

EVALUATION REPORT

Intensive Assistance Model: Implementation and Early Outcomes

Benjamin W. Cottingham

California Education Partners

Jeannie Myung

Stanford University

Marshall Garland

University of Southern California

Abstract: The Intensive Assistance Model (IAM) was a California Collaborative for Educational Excellence (CCEE) support for districts identified for Direct Technical Assistance, designed to strengthen coherent instructional systems through the PLC at Work approach in partnership with Solution Tree. From 2022–23 to 2024–25, IAM engaged eight pilot schools across five districts in structured, evidence-driven collaboration oriented toward student learning, adult collaboration, and results. This Year 3 mixed-methods evaluation drew on 96 semi-structured interviews, Solution Tree’s PLC continuum ratings, and school-level CAASPP and attendance trends. Qualitative analyses found evidence of technical shifts in educators’ mastery of essential standards, development of common assessments, and shared routines for examining student work, alongside adaptive shifts in beliefs about students’ capacity to learn, adult collaborative culture, and mindsets about data use. These shifts were associated with more targeted student interventions and instructional practice, and improvements in students’ learning experiences. However, weighted difference-in-differences and event-study analyses did not detect statistically significant effects of IAM on achievement or attendance, an unsurprising pattern given the pilot’s short duration and the complexity of the improvement contexts in the qualifying districts. Policy implications emphasize strengthening coaching and leadership infrastructure, incorporating process-oriented indicators alongside outcomes, protecting time for collaboration and intervention, and investing in district capacity to coordinate and sustain implementation.

February 2026

Executive Summary

The Intensive Assistance Model (IAM) was part of the California Collaborative for Educational Excellence's (CCEE) broader portfolio of district supports and was aimed at districts where performance concerns persisted over multiple years, across student groups, and across LCFF priority areas—indicating a need for deeper capacity-building support. It was designed to accelerate improvement in districts identified for Direct Technical Assistance (DTA) by building coherent instructional systems that focused on student learning, collaboration, and results. IAM leveraged the PLC at Work model (professional learning community), implemented in partnership with Solution Tree. It engaged eight pilot schools across five districts in a structured, school-level improvement model that emphasized teacher leadership and disciplined, evidence-driven collaboration focused on student learning. The IAM strategy organized adult work around four essential questions: (1) What did we want students to learn? (2) How did we know if they learned it? (3) What did we do if they didn't learn it? (4) What did we do if they already knew it? The intended outcomes for schools in IAM included improving student outcomes, achieving Model PLC at Work recognition, and scaling and sustaining the work from one site to others across the district.

IAM was piloted from 2022–23 through 2024–25 across five districts and eight schools (seven elementary schools and one middle school). This Year 3 report examined how IAM routines matured over time, what district conditions enabled or constrained implementation and scale, and evidence of impacts on teaching and learning. This report was the third in a three-year evaluation series documenting how IAM unfolded—from initial diagnosis and foundation-building to deepening instructional improvement work and efforts to scale. This Year 3 evaluation focused on how improvement capacities and culture evolved over time and how IAM supports translated into stronger instructional coherence and improved outcomes.

This evaluation used a mixed-methods, implementation-focused design. Qualitative data collected over the three-year period included 96 semi-structured interviews with teachers, school leaders, district leaders, Solution Tree coaches, and CCEE leaders; transcripts were analyzed using a grounded theory–informed approach. The study also drew on Solution Tree's PLC continuum ratings, collected three times per year across 14 indicators of implementation quality. A core qualitative analytic lens distinguished between technical changes—improvements in observable skills and processes—and adaptive changes, which required shifts in beliefs, culture, and mindsets. These shifts were expected to strengthen instructional practice and adult collaborative learning conditions that, over time, laid the groundwork for improvements in teaching and student learning. The quantitative analysis estimated IAM effects using weighted difference-in-differences and event-study models,

comparing IAM schools to non-IAM schools within pilot districts. Quantitative analyses used school-level panel data on student CAASPP assessment results and attendance from 2016–17 to 2024–25, publicly available from the California Department of Education.

The qualitative analysis showed evidence of technical shifts in educators’ mastery of essential standards, their ability to develop common assessments, and the establishment of shared systems for reviewing student learning. In the domain of adaptive change, teachers demonstrated stronger commitment to ensuring all students learned, growth in collaborative adult culture, and a shift toward using data as an inquiry tool for improvement. Together, these technical and adaptive shifts contributed to proximal outcomes evident in instructional adjustments aligned to students’ demonstrated learning and in more targeted interventions, often implemented through cross-classroom student grouping.

The quantitative analysis estimated IAM’s incremental effect on publicly available CAASPP results and attendance outcomes, relative to comparison schools in the same districts. Given the pilot’s relatively short time horizon and its focus on districts facing significant, urgent improvement needs, marked short-term changes in distal outcomes such as standardized test performance were not impossible, but they were not expected. Findings from the quantitative impact study, however, provide encouraging evidence of improvements by the third year of implementation. Descriptive analysis examining absolute performance trajectories before and after implementation echo these promising findings. Weighted event-study models comparing IAM schools to non-IAM schools within pilot districts found positive, medium-sized (Kraft, 2020) impacts on state mathematics assessments in the third year of implementation. Combined effects across all three post-implementation were small, but positive in mathematics, and small and negative in ELA. Impacts on student attendance (measured by school attendance rates and chronic absenteeism) were small and close to zero. These patterns in the empirical analysis of distal outcomes were not unexpected given the timeframe and the complexity of these schools’ improvement contexts.

It is important to note that small or imprecise estimates obtained from the impact models do not imply that IAM schools failed to make progress. Rather, these estimates reflect the degree to which IAM schools improved relative to a carefully constructed comparison group. The positive trajectories visible in school-level data and in the later event-study periods are consistent with a program maturing, even if the evidence of the causal effect remains limited at this early stage.

Several policy implications emerged from the evaluation. First, the findings pointed to the value of a stronger state and regional infrastructure for instructional coaching and leadership development so that school-level autonomy was accompanied by the technical and adaptive

capacity needed for high-quality implementation. Second, the results suggested that California's accountability system could have been strengthened by incorporating process-oriented indicators, particularly measures of adult collaboration, alongside a more focused set of outcome measures to better assess readiness for instructional change and better target supports. Third, the evaluation highlighted the importance of structural conditions, especially master-schedule design, in creating protected time for teacher collaboration and timely student intervention. Fourth, the patterns observed underscored the district's central role in coordinating coherent supports, routines, and monitoring processes that could have enabled more consistent implementation across schools.

Introduction

Within the California Collaborative for Educational Excellence's (CCEE's) broader portfolio of district supports, the Intensive Assistance Model (IAM) was designed to help eligible districts accelerate improvement by building coherent, instructionally focused systems for diagnosing challenges, strengthening practice, and sustaining change over time. IAM responded to this level of need with a structured, school-level improvement approach that emphasized teacher leadership and disciplined collaboration around evidence of student learning.

This report was the third in a series of three annual evaluation reports examining IAM implementation and what it took to apply the model in practice to improve schools. Across the series, we traced how IAM unfolded over time—from initial diagnosis and relationship-building, to deepening instructional improvement work, to strengthening the conditions for sustainability and scale. This Year Three evaluation focused on the evolution of improvement routines and structures and the extent to which IAM's supports translated into stronger instructional coherence and improved outcomes. Together, the three reports aimed to inform CCEE, participating districts, and the broader field about the design features, enabling conditions, and practical challenges involved in using IAM's approach to advance equitable, lasting improvement.

Background

California's Local Control Funding Formula (LCFF) established a statewide accountability and continuous improvement system intended to ensure that districts had the resources and support needed to meet the needs of all students—especially those facing the greatest barriers to opportunity. As part of this legislation, the California Collaborative for Educational Excellence (CCEE) was created to provide direct technical assistance to local educational agencies that

demonstrated a sustained need for support, including districts that repeatedly failed to improve performance on the California School Dashboard. Based on California School Dashboard results, districts met the criteria for Direct Technical Assistance (DTA) when performance concerns persisted over multiple years across multiple student groups and more than one LCFF priority area. This pattern signaled a need for more intensive, sustained capacity-building support beyond differentiated assistance.

The first cohort of districts selected for Direct Technical Assistance (DTA) was identified for the 2019-20 school year. At that point, there were no guidelines for how these districts would be supported differently under DTA from the differentiated assistance (DA) they had already received. CCEE leadership recognized that the existing system often incentivized performative accountability for the legislature—check-the-box exercises in which districts appeared stable on paper but remained stagnant in teaching and learning practices. As one CCEE leader noted, “A district can run circles around wrapping up accountability in a nice package and saying, ‘Here, we did it,’ and really nothing changed.” The IAM pilot was designed to bypass this performative layer and focus instead on the daily interactions among teachers and students, where meaningful improvements in teaching and learning were made.

The PLC at Work Model

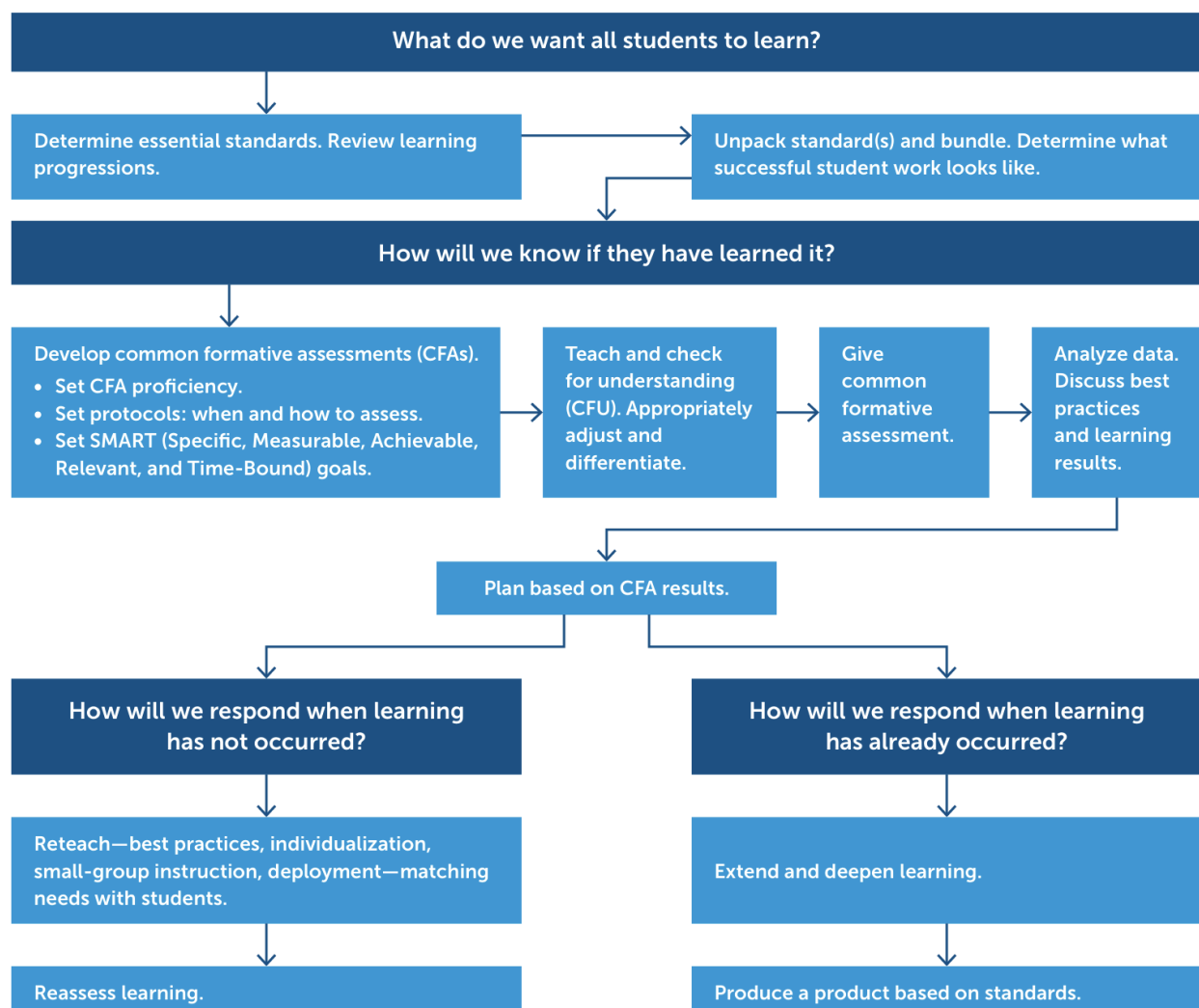
CCEE’s IAM pilot aimed to strengthen teaching and learning in the pilot schools by leveraging a model with a proven track record of school improvement in California: the PLC at Work model. CCEE implemented the IAM pilot in partnership with Solution Tree, an organization with an established record of supporting high-quality PLC at Work implementation. The PLC at Work strategy deliberately simplified adult work by centering it on four essential questions (DuFour et al., 2024):

