers and coaches lacked time to learn how to use existing digital tools to promote conceptual understanding or conduct formative assessment. As a result, teachers continued to commonly use technology for procedural tasks: to present content or to give students opportunities for independent practice. Teachers across districts reported using tools like smartboards and document cameras to walk students through the steps for solving problems. In one district, teachers used online resources (e.g., the graphic design tool, Canva) to post digital versions of textbooks, study guides, and answer keys. In some districts, teachers used programs like i-Ready or ALEKS with mixed results: Some teachers used the tools for procedural practice, and others used them for formative assessment. "ALEKS, we're using for checking for understanding," said one coach. "What I personally love about that is this instantaneous feedback where kids know if they got it right or wrong. I can visually see, like, which kids are getting things right or wrong and go over and give immediate feedback if need to."

In some districts, coaches struggled to get teachers to view technology as a tool that could be integrated with instruction in meaningful ways. To some interviewed teachers, technology was separate from instruction, a tool to use *after* students understand the content. In two districts, teachers believed that math was more effectively taught with paper and pencil because it allowed students to internalize and show their work. One teacher explained her reasoning:

> ***I can't say I use tech in my classroom. They have to do it on paper and pencil. One of the things I struggle with is that they click, they are not focused on showing the work. I want to know your process and your mistakes you're making, what you're writing, processing to arrive at your answer. … I just know it's the correct answer … is what they say [when asked why they picked it]. I do think that you retain the information better when you're writing it down and working out the problems.***

Coached teachers in another district reported having students complete note-taking and practice problems using paper and pencil to ensure the students did the thinking required by the task.

Despite medium to high implementation fidelity across the districts, teacher practice did not change substantially to integrate technology in ways that promote conceptual understanding or support formative assessment.

Conclusion

iCoachTEAM was an ambitious endeavor to build the capacity of coaches to help teachers use technology in fundamentally different ways to improve math teaching, student engagement, and ultimately, student achievement. Achieving this vision required coaches to strengthen not only their coaching skills but also their math content knowledge and understanding of how technology can support conceptual understanding and formative assessments. Teachers also needed to rethink how technology could be used—not simply to present information or support procedural practice, but to deepen students' math thinking and engagement.

Findings from the Cohort 2 implementation study suggest that this transformation had begun but was still in its early stages. Alignment with district math priorities helped generate district support, and coaches reported that their work became increasingly focused on instruction and technology integration over time. Teachers likewise viewed coaching as valuable for expanding their use of technology. However, technology was generally not used in ways that substantially changed math instruction, and many teachers had not yet shifted from procedural to more conceptually focused uses of technology.

Implementation challenges also limited the program's reach. In three of the districts, coaching capacity meant that not all math teachers received coaching. Where coaches also served as classroom teachers, competing responsibilities reduced their ability to participate fully in professional learning and provide consistent coaching.

Consistent with these implementation findings, the impact study found no statistically significant effects on student achievement after two years of iCoachTEAM. There was also no statistically significant relationship between receiving two full coaching cycles and student achievement. Classroom observations similarly did not show statistically significant improvements in instructional practice, although teachers demonstrated modest growth across all observation rubric dimensions, suggesting that practice was moving in the intended direction.

These findings should be interpreted considering several study limitations. Because participating districts preferred to determine which coaches and teachers participated in the program rather than randomly assigning schools, the study shifted from an RCT to a quasi-experimental design. The comparison group was constructed using matched schools in other districts in the same states as intervention districts, and outcomes were measured using publicly available school-level achievement data. Consequently, the intent-to-treat analysis included all tested students in participating schools, even though only a small proportion of teachers received coaching and fewer still received the full dosage of two coaching cycles. Thus, there was reduced contrast between the treatment and comparison schools, and the intent-to-treat estimate likely understates the average treatment effect. In addition, school-level proficiency measures could not capture whether the program benefited students who entered well below grade level—e.g., if their scores improved but did not cross from not proficient to nearly proficient—or explore the contexts, teachers, or students for whom the intervention was most effective.

The iCoachTEAM study findings suggest that improving student engagement and achievement in math through the strategic use of technology requires quite a few shifts in coaching and teaching first. In the treatment districts, coaches first needed to develop expertise in coaching, math content and instruction, and the strategic use of technology. Teachers, too, needed to build their own understanding of conceptually focused math instruction, shift their beliefs about how technology could support learning, and translate those ideas into sustained changes in classroom practice. These foundational shifts were beginning to happen during the study but had not yet reached the scale or depth needed to produce measurable improvements in student achievement. For initiatives that seek to transform instructional practice through coaching and technology integration, meaningful impacts on student learning may require broader implementation, sustained coaching, and a longer time horizon than the two years examined in this study.