1. What do we want students to learn?
2. How will we know if they have learned it?
3. How will we respond when learning has not occurred?
4. How will we respond when learning has already occurred?

In the PLC at Work model, teachers collaborate within and across grade-level teams to develop common, standards-based unit plans and assessments, analyze evidence of student learning, and identify student learning assets and needs. Each team then determines how best to intervene with students and assumes collective responsibility for all students—not only those in individual classrooms. Staff at each pilot school received intensive coaching from a team of five Solution Tree coaches—a lead coach, a math coach, a literacy coach, an assessment coach, and an intervention coach—to develop the structures and processes needed to address the four

essential questions. Coaches support schools through transformation processes that enable continuous improvement in teaching and learning through the model described above. This work was led by two critical groupings of school staff: (a) the guiding coalition, which included site leadership and representatives from each grade level, and (b) grade-level teams, which met regularly to identify key standards, develop unit plans and common formative assessments, examine student outcomes, share instructional strategies, and identify students in need of intervention. The teaching and learning process is elaborated below.

Figure 1: PLC at Work Model Road Map



Implementation of the model generally follows the three tiers of questioning shown in Figure 1 above. In Year 1, schools focused on answering Question 1—“What did we want all students to

learn?”—by establishing essential standards and developing a shared understanding of what student learning looked like for each standard. This first year was supported by the lead coach and the math and ELA coaches. By the end of Year 1 and into Year 2, teams began to answer Question 2—“How did we know if [students] had learned it?”—by developing common formative assessments (CFAs) aligned to the selected essential standards and practicing a shared process for analyzing student work to refine instruction and intervene with students. During this year, school teams participated in the Assessment Academy with their assessment coach to develop rigorous, aligned CFAs; this work also included, in some cases, developing unit plans around the essential standards. Finally, in Year 3, teams developed and implemented intervention systems—supported by the intervention coach—to answer Questions 3 (“How did we respond when learning had not occurred?”) and 4 (“How did we respond when learning had already occurred?”). By the end of the third year, schools should have experienced full implementation of the model and addressed all four questions with support from their Solution Tree coaches. However, the model focused on teaching and learning at the school level. The pilot also aimed to influence one level higher by supporting districts in scaling learning from pilot schools across the district.

The Pilot

Based on California School Dashboard results, districts met the criteria for DTA when performance concerns persisted over multiple years across multiple student groups and more than one LCFF priority area. The first cohort of districts selected for DTA was identified for the 2019-20 school year and initial guidelines for how to support districts in DTA were unclear. CCEE leadership recognized that the existing system often incentivized performative accountability for the legislature—check-the-box exercises in which districts appeared stable on paper but remained stagnant in teaching and learning practices. As one CCEE leader noted, “A district can run circles around wrapping up accountability in a nice package and saying, ‘Here, we did it,’ and really nothing changed.” The IAM pilot was designed to bypass this performative layer and focus instead on the daily interactions among teachers and students, where meaningful improvements in teaching and learning were made.

The same year that DTA began, the Learning Policy Institute published a report and a series of case studies detailing the aspects of positive outlier districts across California, or those districts that outperformed student achievement expectations. The primary finding from this report was that “After controlling for the socioeconomic status of students’ families and district characteristics...teacher qualifications are the most important school-related predictors of student achievement” (Podolski et al., 2019), matching earlier research that indicates teachers are the most important in-school factors impacting student achievement (Chetty et al., 2014;

Rivkin et al. 2005; Rockoff, 2004). With this in mind, CCEE elected to pilot the PLC at Work model as a means to improve teacher and administrator practices that best support student learning in the traditionally struggling districts they served under DTA.

With this model in mind, CCEE brought district leaders to Arkansas in 2022 to observe and learn from Model PLC at Work schools. Leadership interviews underscored that the true marker of success was the quality of adult collaboration in support of teaching and learning. During site visits, leaders noted a complete “absence of excuses and blame.” Rather than looking to an administrator to solve problems, teachers took collective responsibility for all students, and administrators facilitated and protected the resources teachers needed to intervene. As Navo noted during a visit to a Model PLC at Work school:

[The adults] are excited about teaching. They are working around the barriers. They are collectively coming together and saying, 'Give me those kids. I'll take those kids.' That sounds different.

In California, the IAM pilot aimed to replicate this culture of collective responsibility and move beyond performative accountability toward real shifts in teaching and learning. In districts such as Sacramento City, where labor tensions and historical grievances have stalled efforts to improve practice (Hahnel & Melnicoe, 2019), the pilot sought to create a safe harbor for improving practice. The model emphasized shared leadership that balanced classroom and administrative perspectives. For example, districts used a guiding coalition, a shared leadership team of teachers and administrators, to examine teaching and learning, adjust practices, and align resources to meet student needs. In these spaces, external leadership from CCEE and experts from Solution Tree provided the impetus and credibility to protect collaboration time from being repurposed for administrative meetings, ensuring it remained focused on evidence of student learning.

Beginning in the 2022–23 school year, CCEE identified five districts: three already under DTA – Inglewood, Sacramento City, and Vallejo City – due to chronic underperformance or fiscal distress, and two – Fillmore and Madera – that joined voluntarily. To promote shared ownership and long-term sustainability, funding for the 50+ days of intensive annual coaching was split 50/50 between the state and participating districts (see Table 1 for a cost breakdown). The goal was to establish at least one “pilot school,” or model site, within each district. These schools were to receive intensive support from Solution Tree over three years to (1) improve student outcomes and (2) achieve Model PLC at Work recognition,¹ which requires a school to:

¹ “Model PLC at Work” is Solution Tree’s DuFour and Eaker Recognition tier for schools and districts that demonstrate sustained implementation of the PLC at Work process and provide at least three years of

- Demonstrate a commitment to the PLC at Work process
- Implement the concepts for at least three years
- Present clear evidence of improved student learning
- Articulate the culture, practices, and structures of the school or district

Table 1: IAM Program Participation Cost

Districts currently identified for Direct Technical Assistance		
	District Contribution	CCEE Contribution
Year 1	\$0	\$338,800
Year 2	\$164,587.50	\$164,587.50
Year 3	\$135,987.50	\$135,987.50
<i>Total for 1 School/District</i>	<i>\$300,575.00</i>	<i>\$639,375.00</i>
Districts not currently identified for Direct Technical Assistance		
	District Contribution	CCEE Contribution
Year 1	\$169,400	\$169,400
Year 2	\$164,587.50	\$164,587.50
Year 3	\$135,987.50	\$135,987.50
<i>Total for 1 School/District</i>	<i>\$469,975</i>	<i>\$469,975</i>

Because the PLC at Work model is a school-based intervention, CCEE also aimed to (3) scale and sustain the work beyond a single site and across each district. The pilot schools were intended to serve as proof points, demonstrating that a focus on collective efficacy—the belief that educators can impact student learning regardless of external factors—can meaningfully improve student outcomes. The first-year steps toward these goals were documented by PACE in *What Does It Take to Accelerate Learning for Every Child?* (Cottingham et al., 2023). A pathway for scaling this kind of intervention from a single school to districtwide implementation was also detailed in *Improving the Instructional Core at Scale A Case Study of Chino Valley Unified School District* (Cottingham & Myung, 2025), a case study of Chino Valley Unified School District, which earned Model PLC at Work district recognition in 2024–25. This report provides a final analysis of the IAM Pilot’s impact in achieving the goals described above.

student-achievement data showing growth (reviewed through Solution Tree’s Evidence of Effectiveness process). See more here: <https://allthingsplc.info/what-is-dufour-and-eaker-recognition>.

Methods

This study used a mixed-methods, implementation-focused design to understand both how the IAM was carried out across participating sites and whether participation was associated with changes in outcomes. The qualitative strand documents how key stakeholder groups experienced the IAM, how core components were enacted in practice, and what conditions supported or constrained implementation at the school and district levels. The quantitative strand complements this analysis by estimating the IAM's association with changes in school-level achievement and attendance over time, using publicly available data to situate implementation findings alongside outcome trends.

Together, these approaches allow us to (1) identify cross-site patterns and meaningful variation in IAM enactment; (2) clarify the mechanisms through which the IAM appears to influence collaborative routines, instructional practice, and intervention design; and (3) assess whether outcomes expected by the IAM's theory of action show evidence of improvement within the study's three-year timeframe. We also apply a conceptual framework to integrate the qualitative and quantitative analyses, distinguishing among foundational district and school conditions, the IAM's mechanisms of technical and adaptive change, and proximal impacts on teacher practice, and longer-term impacts on student learning.

Qualitative Approach: Semi-structured interviews

We conducted 96 semi-structured interviews with a purposive sample of stakeholders positioned to speak to IAM design, delivery, and uptake. Interviewees included educators (teachers and other instructional staff), school administrators, district administrators, Solution Tree coaches, and CCEE leaders. Interviews focused on (1) local context and readiness for IAM; (2) the nature and consistency of coaching and support; (3) changes in collaborative routines (e.g., PLC processes), instructional practice, and improvement planning; (4) enabling conditions and barriers (e.g., time, leadership stability, data systems); and (5) perceptions of IAM's value, sustainability, and requirements for scaling. We used a common interview protocol tailored by role, enabling cross-site comparability while preserving flexibility to probe emergent issues. All interviews were recorded with participant consent and professionally transcribed for analysis.

Additionally, we used qualitative data collected three times per year by Solution Tree Associates using the PLC at Work continuum, which tracks school-level implementation across 14 indicators defined in the PLC literature (see DuFour et al., 2016). The indicators include: (1) establishing a clear purpose for the work; (2) communicating that purpose to all staff; (3) establishing a shared mission; (4) a shared vision; (5) collective commitments; and (6) common school goals. They

also include developing (7) interdependent teams with (8) established norms for collaboration and (9) SMART goals for student learning. Teams then (10) determine essential standards; (11) establish a shared process for analyzing data; and (12) clarify what proficiency looks like for the essential standards. Finally, teams develop (13) common formative assessments and (14) response-to-intervention plans to support students who need additional help. Across these 14 indicators, associates rated the depth of implementation at the beginning, middle, and end of each school year, documenting evidence to support each rating. Implementation levels were defined as follows:

- (1) Pre-Initiating Stage - The school has not yet begun to address this PLC principle or practice.
- (2) Initiating Stage - The school has made an effort to address this principle or practice, but the effort has not yet begun to impact a critical mass of staff members.
- (3) Implementing Stage - A critical mass of staff members is participating in implementing the principle or practice, but many approach the task with a sense of compliance rather than commitment. There is some uncertainty regarding what needs to be done and why it should be done.
- (4) Developing Stage - Structures are being altered to support the changes, and resources are being devoted to moving them forward. Members are becoming more receptive to the principle, practice, or process because they have experienced some of its benefits. The focus has shifted from “Why are we doing this?” to “How can we do this more effectively?”
- (5) Sustaining Stage - The principle or practice is deeply embedded in the culture of the school. It is a driving force in the daily work of staff. It is deeply internalized, and staff would resist attempts to abandon the principle or practice.

We converted these ratings to scale scores of 1 at Pre-Initiating up to 5 at Sustaining to observe overall changes to adult collaborative practice over the course of pilot and understand the sequencing of changes within the schools. While subjective, these ratings informed the development of our interview protocols over the IAM, helped triangulate qualitative evidence across multiple sources, and minimized interruptions for staff working with coaches for 50 days each year. They also help to understand the sequencing of the work and what aspects of the model were prioritized in each year. We see significant variation in implementation across the schools and speed of uptake of the model. It is also worth noting that different lead associates provided these ratings so there is variation in how they were scored. Associates did provide their full ratings through an aligned rubric that included further language describing what

constituted evidence at each indicator at each rating level. Of the eight schools participating in the pilot, six shared complete continuum data across all three years, one school was closed after year three, and another provided no data. The six continuum data tables can be found in Appendix A.

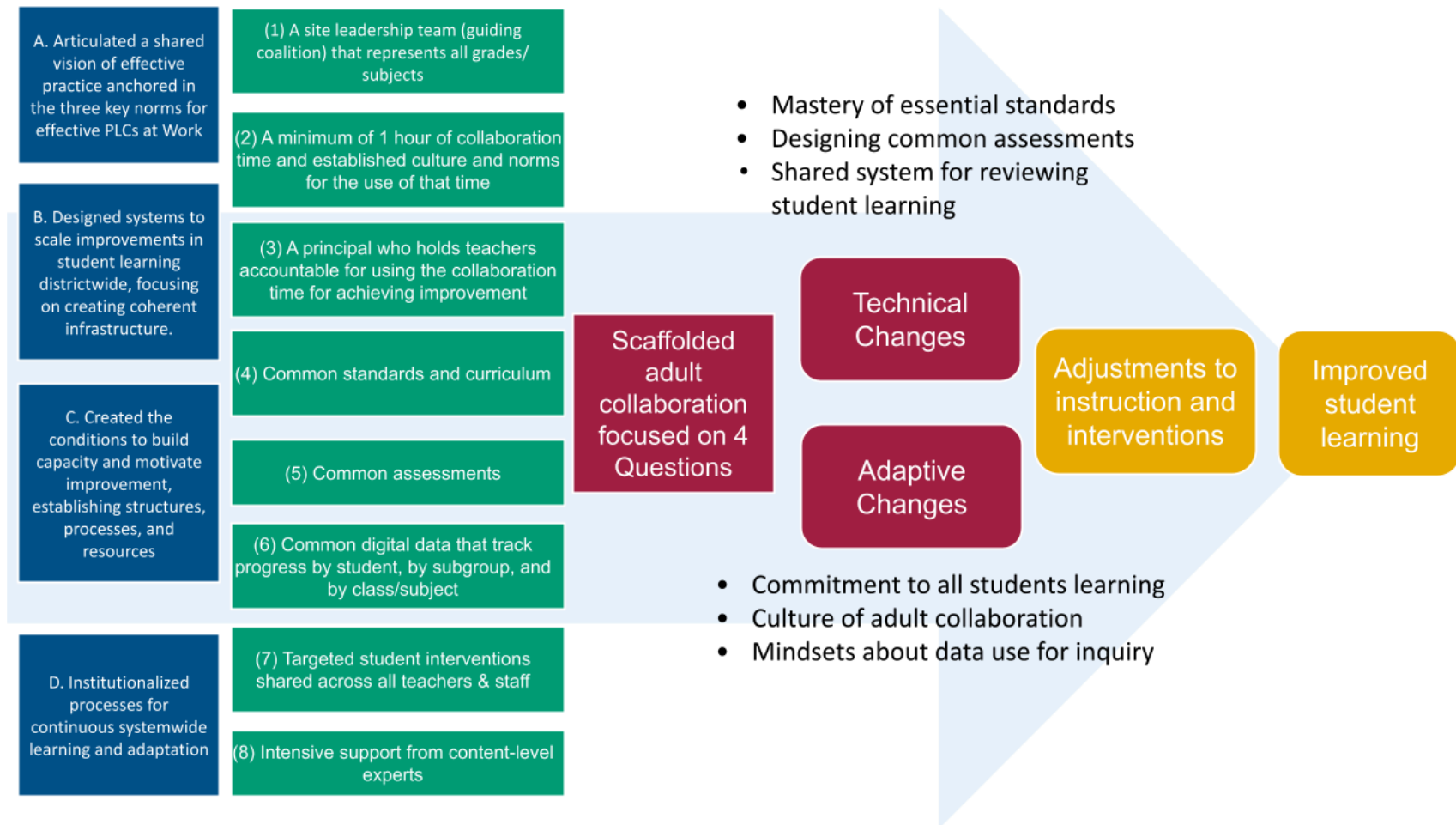
Qualitative Methodology: Analytic Strategy

We analyzed interview transcripts using a grounded theory–informed approach, developing analytic categories and explanations iteratively from the data rather than imposing them in advance. The team engaged in cycles of close reading, memo writing, and collaborative discussion to identify recurring concepts, relationships, and variation across roles and sites. We then developed and refined a codebook through iterative testing and reconciliation and applied it systematically across the full corpus of transcripts. As patterns stabilized, we used constant comparison to examine differences across stakeholder groups and implementation contexts and to identify hypothesized mechanisms linking IAM activities to changes in practice and organizational conditions.

Findings from this analysis informed the development of a conceptual framework describing IAM implementation and the factors shaping its scalability. The framework synthesizes district-level inputs and activities, school-level implementation conditions, educator-based changes that support implementation, intermediate outcomes for school staff and students, and longer-term impacts on student learning. We use this framework throughout the report to organize results and surface actionable implications for strengthening IAM and supporting expansion to additional sites.

The conceptual framework in Figure 1 presents a multilevel theory of action for how the CCEE IAM contributes to improved student learning. It links (a) district-level system building (blue), (b) school-level enabling conditions that make high-functioning PLCs feasible (green), and (c) IAM-mediated mechanisms of intermediate-term change observed in Year 3 (red), culminating in shorter-term changes to instructional practice, and longer-term improvements in student learning (gold). The framework is intended to organize the evaluation by distinguishing among upstream infrastructure and coherence-building efforts, the site conditions required for disciplined collaborative improvement, and the proximate instructional and intervention changes hypothesized to affect student outcomes.

Figure 2: IAM Qualitative Evaluation, Conceptual Framework



District-level drivers of coherence and scale

The leftmost column (A–D) represents district actions intended to build a coherent instructional improvement infrastructure and enable reform at scale. Consistent with prior PACE analyses of districtwide improvement efforts (Cottingham & Myung, 2025), these actions include: (A) articulating a shared vision of effective practice and anchoring collaboration in clear norms; (B) designing systems to scale improvements to the instructional core across schools; (C) creating organizational conditions that build capacity and motivate improvement through structures, processes, and resources; and (D) institutionalizing routines for continuous system learning and adaptation. Collectively, these district-level strategies provide the enabling environment for IAM work to be taken up consistently across sites and sustained beyond external support.

One example of a district driver that enabled IAM work to take hold in one district was its deliberate investment in protected collaboration time. District leaders framed collaboration not as an add-on, but as a core lever for strengthening school leadership and improving instruction, grounded in a broader leadership development strategy. Importantly, prior to the introduction of the IAM approach, this district had paired this shift toward increased time for teacher collaboration with expanded access to enrichment “specials” (e.g., art, STEM, drama), so that the same schedule changes that created consistent time for adults to work together also increased students’ opportunities to engage in well-rounded learning experiences during the school day. As one school principal explained, this emphasis on collaboration was shaped by districtwide learning prior to the pandemic that ultimately translated into structural changes that significantly increased educators’ time to work together in meaningful, focused ways:

Even before COVID, our superintendent and board were really adamant about bringing in NISL (the National Institute of School Leadership), and having all of the district and site leadership eventually go through that training that really focused on developing and looking at school leadership. And one of the first things we did was really look at high-performing countries and what they did differently. One of the big takeaways was they actually have less time in front of their students, but more time for collaboration. And that was something that really stuck with our district leadership, and they were really committed to providing that... And I think our superintendent and board are also really passionate about giving our students the experiences at school that maybe they're not in a place, socioeconomically, to have—like art, STEM, and drama. So it's been a huge investment... But it's been game-changing as far as the time our teachers have to collaborate. Because prior to having that, it was basically that we had early release once a week, where the kids got out like an hour and a half early. But by the time you really got the kids out, then you had maybe an hour before the end of the duty day. And so now to have this built in—it's been a game changer for me as a leader and for my vice

principal, because now all of my collaborative teams meet at different times. So I've been able to be a part of the process with all of my teams. In some teams, I'm a lot more involved than in others, but some of them now are at the point where I just kind of sit back and I'm just like, "Wow, you guys, you're on it. I'm learning things from you!" But before, when we just had an early out, all the teams were meeting at the same time. And so I'd pop in—and what happened was, I'd pop in and they'd be like, "Oh, we have all these questions," and then it would get off track... It wasn't focused on what collaboration really looks like.

In response to learning about this district's approach to pairing teacher collaboration time with student specials, one principal in a different IAM district was drawn to the idea but questioned whether it would be feasible in his own district: "The challenge is finances. It would be wonderful if we had a credentialed PE teacher on campus who could cover that time, and maybe a credentialed STEM or art teacher, too. But typically we don't...so that would be my challenge. Otherwise, I love it."