References

Campbell, P. F., & Griffin, M. J. (2017). Reflections on the promise and complexity of mathematics coaching. *The Journal of Mathematical Behavior*, *46*, 163–176. https://doi.org/10.1016/j.jmathb.2016.12.007

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88–113. https://doi.org/10.1016/j.edurev.2013.01.001

Danielson, C. (2013). *The Framework for Teaching: Evaluation instrument* (2013 ed.). The Danielson Group.

The Danielson Group. (2022). *The framework for teaching*. https://danielsongroup.org/the-framework-for-teaching/

Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?. *International Journal of Public Health, 58(4)*, 637-642. https://doi.org/10.1007/s00038-012-0416-3

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, *28*(5), 524–549. https://doi.org/10.2307/749690

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, *41*(5), 513–545. https://doi.org/10.5951/jresematheduc.41.5.0513

Hodkowski, N. M., Hardy, A., McMahon, L., & Mielicki, M. (Under Review). A Tool for Observations of Mathematics Instruction that Integrates Technology to Promote Conceptual Understanding. *Journal for Research in Mathematics Education*.

Iacus, S., King, G., & Porro, G. (2009). cem: Software for coarsened exact matching. *Journal of Statistical Software*, *30*(9), 1–27. https://doi.org/10.18637/jss.v030.i09

Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research*, *43*(4), 489–510. https://doi.org/10.2190/EC.43.4.d

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teaching coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547–588. https://doi.org/10.3102/0034654318759268

Kuhfield, M., Soland, J., Lewis, K., & Morton, E. (2022). *The pandemic has had devastating impacts on learning. What will it take to help students catch up?* Brookings. https://www.brookings.edu/articles/the-pandemic-has-had-devastating-impacts-on-learning-what-will-it-take-to-help-students-catch-up/

National Center for Education Statistics. (n.d.). *NAEP long-term trend assessment results: Reading and mathematics*. https://www.nationsreportcard.gov/highlights/ltt/2022/

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE).

Appendix

Framework for Teaching

Table A1. Components and Elements of the Framework for Teaching

Component	Element	Description
Communicating with Students	Expectations for Learning	Teacher clearly communicates goals for learning to students and situates learning goals in broader context.
	Directions for Activities	Teacher communicates lesson expectations and addresses possible challenges so students understand what they are expected to do during a lesson.
	Explanations of Content	Teacher provides clear and engaging explanations, with appropriate scaffolding, anticipates possible student misconceptions, and encourages students' intellectual engagement.
	Use of Academic Language	Teacher uses precise, academic vocabulary and explain the use of this vocabulary to students.
Using Questioning and Discussion Techniques	Quality of Questions/Prompts	Teachers ask only a few high-quality questions and provide students with sufficient time to think about their responses, to reflect on the comments of their classmates, and to deepen their understanding.
	Discussion Techniques	Teachers promote learning through discussion, challenging all students to explain and justify their reasoning and conclusions, enabling students to engage in discussion directly with one another, and building on student responses.
	Student Participation	Teacher uses a range of techniques to encourage all students to contribute to the discussion and enlists the assistance of students to ensure this outcome.
Engaging Students in Learning	Activities and Assignments	Teacher implements activities and assignments that require student thinking that emphasizes depth over breadth and encourage students to explain their thinking.
	Grouping of Students	Teachers group students in ways that provide opportunities for all students being intellectually engaged in the lesson.
	Instructional Materials and Resources	Teachers use instructional materials and resources that provide opportunities for all students being intellectually engaged in the lesson.

Component	Element	Description
	Structure and Pacing	Teachers structure and pace lessons such that all students are intellectually engaged throughout the lesson and provides opportunities for students to reflect on their learning.
Using Assessment in Instruction	Clear Standards for Success	Teachers ensure that students understand the characteristics of high-quality work, or students help establish the evaluation criteria.
	Monitoring of Student Learning	Teacher constantly monitors student understanding, makes use of strategies to elicit information about individual student understanding, and provides opportunities for students to monitor their own understanding.
	Feedback to Students	Teacher provides specific feedback to students focused on improvement and creates opportunities for students to provide feedback to each other.

Note. Adapted from the *Danielson Framework for Teaching Evaluation Instrument* (2013) by Charlotte Danielson, with modifications incorporating langauge from the *Framework for Teaching* (2022) by the Danielson Group.

Observer Training

Four researchers not affiliated with any school district served as observers and received extensive training on the observation tool for conceptual understanding to ensure interrater reliability. As part of the training, observers watched and scored classroom observation videos and met to discuss scores and resolve any discrepancies. The observers watched five videos, scored 21 scenarios, and met four times before baseline observations, and they scored two videos and met two times before outcome observations.

The same four observers were also trained to use the Framework for Teaching, using a similar approach as described above. The observers scored eight videos and met four times to discuss and calibrate scores before baseline observations, and they scored 10 videos and met five times before outcome observations.