School-level enabling conditions for disciplined PLC improvement

The green elements represent core elements that a school must have in place to translate improvement intentions into routine collaborative practice. These include: (1) a representative site leadership team (guiding coalition); (2) protected collaboration time with established norms for its use; (3) principal accountability for using collaboration time to produce measurable improvement; (4) common standards and curriculum; (5) common assessments; (6) shared digital data systems that track progress by student, subgroup, and class/subject; (7) agreed-upon intervention approaches and responsibilities for students not meeting learning targets; and (8) intensive support from content-level experts. In the framework, these conditions function as a schools' *implementation capacity*: they determine whether collaborative inquiry is feasible, whether data is available and shared, and whether teams can move from analysis to coordinated instructional responses—features emphasized in PACE's work on learning acceleration and school-based improvement (Cottingham et al, 2023).

IAM mechanisms observed in Year 3: collaboration, technical and adaptive change

Building on district coherence and school readiness, the IAM operates primarily through scaffolded adult collaboration focused on the four guiding PLC questions outlined above, which structure how educators interpret evidence and make collective decisions. Within these collaborative routines, the framework differentiates between two classes of change:

- Technical changes, which involve the improvement of educators' instructional tools and routines:

- mastering essential standards,
- designing common assessments, and
- developing systems for reviewing evidence of student learning evidence to inform instruction
- Adaptive changes, which involve shifts in beliefs, roles, and norms necessary for sustained improvement:
 - re-examining assumptions about students' learning capability
 - redefining roles and expectations for collaboration and collective responsibility, and
 - mindsets on using data as inquiry for improvement rather than as judgment.

This technical–adaptive distinction comes from the adaptive leadership framework (Heifetz & Laurie, 1997), which argues that technical challenges have known solutions that can be addressed with existing expertise, while adaptive challenges require learning and often involve shifts in values, roles, relationships, and behaviors.

A central implication of this framework is that meaningful improvement depends as much on adaptive change as on technical change, and often more. Technical changes, such as identifying essential standards, creating common assessments, adopting common lessons, or strengthening data routines, are necessary because they provide the concrete tools and practices through which instruction is organized. However, the IAM is premised on the idea that these technical developments will not be implemented with fidelity or sustained over time unless schools also engage in the adaptive work required to shift how adults make sense of the work, relate to one another, and share collective responsibility for student learning.

Broadly speaking, improvement efforts often gravitate toward the technical because technical solutions are more known, easier to specify, and more readily measured. As a result, schools and districts may focus on the trappings of collaboration—pacing documents, assessment calendars, data trackers, or meeting checklists—without equally investing in the underlying shifts in beliefs and norms that make those tools consequential. This dynamic can produce a form of “surface compliance,” in which teams complete required processes but do not use them to challenge assumptions, interrogate inequities, or change instructional decision-making in durable ways.

Taken together, technical and adaptive changes are expected to generate the proximate outcome of *adjustments to instruction and interventions*—that is, changes in core instruction, differentiation, and intervention design and delivery—which, in turn, are hypothesized to improve student learning. In evaluation terms, the framework supports analyses that trace how district and school conditions shape the strength and consistency of collaborative routines, and

how those routines translate into instructional and intervention adjustments that plausibly influence student outcomes.

The conceptual framework is intentionally structured to distinguish between (1) foundational district- and school-level conditions and (2) the proximate mechanisms through which IAM work is enacted in schools. Accordingly, the qualitative approach places primary emphasis on the red components of the framework, which represent the core processes observed through Year 3 data collection: scaffolded adult collaboration organized around key PLC questions, the balance of technical and adaptive change, and how these processes translate into the near term outcomes of adjustments to instruction and interventions.

This focus reflects the evaluation's sequencing across years. Years 1 and 2 documented the extent to which districts articulated a shared vision, built coherence-supporting infrastructure, and created enabling conditions for continuous improvement (blue), as well as whether participating schools had the prerequisites to support disciplined collaborative inquiry (green). These earlier analytic strands examined how systems and structures—such as leadership teams, protected collaboration time, common instructional resources, assessment routines, and data infrastructure—were established and strengthened to support implementation.

In contrast, Year 3 data collection was designed to examine what happens once these foundational elements are in place: how the IAM is experienced within schools, what the model looks like in day-to-day practice, and what changes occurred as a result of the program. The red portion of the framework therefore serves as the primary organizing lens for Year 3 analyses, capturing the collaborative routines and adult learning processes that mediate between enabling conditions and changes in classroom instruction and student support. By foregrounding these mechanisms, the Year 3 analysis aims to clarify how the IAM operates on the ground, the adaptive and technical shifts it appears to catalyze, and the plausible pathways through which those shifts contribute to improved student learning.

Finally, it is important to interpret findings in light of the evaluation's three-year timeframe. The IAM is designed to support change that is inherently multi-year: developing shared instructional expectations and collective responsibility for results; shifting norms so that data function as a tool for inquiry rather than judgment; strengthening teacher leadership and redefining roles within PLCs and guiding coalitions; building and routinely using common assessments and shared data routines; and coordinating timely, consistent intervention systems for students not yet meeting learning targets. Transformations of this kind typically require sustained implementation, repeated cycles of learning, and stable organizational support before they are reflected in large shifts on more distal indicators. As such, while the evaluation does not assume that three years is sufficient to produce dramatic changes in longer-term outcomes, we would

expect to see early changes in more proximal outcomes—for example, greater consistency and quality of collaborative routines, more coherent instructional and intervention adjustments, and increased alignment in the use of evidence to guide improvement. These nearer-term indicators are therefore treated as critical signals of whether the IAM is taking root in ways that can plausibly generate stronger student outcomes over time.

In this report, we first describe the technical and adaptive factors evident in Year 1 of implementation and then revisit these factors in Year 3 to examine how they evolved over time. We treat Year 3 conditions as an indicator of the IAM’s proximal contributions to educators’ professional learning and growth—capturing changes in routines, beliefs, and organizational conditions plausibly shaped by the model, as well as evidence of changes in instructional practice associated with the IAM. We then complement this qualitative analysis with a quantitative assessment of the IAM’s association with student learning outcomes, conceptualized as a more distal measure of impact that may begin to reflect the cumulative effects of instructional and organizational change.

Quantitative Approach

The quantitative section of the report presents findings from the effect of IAM implementation in the 2022–23 school year across schools in five California districts: Vallejo City Unified School District, Madera Unified School District, Inglewood Unified School District, Sacramento City Unified School District, and Fillmore Unified School District. The study leverages up to eight years of publicly available, school-level aggregate data on student enrollment, test performance, and school attendance from the California Department of Education (CDE). The panel contains between four and five years of pre-intervention data—though assessment data were not available for two COVID-impacted school years (2019-20 and 2020-21)—and three years of post-implementation data, allowing us to estimate the effect of IAM participation on student achievement in math and English Language Arts (ELA) and student attendance outcomes. Furthermore, we explore how the effects of IAM participation differentially affected students in different grade levels.

The quantitative analysis addresses three research questions:

1. What is the effect of IAM participation on student achievement and attendance?
2. To what extent did the effect of IAM participation vary across post-implementation years?
3. Are the effects of IAM different for students in particular grade levels?

Quantitative Data

This study relies on publicly available data published annually by the California Department of Education (CDE). These data are aggregated summaries of student-level accountability data submitted by school districts to CDE; the level of aggregation varies by the domain of school and district characteristics examined. Depending on the data type, the files used for this evaluation span the 2016–17 through 2024–25 school years. We linked these files to the IAM schools and districts using the 14-digit County-District-School (CDS) code, a unique and stable identifier for each California public school. The table below summarizes the data files collected for this study. Additional information on the source data can be found in Appendix B.

Table 2: Quantitative Data Sources

Data File	Content	Source	Years
CAASPP assessment results	Smarter Balanced Summary Assessments: math and ELA scale scores and proficiency rates, grades 3–8	CDE	2016-17 to 2018-19; 2021-22 to 2024-25
Attendance	Average days absent, attendance rate	CDE	2017-18 to 2018-19; 2020-21 to 2024-25
Chronic absenteeism	Rate of students missing ≥10% of instructional days	CDE	2016-17 to 2018-19; 2020-21 to 2024-25
Cumulative enrollment	School enrollment (overall and by student subgroups)	CDE	2016-17 to 2018-19; 2020-21 to 2024-25
Public Schools and Districts	Operational school characteristics	CDE	NA

Note: The 2019-20 school year is excluded from all analyses, and the 2020-21 school year is excluded for the testing outcomes due to school closures and testing disruptions during the COVID-19 pandemic.

Inclusion Rules and Sample Restrictions

All eight IAM schools were included in the impact models, except for grade-by-year combinations where data were not available (for example, analyses of Grade 8 test scores necessarily exclude elementary schools with K–6 configurations). Our primary sampling frame is limited to districts where IAM was introduced as part of the pilot program in the 2022–23 school year. Accordingly, the pool of potential comparison schools includes campuses in the five IAM districts. Because IAM was implemented only in elementary and middle schools, we restricted the comparison pool—applying the same rules to the statewide data used in

sensitivity analyses that draw comparison schools from across California—based on the “soctype” of IAM schools. Specifically, we included schools with soctype codes 60 and 62, which denote Elementary Schools (Public) and Intermediate/Middle Schools (Public), respectively.

Outcome Measures

Our primary measures span two dimensions of student outcomes:

- (1) Performance on state standardized assessments in math and ELA, operationalized as standardized scale scores and the percentage of students who met or exceeded the performance standard in the respective grade and year. Test scores were standardized by subject, grade, and year using the statewide means and standard deviations, weighted by the number of test-takers in a given cell; and
- (2) School-day attendance. CDE publishes two measures of school-day attendance: (1) the percentage of days attended, a derived metric since the attendance data do not directly report this measure. We estimate attendance rates using the total number of days missed and cumulative student enrollment. Specifically, we multiply the cumulative number of students enrolled by 180 to obtain a denominator necessary to calculate the attendance rate, and the share of students who were chronically absent, operationalized as the percentage of students who missed 10 percent or more of instructional days. We summarize these outcomes in Table 3 below.

Table 3: Description of outcome measures

Achievement outcomes	
Variable	Description
Math scale score	Mean CAASPP math scale score among tested students (standardized)
ELA scale score	Mean CAASPP ELA scale score among tested students (standardized)
Percentage proficient in math	Share of tested students meeting or exceeding the standard in math
Percentage proficient in ELA	Share of tested students meeting or exceeding the standard in ELA
Attendance outcomes	
Variable	Description
Attendance rate	Percentage of eligible (180) days attended

Chronic absenteeism rate	Percentage of enrolled students missing 10 percent or more of instructional days
--------------------------	--

Grade Levels and Subjects for Testing Outcomes

CAASPP administers math and ELA assessments in grades 3–8. In this report, the primary level of analysis for the assessment outcomes are the school-by-year level, aggregating across all tested grades within a school. We present the findings of subgroup analyses in the Quantitative Results section below that disaggregates results separately by grade level. We restrict this analysis to elementary grades (3-6) due to limited treatment exposure in non-elementary grades (only one middle school, Fillmore Middle School in Fillmore Unified School District, adopted the IAM model during the pilot phase covered in this study.)

Quantitative Methodology: Analytic Strategy

Our identification strategy is a quasi-experimental difference-in-differences (DiD) design, paired with an event-study specification that harnesses between seven and eight years of longitudinal school-level data. Given that participating schools were not randomly assigned, our approach draws on panel data from participating and non-participating schools to construct a counterfactual scenario of academic performance among statistically similar non-IAM schools. The deviation from this prediction is the estimated treatment effect of IAM — that is, using data from all schools allows us to estimate the extent to which IAM schools' outcomes differed from their predicted trajectory had the intervention not been implemented. We estimate two types of treatment effects. First, using two-way fixed effects (TWFE), with year and school fixed effects to absorb time-invariant characteristics that may jointly influence the decision to implement IAM and school performance, we estimate the composite effect of IAM participation on student outcomes across all post-intervention periods. Second, we relax the assumption that a single "IAM" treatment effect exists, estimating dynamic treatment effects for each post-intervention period using an event-study specification. We provide more formal descriptions of the econometric approach in Appendix C.

This design is well-suited for policies that are implemented at a cluster level — in this case, a school — in that the program affects the entire unit. Given the availability of extensive baseline data, this design enables us to more precisely estimate the counterfactual for untreated units. We further bolster the strength of the design by limiting the candidate pool of comparison schools to schools within targeted districts to minimize the threat of unmeasured district factors compromising the internal validity of our estimates. Together, these factors allow investigation into the plausibility of the assumptions required for credible causal estimates of the introduction of the program in a non-experimental setting: most importantly, that prior to

implementation, the trajectories for the outcomes of interest were equivalent between conditions.

An important consideration when interpreting these two sets of impact findings is to understand the question they answer: Did IAM schools improve more than similar non-IAM schools would have absent the program? Near-zero effects can occur *even when IAM schools improve in absolute terms*, if comparison schools improved at a similar rate. Or, conversely, effects can be negative *if IAM schools improved more slowly* than non-IAM schools. Indeed, the study documents steady improvements in student attendance in the post-implementation period: however, these secular improvements occurred during the post-COVID recovery period, as school attendance rates improved nationwide, though not fully restored to pre-epidemic levels.

Weighting to Achieve Comparability Between IAM Schools and Non-Participants

IAM schools in the five participating districts vary considerably in their student populations and prior academic trajectories. Directly comparing outcomes at IAM schools to all non-IAM schools — without accounting for these differences — could confound program effects with pre-existing differences between schools.

To address this, we construct a weighted comparison group using balancing methods that rely on the probability — or propensity — to implement IAM. The model estimates each school's probability of being an IAM school based on baseline characteristics measured in the year before IAM launched (2021-22), including prior achievement (math and ELA), prior attendance, school enrollment, and student body composition (shares of economically disadvantaged students, English learners, students with disabilities, and students by race/ethnicity). We also incorporate multi-year averages (up to six baseline years) for each covariate to alleviate concerns about mean reversion injecting bias into the estimated effects of participation. Comparison schools are then assigned weights that make the weighted comparison group resemble IAM schools on these observed characteristics. Comparison schools are drawn from the same district as the IAM schools they are compared against, ensuring that IAM and comparison schools face the same district-level policies, resources, and demographic trends throughout the study period. The balance achieved by this weighting approach is documented in Appendix C.

How should I interpret the impact findings in this report?

In the Quantitative Results section of the report, we present findings from the impact models using tables and figures. The individual estimates from this model summarize two differences between IAM and non-IAM schools. That is, these coefficients do not represent simple cross-sectional, regression-adjusted differences between IAM and non-IAM schools. Rather, they represent **differences-in-differences**: the change in outcomes at IAM schools relative to the year just before IAM launched (2021-22), minus the corresponding change at comparison schools over the same period.

For instance, a coefficient of +2 percentage points for the share of students meeting or exceeding the proficiency standard for math in the 2023-24 school year (so, T+1, or one year post-implementation) means that IAM schools improved by two percentage points more than comparison schools did relative to the shared baseline period.

Analytic Sample

The study includes eight schools in the five IAM districts presented in Table 4. Of the eight schools, the vast majority (7) serve elementary grade students, while one (Fillmore Middle in Fillmore Unified) serves middle school students.

Table 4. Intensive Assistance Model schools and enrollment, 2022-23

District	School	School Type	Number Enrolled
Inglewood Unified	Highland Elementary	Elementary	502
Madera Unified	Nishimoto Elementary	Elementary	753
Madera Unified	George Washington Elementary	Elementary	673
Madera Unified	Sierra Vista Elementary	Elementary	768
Sacramento City Unified	Elder Creek Elementary	Elementary	816
Sacramento City Unified	H. W. Harkness Elementary	Elementary	319
Vallejo City Unified	Federal Terrace Elementary	Elementary	349
Fillmore Unified	Fillmore Middle	Middle	878

The table below presents baseline characteristics of IAM schools and the unweighted comparison pool within the five IAM pilot districts. Before reweighting, the two groups differ on student demographics and prior achievement. IAM schools have lower shares of socioeconomically disadvantaged students (76 percent versus 90 percent) and Black and Hispanic students (70 percent versus 82 percent). On academic outcomes, IAM schools entered the study period with higher standardized scores, approximately 0.3 standard deviations above comparison schools in both ELA and math, and higher rates of students meeting or exceeding proficiency standards. Attendance and chronic absence rates were similar across groups.

Table 5. Characteristics of Intensive Assistance Model and comparison schools at baseline (2021–22)

School Characteristics	IAM Schools	Comparison Schools
Average of socioeconomically disadvantaged, percent: school-level (t-1 to t-6)	76.17	89.93
Average of percentage Black and Hispanic students (t-1 to t-6)	69.79	81.58
Average of standardized ELA and math scores (t-1 to t-6)	-0.59	-0.89
Average of all students:count: school-level (t-1 to t-6)	548.73	674.10
Average of chronic absence rate: total (t-1 to t-6)	20.34	21.04
Average of attendance rate: total (t-1 to t-6)	93.65	93.41
Proportion of elementary schools	0.87	0.88
Mean scale score (standardized): ELA (t-1)	-0.58	-1.01
Mean scale score (standardized): math (t-1)	-0.57	-0.87
Attendance rate: total (t-1)	90.24	89.19
Chronic absence rate: total (t-1)	40.48	44.08
Percentage standard met and above: ELA (t-1)	32.58	24.35
Percentage standard met and above: math (t-1)	22.93	16.90

Study Scope and Analytic Lens

To maintain analytic focus on IAM as a programmatic intervention, and to avoid implying that the report evaluates any single local education agency, we anonymize participating districts and, where necessary, schools and individual roles throughout the report. This approach aligns with the study's purpose: identifying cross-site patterns in implementation, the conditions that enable or constrain improvement, and implications for scale, rather than attributing specific outcomes to particular districts.

A persistent tension remains between policymakers' demand for clear, comparable numbers and the slower, less visible work of cultural and instructional transformation. As one leader reflected, "If we can figure out how to measure the foundational building blocks of school transformation, we'll be better off." The challenge is making the case that the everyday signals of collective efficacy—such as a teacher saying, "Can you help me? Your students learned this better than mine"—are not anecdotal, but essential indicators of improvement, alongside standardized test results.

The IAM pilot reflects a shift in California's improvement strategy from "proving" to "learning." By organizing adult collaboration around the four PLC questions and creating conditions for honest, evidence-based reflection, CCEE tested an approach to state support that engages the instructional core—teaching and learning—rather than operating primarily through compliance structures. The pilot's legacy, then, is not only potential gains in a subset of schools, but an emerging, more systematic approach to building the infrastructure for internal accountability and student-centered collaboration that can be sustained beyond external oversight.

Qualitative Analysis of Year 1

In this section, we describe qualitatively where each pilot school and district began and examine the technical and adaptive changes they advanced by Year 3, with attention to the collaborative mechanisms through which the IAM was enacted. Building on district coherence and school readiness, the IAM operates primarily through scaffolded adult collaboration organized around the four guiding PLC questions, which structure how educators interpret evidence, surface problems of practice, and make collective instructional decisions.

The continuum data collected indicated that teams moved most quickly on the development of their "why" of the work and setting up their PLC teams (rubric numbers 1-8). This is consistent with the literature around the PLC at Work model and what associates shared with us at the initiation of the project. These steps to establish the school's rationale for the work and establish their expectations for collaborative work are to set the foundation for implementation

of the technical and adaptive shifts required by the model. Of the technical shifts we focused on, the development of essential standards moved much more quickly (1.86 to 3.29) than creating common formative assessments (1.71 to 2.86), reviewing student data (2 to 2.57), and by extension response-to-intervention systems (1.57 to 2.57) (See Table 6).

Table 6: Year 1 Aggregated District Change Continuum Data

Rubric #	Indicator	Year 1		
		Start	Middle	End
1	<i>Clear Purpose</i>	2.14	3.57	4.00
2	<i>Communicate Purpose</i>	2.43	3.29	4.00
3	<i>Shared Mission</i>	1.86	3.14	3.57
4	<i>Shared Vision</i>	1.86	3.14	3.86
5	<i>Collective Commitments</i>	1.86	2.71	3.29
6	<i>Common School Goals</i>	1.71	2.86	2.86
7	<i>Interdependent Teams</i>	2.57	3.57	3.71
8	<i>Collaborative Norms</i>	1.86	3.00	3.14
9	<i>Smart Goals</i>	1.43	2.29	2.29
10	<i>Determine Essential Learning</i>	1.86	3.29	3.29
11	<i>Analyze Data</i>	2.00	2.43	2.57
12	<i>Clarify Proficiency</i>	1.43	2.29	2.57
13	<i>Common Formative Assessments</i>	1.71	2.57	2.86
14	<i>Response-to-Intervention</i>	1.57	1.86	2.57

The following quotes provide further description around the adaptive and technical changes that teams experienced by the end of year one.

Mastery of essential standards

In the IAM, collaborative work begins with the PLC guiding question, “What do we want all students to learn?” Teacher teams address this question by focusing on essential standards—the priority learning outcomes identified as most critical for student success and future learning. A foundational technical requirement is that teachers develop strong mastery of these standards themselves. Building this expertise involves deeply understanding the intent

and rigor of priority standards, unpacking them into clear learning targets, and developing a shared, accurate conception of proficiency through rubrics, exemplars, or proficiency scales.

This technical skill enables teams to design and use common formative assessments aligned to the essential standards and to interpret evidence of student learning consistently across classrooms. With a shared understanding of the standards and what constitutes mastery, teams are better positioned to identify learning gaps, plan targeted reteaching and supports for students who are not yet proficient, and provide meaningful extension for students who have already mastered the content. Over repeated cycles, this teacher-level mastery strengthens the quality and coherence of instruction and supports continuous improvement toward higher levels of student mastery of essential standards.