Reliability and Validity of Observations Instruments

To establish interrater reliability among the four observers, pairs of observers scored six out of 34 (17.65%) observations at baseline and 13 out of 30 (43.33%) observations at outcome, before independently scoring the rest. Each pair of observers met shortly after co-conducting the observations to discuss scoring and reach consensus on the final scores. Measures of interrater reliability were computed on the 19 double-coded observations, prior to resolving disagreements. Table A2 shows percent agreement (both exact agreement and scores within one point of each other), Cohen's kappa (quadratically weighted), and Pearson correlation values for both the Digital Promise rubric and Framework for Teaching.

Table A2. Measures of Interrater Reliability for the Digital Promise Rubric and the Framework for Teaching

Domain/dimension or component/element	Percent agreement (exact)	Percent agreement (within one)	Cohen's kappa [95% CI]	Pearson correlation [95% CI]
Digital Promise rubric				
Conceptual Understanding: Depth of Mathematics/ Cognitive Demand	84	100	.83 [.65, 1.00]	.83*** [.61, .93]
Conceptual Understanding: Student Productive Struggle	74	95	.65 [.40, .89]	.70*** [.36, .87]
Formative Assessment: Eliciting Students' Mathematical Reasoning	74	100	.76 [.56, .97]	.77*** [.48, .91]
Formative Assessment: Responding to Elicited Student Reasoning	53	95	.51 [.13, .88]	.54* [.11, .80]
Framework for Teaching				
Communicating with Students: Expectations for Learning	53	89	.54 [.24, .84]	.55* [.13, .80]
Communicating with Students: Directions for Activities	58	89	.22 [−.05, .49]	.27 [−.21, .65]
Communicating with Students: Explanations of Content	37	89	.22 [−.07, .52]	.29 [−.19, .66]
Communicating with Students: Use of Academic Language	74	100	−.09 [−.24, .06]	−.12 [−.55, .35]
Using Questioning and Discussion Techniques: Quality of Questions/Prompts	63	95	.19 [−.31, .69]	.20 [−.28, .60]
Using Questioning and Discussion Techniques: Discussion Techniques	84	100	.84 [.67, 1.00]	.84*** [.63, .94]

Domain/dimension or component/element	Percent agreement (exact)	Percent agreement (within one)	Cohen's kappa [95% CI]	Pearson correlation [95% CI]
Using Questioning and Discussion Techniques: Student Participation	47	95	.18 [−.41, .76]	.19 [−.29, .59]
Engaging Students in Learning: Activities and Assignments	63	95	.35 [−.09, .80]	.38 [−.09, .71]
Engaging Students in Learning: Grouping of Students	47	95	.50 [.21, .80]	.57* [.15, .81]
Engaging Students in Learning: Instructional Materials and Resources	53	95	.44 [.05, .83]	.45 [−.01, .75]
Engaging Students in Learning: Structure and Pacing	68	100	.61 [.40, .82]	.67** [.31, .86]
Using Assessment in Instruction: Clear Standards for Success	53	100	.27 [−.11, .66]	.30 [−.18, .66]
Using Assessment in Instruction: Monitoring of Student Learning	42	89	.13 [−.25, .51]	.15 [−.33, .57]
Using Assessment in Instruction: Feedback to Students	37	100	.39 [.03, .75]	.41 [−.05, .73]

*** $p < .001$. ** $p < .01$. * $p < .05$.

To determine internal consistency of the Digital Promise rubric, SRI calculated nonstandardized Cronbach's alpha separately for each domain at each time point. Because the rubric had only two themes and Cronbach's alpha may be less reliable for measures with two items (Eisinga et al., 2013), SRI also calculated Pearson correlations separately at each time point. Finally, to increase the power for each measure of internal consistency, SRI calculated each measure for the observations combined across time points ($N = 64$). All internal consistency measures are in Table A3. The Measures of Internal Consistency for Conceptual Understanding demonstrates strong internal consistency, suggesting that it measures a cohesive underlying construct—instructional practices that support students' conceptual understanding through the use of technology. To achieve a high score on either theme, students must be given opportunities to independently interact with technology while engaging in classroom activities that support exploration and sense-making of mathematical principles and relationships.

Table A3. Measures of Internal Consistency for Conceptual Understanding and Formative Assessment Tools at Baseline, Outcome, and Combined

Domain	Time point	Cronbach's α [95% CI]	Pearson's *r* [95% CI]
Conceptual Understanding	Baseline	.94 [.86, .99]	.90 [.81, .95]
	Outcome	.95 [.86, 1.00]	.91 [.82, .96]
	Combined	.95 [.90, .99]	.91 [.85, .94]
Formative Assessment	Baseline	.91 [.83, .96]	.85 [.72, .92]
	Outcome	.74 [.44, .87]	.59 [.29, .78]
	Combined	.84 [.71, .91]	.73 [.59, .83]

Silicon Valley
(SRI headquarters)
333 Ravenswood Avenue
Menlo Park, CA 94025
1.650.859.2000

Washington, DC
1100 Wilson Boulevard
Suite 2700
Arlington, VA 22209
1.703.524.2053

education@sri.com

www.sri.com/education-learning/