Across all pilot districts and schools, the Year 1 focus was on selecting and deconstructing essential standards. Interviewees at the district office, school leadership, and teacher levels described this process as critical for aligning grade-level teams and school sites around shared instructional priorities. Narrowing down to the essential standards in each subject—math and ELA—created space for teachers to go deeper on essential concepts, rather than trying to teach everything with the same level of emphasis. One principal noted the “relief” this narrowed focus provided to teaching staff:

Our school mainly focused on identifying our essential standards...every team kind of started with math... our team got more support, so we also have our essential standards for ELA... that was a huge step, kind of taking a weight off our shoulders, knowing that I have to make sure my kids know these 10 to 12 standards. And yes, I am still in charge of teaching that 30 something in total...but not at the same depth that I have to teach just 12 standards. Knowing that our teams were or at least our team was aligned in that. That was a huge relief.

The focus on essential standards enabled deeper, more sustained engagement with a smaller set of critical topics. Students received more instructional time in these areas, and teachers reported that they could provide intervention more easily when needed. Teachers and principals also noted that the essential standards supported conversations about vertical alignment across grade levels in guiding coalition and PLC meetings. Staff in participating schools understood that the selected essential standards were also a prerequisite for developing common formative assessments (CFAs), which would later enable teams to examine and discuss student outcomes. Most pilot schools began by identifying math essential standards in Year 1, with the goal of engaging in the same process for ELA the following year.

The essential-standards process helped teacher teams surface misalignment across classrooms and across the instructional resources they were using (e.g., assessments, curricula, etc.). One district administrator noted how the process of selecting essential standards differed from the district's earlier work to develop scope-and-sequence guides:

When teachers are going through the process of taking the standards and putting them out into the scope and sequence, and then picking the essential standards, and picking learning targets, and then unpacking the standards, and really understanding what the standards are asking students to be able to know and do. Then they pick the materials that match that aspect of the standard. And so you're there times where if you take this massive textbook, you might only be using half of it now. Yeah, look from page to page chapter to chapter... I noticed that was different from when they built the scope and sequence. And I think the architecture for collaboration is new at the school, where you have the ethos in place, and you have now scheduled collaboration, you have the PLC questions, you have the touchstone of the standards, and the elements of the standards are being focused as the touchstone for the conversation. And then it naturally leads to conversation around. Okay, well, how will we even know that the standards are being learned? Okay, well, that leads you naturally to an assessment. And also now you're having to look at assessments. And the and the assessments have to align to what the students are learning. So now you're looking at your old assessments going...these don't even reflect the standard. The question doesn't match the rigor of the standard or the nature of the standard. So it's really forcing them to reflect on the quality of the intellectual work that they're required to do. And they're discovering that there's a lot of misalignment and system with that process of discovery.

This discovery process—selecting essential standards and surfacing instructional misalignment—began to build teachers' sense of agency in choosing the resources and practices most likely to support student learning of the essential standards. Several interviewees noted that teachers were less likely to go “page to page, chapter to chapter” and instead were becoming “critical consumers of curriculum,” carefully considering which instructional tools would best support student learning.

Districts that encouraged teacher teams to adapt and build on existing instructional resources—such as scope-and-sequence documents and pacing guides—appeared to move more quickly from essential-standards selection to developing common formative assessments and instructional units. The district lead for curriculum and instruction in one district described the conversations her team had with grade-level teams at each site, noting that supporting teams to refine materials represented a shift in how the district developed resources to support teachers:

The district had created the scope and sequence with materials and units. But it was every standard in every grade level in every content area. And Solution Tree took our stuff and went, this is beautiful but there's no way you can teach all this and if your teachers are trying, they're never going to get to tier two. They're never going to be able to scaffold, they're never going to ever try to meet kids where they are because they're simply trying to teach new content every single day. I was like, Oh my gosh, finally, validation. So from about October to now, my elementary team got on board.

Supporting site teams to adapt district resources as needed helped teams move more quickly—rather than reinventing the wheel—and positioned the district office as a collaborator during the project's first year.

Establishing essential standards also prompted a cascade of changes in how adults approached their practice and its impact on students. The ability to select essential standards is an enabling condition for developing CFAs and examining student data. Starting with essential standards refocused staff on what students needed to learn and supported the development or refinement of instructional resources aligned to those learning goals, informed by evidence of their impact on student learning. It was a seemingly small shift that, over time, led to meaningful changes in how adults approached their work, as one principal described:

What happened was a shift, and it's a simple shift, that I'm almost kicking myself that I hadn't seen it, or was aware that that needed to happen. So the shift was from constantly focusing on what we're teaching, to, get this, to what students are actually learning. From the preoccupation of, "Have I taught this? Have I taught that? Have we presented this lesson? Have we gotten to this unit?" We got rid of all that and really shifted to what our students are learning, and is what they're learning essential. So I mean, it's simple, it's the simplest of shifts, but it's made a huge difference. And I'm not gonna lie, it was very overwhelming at first. Because many of us, including myself, are very hung up on the, well, the district is telling us to cover these seven units in these three months, and, you know, we're being told by the district to use this book and this book only. And so there's a lot of, let's say, well, meaning and intention, guardrails that our district puts in place and so we're always of that mindset, make sure I teach this make sure I teach that. This year, we shifted that focus to okay, what are our students learning and is it essential? It almost kind of focused us even more, you know, it allowed us to, yes, focus on what needs to be taught, but focus that really on what are the students learning and monitoring it.

Across sites, we heard a consequential development in Year 1: a shift from focusing primarily on what adults taught to focusing on what students learned. Teachers moved away from a checklist

mindset—“Have I taught this?”—and toward collective attention to evidence of student learning: “What are our students learning, and is it essential?” While straightforward, this reframing is profound. It reflects a simple but foundational shift that unlocks the power of PLCs: when teams organize around student learning rather than lesson coverage, common expectations, common assessments, and shared responsibility for outcomes become central to advancing the work. Providing teachers with dedicated time and a structured process to align their understanding of student learning helped lay the groundwork for deeper conversations about how to support students more effectively and refine instructional practice.

Designing common assessments

A second technical shift in the IAM is strengthening teachers’ capacity to answer the PLC guiding question, “How will we know if students learned it?” in relation to essential standards. This involves designing common assessments—specifically, co-developing CFAs aligned to shared essential standards—and using the resulting evidence to guide instruction. Developing this skill requires concrete, learnable practices: translating standards into assessable learning targets, writing or selecting items that match the intended rigor, agreeing on scoring criteria, and establishing routines to analyze results consistently across classrooms.

While the essential-standards work was consistent across sites in Year 1, the development of CFAs—and the effective use of CFA data—varied across the pilot. Schools with new leadership, limited collaborative structures, or low-trust cultures often struggled to develop, analyze, and act on evidence of student learning from CFAs. One Solution Tree associate described these Year 1 challenges as needing to “get all the pieces” in place before practice could meaningfully change:

Schools have identified their essential standards and that teams are using those to guide their work within a cycle of instruction. Unit by unit, I want teams to be clear on what their students need to know and be able to do to have some means of monitoring that using common assessments. Then, and this part is a little fuzzy in year one, really using that information to help kids learn more. There are a lot of pieces to put in place to get to that place, but most teams now are really, truly going through those cycles. They're not at the place where they're feeling confident about analyzing and using the information coming from their assessments, but they, definitely, know and we call it the Plan, Do, Study, Act Cycle, getting them into that ongoing process as a team. Most of the time that they are spending right now is focused on critical questions one and two, "What do we want kids to know? How will we know they've learned it?" They're doing backward planning of units. All of that work, by the end of the year, they should have a

lot more confidence...in the process...so that they can hit the ground running next year. They're in the process of doing a lot of getting the pieces.

Across schools, a common challenge was developing instructional resources while simultaneously building the structures and skills needed to implement PLCs at Work. By the end of Year 1, implementation varied notably—particularly in the progression from establishing essential standards to developing common assessments.

As with essential standards, some schools were able to build on existing district resources to create common assessments. In Fillmore, district leaders had been working on essential standards and aligned assessments prior to the pilot; the pilot prompted them to bring more teacher voices into that work. School staff began developing common formative assessments that were then taken up by the district team. Relatedly, the district reduced the number of assessments it asked schools to administer so teachers could focus more tightly on common assessments that directly informed decisions about teaching and learning.

Schools focused on internalizing a process for developing and refining CFAs for each unit. In one district, the school began by selecting essential standards and quickly moved to creating CFAs. Many teachers initially assumed that developing, administering, and reviewing CFA data would be a one-time exercise. Instead, the school spent the second semester establishing a shared process for using CFAs consistently through the remainder of Year 1. Developing CFAs helped solidify vertically aligned essential standards within instruction, but the more consequential shift was how regular review of CFA data began to shape teachers' instructional decisions.

Shared system for reviewing student learning

In the IAM, the third guiding question—"How will we respond when some students do not learn?"—depends on having a shared, reliable process for reviewing evidence of student learning and coordinating timely action. Accordingly, a third technical shift is the development of a shared system for reviewing student learning: regular team routines, common tools and protocols, and agreed-upon decision rules that enable educators to examine formative evidence together and plan timely, aligned responses when students are not yet meeting expectations.

In all but one district, the district first had to create time for teams to collaborate regularly and analyze student data. All Solution Tree Associates noted that the PLC at Work model requires at least one hour of collaboration time per week to succeed. Securing this time was difficult in four of the five participating districts and often required a patchwork of coverage strategies, including principals covering classes, one teacher covering multiple classes simultaneously, coaches providing coverage, requests for dozens of substitute teachers, and college student

volunteers. Overcoming time constraints was a critical prerequisite for developing a shared process to review student learning.

When CFAs are aligned to essential standards, the resulting data provide a snapshot of student learning that can inform instructional next steps and, when used consistently, can help build trust across a school. One of the primary challenges at Highland Elementary was a lack of trust—between administrators and teachers and among teachers themselves. The principal used regular data conversations to celebrate any observed movement in student learning and to problem-solve how best to support students who needed additional help. Over time, teachers began to experience data analysis as a form of administrative support rather than a punitive post-mortem. Shifting this entrenched data-use culture was a key Year 1 hurdle, as the associate at on school described:

One of the things we always have to break through with schools that have traditionally underperformed, is they don't like looking at data, because using data is a piece around which they get beat up. Right, who's responsible for this data? ...even talking to [the principal], at the end of the year, when we looked at his iReady data. He's like, "This is the only thing we improved?" But is it? It's not where it needs to be? Absolutely..., but it's better than what it was. So let's celebrate that. Let's recognize the work that we've done there, right? Continuous improvement is what a professional learning community is based on...If you don't stop and celebrate along the way...oftentimes people get worn out, they get ground down. So we really encouraged him at the end of last year at the beginning of this year. When teachers get back there, let's celebrate. This is where we were, this is where we are. It's not where we're going to end up. But it's where we are and recognize that path is important to be on...That's going to be an important piece to get teachers to more honestly and ably look at data...whether it's good data or bad data, it's how do we react to it? It's the most important question. It's not something to be beat up over. It's not something to be, Alright, who's responsible? Right. And that's going to be that's a little bit different than what they've experienced in the past.

The administration worked to build trust with teachers through constructive data conversations and by expanding the guiding coalition to include teachers who wanted more of these types of discussions. As the school's culture shifts—often slowly—staff hope to continue strengthening the technical tools related to standards, assessment, and data-monitoring processes. At this school, its commitment to consistent processes for developing and using instructional resources (essential standards, CFAs, and data-monitoring routines) began to pay dividends by the end of Year 1, with teachers beginning to share responsibility for students across grade levels based on CFA data conversations.

Commitment to all students learning

The first adaptive change in the IAM is building a shared commitment to all students' learning, which requires changes in beliefs, expectations, and collective responsibility—not just new structures or tools. As an initial step, each pilot school defined its mission, vision, values, and commitments to the work in the guiding coalition and then carried these conversations into PLCs through grade-level teams. Interviewees consistently described these conversations as the starting point of their work with Solution Tree and as defining the “why” of the pilot. Participants spoke readily about this process, and none reported disagreement with the overall purpose of the work.

Some teachers, however, raised concerns about the time and energy required to carry out other aspects of implementation, including the technical changes described earlier. In addition, some staff who wanted to protect their autonomy expressed tension with the collaborative learning expected in the PLC model. Still, the need to improve teaching and learning was not contested in any district, and establishing shared commitments to the work generally went smoothly.

Culture of adult collaboration

Shifts in adult collaborative culture emerged early in the pilot as a second core adaptive change. Teachers began bringing instructional resources and evidence of student learning into collaborative meetings. As one principal described, she observed a clear shift in teacher collaboration during PLC meetings:

Where I'm really impressed, and I really have to take a step back sometimes I just and I marvel at it is the improvement in teacher professional practice. You know, I was really concerned when we first started on this path that the teachers would question, you know, what, why us and they would be resistant...they come with their evidence of student learning...I would be so nervous on those days when [our associate] was here, and like, Oh, my God, they're not going to bring anything they're going to come empty handed, but nope. Never. They never let me down. It was they, they would bring their work, and they would sit there and they would look at it. And they've been great together...I know students' scores are going to come along. But where I'm just really impressed is that shift in professionalism, and in truly being engaged with the work on the part of my teachers.

This shift also included teachers feeling freer to use instructional supports—such as pacing guides and curricular materials—strategically, rather than complying with every resource as written. As one teacher shared, “We were given permission by the person at our district level

who was in charge of this project... that we could basically toss out the planning guide if we needed to and start from scratch.” While this autonomy was helpful, it also represented a shift in how teachers approached instruction, and some noted it was challenging to adjust.

Implementation of the PLC at Work model also enabled teachers to support one another more quickly when student learning began to lag. In one small school with only one teacher per grade level, in the second year of the pilot, a long-term substitute—whose background was in physical education—had to take over a third-grade classroom. Because teachers had spent the previous year establishing essential standards and vertically sequencing CFAs, staff were able to step in, provide instructional support, and take students for intervention even across grade levels. As the Solution Tree associate described the support that emerged for the teacher and students:

All this didn't come from the admin. This came from the teaching team of like, how can we get these kids? We can't let these kids slide. We got it, we got to fix it, we got to help them out, you know, and there was no judgment, there was no ill will. It was just, it is what it is, how do we fix it? A real sense of urgency.

Over the course of the first year, the school quickly developed a culture of collective responsibility in which all teachers supported the learning of all students. The school now has a data wall in the staff room that displays student proficiency across essential standards. Any staff member can be tapped to intervene with any student, and instructional resources are made available to support intervention or extend learning as needed.

Principals also noted taking a more active role in PLCs through the development of guiding coalitions and having time to join grade-level team meetings. Principals facilitated guiding coalition meetings to set commitments for the work and align grade-level leads on the focus for their team meetings. A consistent message from Solution Tree Associates was for principals to delegate operational and logistical responsibilities to other staff so they could more fully support the instructional conversations happening in grade-level meetings. This, in turn, strengthened alignment between the guiding coalition (leadership meetings) and the work teachers were doing in PLCs. Several principals described this as a significant shift in how they approached instructional leadership at the school site.

Principals in multiple districts also described having to “manage up” by asking district leaders to remove initiatives so schools could focus on the intensive PLC work. In several cases, principals negotiated with district offices to opt out of the many initiatives their schools had been selected to participate in. Principals in two districts noted having upwards of 15 different coaches, including Solution Tree Associates, supporting their schools, which at times overwhelmed staff with competing requirements. One principal described the need to limit initiatives if the district

wanted the work to succeed, and to leave decisions about participation to school administration:

The distribution of initiatives or equitable treatment in terms of like, let's give some of the schools that are have all these different initiatives plus DOI, you know, some some autonomy to say, hey, this EI roadmap, we're gonna hold off on that or just kind of do like, not as intense implementation. Or, you know, let's give our teachers that are trying to clear the credential, let's take some of those tasks off their plate and kind of combine some of those requirements with the unit planning process. So we are the ones really pushing for that. And like I said, sometimes we get the win, and sometimes we're told not yet. And so that is a challenge. And that's something that for, district staff, that would be a piece of feedback that they could really learn, you know, learn from or hear it really hear us like to say, let's really be intentional and equitable. When we're assigning initiatives to all the schools, and really understand the scope of what some of these projects entail, so that we don't overwhelm the schools, because if you're trying to do ten things well, as opposed to trying to do three things well, yeah, right, like what is the priority is, in other words, like prioritize the initiatives, and give the schools a little bit more autonomy, when, like you said, what are we going to focus on? What's going to be our priority?

Removing initiatives from schools' and teachers' plates is one way to create the space and resources needed to focus deeply on instruction, rather than diffusing attention across competing priorities.

County and district leaders noted shifts in school culture during the pilot, which contributed to four of the five districts developing scaling plans for the pilot's second year. As an associate at on school described, county and district administrators wanted to support the changes they observed at the and extend them across the district:

The [district leader] showed up one day, came in, he just said, I'm gonna walk through some classrooms if that's okay. And [the principal] said, Okay, go ahead. And he goes, Well, do you want to come with me, you don't have to. And [the principal] said, I trust what's going on in our classrooms, you can go ahead and go. And so he went, he came back about two hours later. And he said, I just want to tell you, this is the first building I'd been in where everybody had learning targets posted in their classroom, where everybody's teaching was matching the learning targets. And the kids were largely engaged. So that's just one example of what they were seeing on campus was what they wanted from everywhere. The intentionality of instruction, the intentionality of purpose, had grown significantly, not just last year, but also through this year, too...

Towards the end of last year, when they were really talking about making some significant changes across the district, that's what they wanted...The senior Cabinet members, that essentially they were like, "What do we want for teaching in our district?" And somebody said, "Well, we essentially want to be like [the IAM pilot school], right, and how intentional it is." And they said, "Okay, well, how do we make that happen?" And so trying to replicate the support that [the IAM pilot school] had last year and this year to help the principal grow in that way, all those pieces they want to do at every site, in the district to we're working with Ed services team as well, to build to build their understanding of what a Professional Learning Community is and to build their understanding of how to support that and grow that, because they can't have it if they don't understand it.

Focusing on the technical work of essential standards, CFAs, and data-monitoring routines contributed to a more "intentional" culture of teaching and learning—one that outside observers could see and feel in daily instruction. The district has likewise been intentional about how to leverage the capacity being built at the IAM pilot school to support schools across the district.

Several districts are considering ways to leverage principals and teacher leaders from pilot sites to spread learning to other schools. District principal leadership meetings, in particular, emerged as a key venue for sharing lessons from pilot sites. Several districts planned for pilot-site principals to present to their peers over the summer in preparation for the fall, with additional opportunities to sustain that learning during the school year through ongoing leadership meetings. At least one district is considering contracting with Solution Tree to provide broader essential-standards training for additional staff as the pilot enters Year 2.

In Year 1, two of the five participating districts had district office staff attend coaching sessions led by Solution Tree alongside school-based teams, signaling a shift toward more collaborative relationships with schools. District administrators in these districts noted that, for the model to work, they needed to understand it firsthand—even though some had never worked in or led PLC at Work schools. One administrator explained why he prioritized attending coaching sessions, given his prior experience as a teacher and administrator in a PLC school:

Yeah, I think the challenge we have is that most of us who are in this system have been raised educationally in a model where we haven't had a healthy collaboration, we just haven't had it. I attempted as a World History PLC. And so I was like the blind leading the blind, I'm reading the book trying to figure it out. And so most of us haven't had that experience or most can't really lead it from a place of saying, "I've done this work as a teacher as a PLC lead." So most of us are trying to figure it out... If I can go back and do

it all over again, what I would have done is I would have been in there learning with teachers and I would have embedded myself in probably a grade level team. Or maybe you float around between grade level teams and see elements of different parts of the process with different grade levels and an understanding because what I relied on is the principal then debriefing with me, and I said, Okay, so what are you guys doing? Like what's the process they follow?

Building a shared, aligned understanding of the model is critical for any change effort—especially given that pilot support lasts only three years. More frequent on-site engagement helps build the capacity of district staff who will be responsible for leading the work going forward and supporting its spread to other schools.

Fillmore Unified has decided to extend the work to all schools in the district, but giving each school autonomy over technical decisions—such as which standards are essential or what content is included on common formative assessments—has made coherent district support more difficult. District benchmarks are not consistently aligned with school-level CFAs, and schools are using pacing guides differently based on site-level decisions. One district leader described the challenge of coordinating support across schools that are undertaking the same work autonomously, without regular communication with one another:

We look at our CAASPP data coming in trying to identify what happened in that classroom, or what happened at that site, or what happened at that grade level, can we identify what it was that made that group move, versus this other group who either has not grown or, you know, or limited growth. And I think that as for me, when I'm looking at that big picture data, there's so much to consider, there are so many pieces, because I gotta go back now and look at all the essential standards [each school] selected. So I have to look at the essential standards that they picked versus this [school], and then I'd have to really go back down and identify their CFAs and look, what kinds of assessments did they have? What were the scores on those assessments that led to that success or not? And then you've got the cohort data between the kids coming in from the previous year, like so, for me at the district level perspective, I'm still living in that kind of, where am I going now? Because the amount of things I need to analyze is significant, where at least if we're all on the same page, and we're on the same train moving forward, I can say, Hey, okay, how did you teach this standard? Because I know what everyone taught. And I know, we all agreed on [these standards]... I can go straight to instructional strategies, and see what was taught. But now I'm having to go through and filter, the assessment, the standard, all the pieces to determine where they got the bang for the buck.

The district is attempting to ensure coherence with regular leadership check-ins, a structure similar to a district-level guiding coalition, to coordinate across the school sites. However, this structure doesn't include all of the external technical assistance providers – at the initiation of the project there were 19 different coaches working in schools across the district – suggesting that coherence will be difficult to achieve even with district leadership bringing site leaders together regularly.

While some districts intentionally pushed in to support work at sites other teams expressed confusion over what the district role was in supporting this work. While district leads provided autonomy for teachers to make decisions around the use of instructional resources there was confusion over what support the district would provide to facilitate the model. Several of the pilot schools struggled with staffing for classrooms full time (e.g. the fourth grade class was a series of long term substitute teachers) and to cover for teachers so that they could collaborate. Schools that did not have any specials (e.g. PE, music, art, ect) struggled to provide a consistent time for teachers to meet. The lack of support made implementation of the PLC at Work model unsustainable with one principal stating, “there was a point I wanted to quit” and when they asked for sub support they were “slapped on the wrist” by the district office. There was no clarity in the district role for supporting implementation of the technical and adaptive changes necessary to support the school. To the point that district staff, including the superintendent, did not appear to know what the school was experiencing or that resources were needed to support their work. Clear communication of the expectations for teachers, principals, and district admin to collaborate on the work will be needed for the models to scale across entire district systems.

Mindsets about data use for inquiry

The third adaptive shift centered on reframing data as a tool for improvement and reflection, rather than evaluation or blame. In the first year of implementation, teams were only beginning to experience the routine of using data consistently to inform improvement efforts. Interviewees in several districts noted the need to build a productive data-use culture in schools where trust was low among district, school leadership, and classroom staff.

At one school, school leaders prioritized establishing a shared process for reviewing student learning in Year 1, with the explicit goal of building trust. They aimed to help teachers become more comfortable examining data with candor and then using it to adjust instruction and activate additional supports as needed. These efforts began to pay off by the second year. School administrators described how data monitoring and analysis increasingly informed targeted interventions, with some teachers sharing intervention responsibilities across

classrooms, regrouping students based on need and on teachers' demonstrated success teaching specific essential standards. Leaders also noted that teachers became more targeted in supporting students who were far below grade level, using essential standards from prior grades to guide interventions that addressed missing foundational skills rather than repeatedly drilling on-grade-level content that students could not yet access. Over time, data became the lens through which teachers planned instruction, intervened, extended learning, and celebrated progress. Teachers were not penalized for lower scores; instead, staff increasingly understood themselves as collectively responsible for identifying and closing learning gaps.

District offices similarly sought to cultivate more transparent, trusting data-use cultures with the schools they supported. In one district, the district administrators gave principals autonomy to enact the PLC at Work model and make adjustments as needed. However, that autonomy was coupled with monitoring and feedback on implementation quality. The district administrator leading the work described the challenge of maintaining this balance:

Do we need to push it a little bit more? Do we need to pull back? Or where can we monitor this progress? And it's by allowing principals to have that space and then having them come together during leadership [meetings] and share out their progress? It gives us a quick check, almost like our formative assessments write a quick check of are they – are we – on the right pathway? Or do we need to pull back a little bit? It also puts that little positive pressure on principals as they're sharing out so that they are all responsible for moving this forward and not being a sideline complainer...making sure that they're in that process of making this successful and then creating a safe space for them where that if they're vulnerable, and they say, "I don't know how to do this"...Again, remember, we're emulating what they're expecting their teachers to do during PLCs during collaboration time, "Show me how you're successful because it wasn't successful in my class," or "Hey, guys, it's successful in my class. Let me show you how."

Encouraging principals to take up a new process while simultaneously establishing a new culture of vulnerability has been difficult for some school site staff as well as at other schools in the district that the work is being spread to. Nonetheless, district leadership was committed to emulating the type of vulnerable leadership they want principals to lead at their sites. An example of this were the common data reports that the district generated and shared with principals to support conversations about the impact of the model on teaching and learning at each district leadership meeting.

Tiered interventions based on student need

Most schools understood the PLC work in Year 1 as primarily focused on Tier 2 instruction. All but one school described the work as building the “pieces” necessary for the model to function—essential standards, CFAs, data-analysis protocols, and structures for collaborative PLC time—in order to intervene effectively with students. One school, however, conceptualized the work as a systematic process for redefining how teachers plan and deliver high-quality Tier 1 instruction. The principal described how the PLC process was changing how teachers approached instruction:

We have really reinvented how we plan instruction, and it's really impacting the delivery of effective instruction as well...Every team does a unit plan similar to this for...math and literacy. And we're very intentional about unpacking our standards, identifying the skills that students have to be able to do at the end of the unit, identifying supporting standards, developing learning targets, and then creating assessments that are aligned to the rigor of the standard. And then we our teachers have been getting really really good at making the the teaching visible, making the how you teach it visible through unit planning...For every day of the week during that unit, they are very aligned and consistent throughout the grade level to be teaching the same thing, and having autonomy with how they're going to teach it... Looking at instruction together, looking at effective instruction effects, effective instructional resources, developing common assessments, and looking at the data together, and then in responding to that data...For the essential standards they are going through the cycles of inquiry, using this tool....that's for their unit, and then then they analyze the data...They discuss after each common assessment, skills that proficient students demonstrated skills that students that are not proficient needed to demonstrate and how they can help those students and who they are...Each grade in each class, and then they reflect on their unit so that by the same time next year, as they're preparing that same unit, they have some reflection on how to improve that unit for the next time they teach that unit. And so this has been a game changer for us, we're seeing some positive growth in results.

Teachers at this school moved much more quickly into developing unit plans, common formative assessments, and data-analysis processes. By the end of Year 1, they had completed most units in both ELA and math and had also begun that work for dual-immersion Spanish ELA and math. Most other schools began developing unit plans in the second semester, and few expressed strong confidence in their processes for analyzing student data. Reframing the work as strengthening high-quality Tier 1 instruction—rather than primarily as a process for intervention—offered a different lens on collaborative culture at this site compared to others.

Notably, this perception of the work was not shared by other principals or district administrators supporting implementation at other pilot sites.

Qualitative Analysis of Year 3

After three years of support from Solution Tree through the IAM, we revisited participating districts to assess their progress in developing and sustaining intensive PLCs at Work. This Year 3 follow-up examined the extent to which districts had advanced along the conceptual framework's technical and adaptive dimensions—strengthening core technical practices (e.g., mastering essential standards, designing common assessments, and analyzing evidence of student learning to inform instruction) alongside adaptive shifts (e.g., building collective responsibility for all students' learning, establishing norms and roles that support disciplined collaborative practice, and reframing data use as inquiry rather than judgment).

Anchored in the four guiding questions and reinforced through common formative assessments, the PLC process of sharing and reviewing evidence of student learning creates a clear throughline from standards to instruction to evidence to response. Through PLCs, student learning becomes something educators can observe, discuss, and act on together. As one leader put it, the “real difference” is not will or effort, but the presence of a coherent process embedded within a system for improvement:

In the absence of a system for what you should be doing during collaborative time, you are griping because you need an outlet, and no judgment because I have those times too. You're griping and spending time on talking about back to school and maybe spend 45 minutes on that, when really it just needs to be like a five minute conversation. I think that that's the difference and what I was sharing and keep sharing about how teachers just need a system for it. It's not that they can't do it or don't want to do it. They need a system. That's the real difference. There's a system. There are the four questions and if you do these things, you're going to see growth and student results.

We see evidence of these shifts in practice for both technical and adaptive changes in the continuum data as well (Table 7). The indicators for establishing the why of the work and setting expectations for adult collaboration were at or approaching levels of sustaining in all participating schools. We similarly observed the continued development of essential standards (increasing from 1.86 to 4.71) to approaching the level of sustaining. The remaining technical changes – common formative assessments (4.57) and analyzing data (4.71) – reached levels of

implementation that also approached sustaining levels by the end of year three. Response to intervention was unsurprisingly the indicator that was the least developed by the end of the IAM with a score of 4. This is unsurprising given most of the schools only began implementing this aspect of the model midway through year two of the IAM. In the following section we detail the technical and adaptive shifts we observed over the course of the three year IAM.

Table 7: Full Aggregated District Change Continuum Data

Rubric #	Indicator	Year 1			Year 2			Year 3		
		Start	Middle	End	Start	Middle	End	Start	Middle	End
1	<i>Clear Purpose</i>	2.14	3.57	4.00	4.00	4.14	4.43	4.43	4.60	4.86
2	<i>Communicate Purpose</i>	2.43	3.29	4.00	4.00	4.29	4.71	4.86	5.00	5.00
3	<i>Shared Mission</i>	1.86	3.14	3.57	3.43	4.14	4.43	4.43	4.60	4.86
4	<i>Shared Vision</i>	1.86	3.14	3.86	3.86	3.57	4.00	4.57	4.80	4.71
5	<i>Collective Commitments</i>	1.86	2.71	3.29	3.14	3.57	4.29	4.29	4.60	4.71
6	<i>Common School Goals</i>	1.71	2.86	2.86	2.71	3.43	3.86	4.00	4.40	4.43
7	<i>Interdependent Teams</i>	2.57	3.57	3.71	3.86	4.29	4.29	4.29	4.60	4.71
8	<i>Collaborative Norms</i>	1.86	3.00	3.14	3.29	3.86	4.00	4.14	4.60	4.71
9	<i>Smart Goals</i>	1.43	2.29	2.29	2.29	3.43	3.86	3.86	4.00	4.29
10	<i>Determine Essential Learning</i>	1.86	3.29	3.29	3.71	4.14	4.57	4.71	4.80	4.71
11	<i>Analyze Data</i>	2.00	2.43	2.57	3.00	3.71	4.14	4.14	4.40	4.71
12	<i>Clarify Proficiency</i>	1.43	2.29	2.57	3.14	3.57	4.00	4.00	4.20	4.57
13	<i>Common Formative Assessments</i>	1.71	2.57	2.86	3.29	3.57	4.14	4.14	4.40	4.57
14	<i>Response-to-Intervention</i>	1.57	1.86	2.57	3.00	3.43	3.71	3.71	3.80	4.00

Rating	Pre-Initiating	Initiating	Implementing	Developing	Sustaining
Score	1	2	3	4	5

The Year 3 findings that follow illustrate how the IAM developed this shared system through technical and adaptive changes that transformed the quality of collaboration and culture among school staff, with indications of positive effects on student learning experiences.

Mastery of essential standards

By Year 3, IAM teachers had moved beyond initial implementation to develop stronger mastery of essential standards—a key technical shift in the work. Educators described a clear move from “covering the curriculum” to using curriculum materials as flexible resources rather than the instructional goal. This shift required greater understanding of how to identify and deconstruct essential standards, planning discipline, and instructional judgment: unpacking standards into learning targets, designing assessments and lessons backward from end goals, and making real-time adjustments while staying anchored to shared expectations. Over multiple years, essential standards became the “tight” core that enabled stronger coherence and collaboration without sacrificing teacher autonomy in how instruction was delivered.

One principal captured how deeply this shift challenged long-standing routines organized around pacing guides and page-by-page coverage. Reflecting on decades of instruction shaped by calendars and textbooks, the principal described it this way: “For years, for my entire career, it was always, ‘Here’s your book, teach it. Follow this calendar.’ Then it became, ‘It’s not a pacing guide, it’s a pacing calendar,’ but the focus was still the same—get through page 36 by April 25, page 99 by May 2.” In contrast, the principal emphasized that the work is breaking larger standards into meaningful learning targets and ensuring students reach proficiency by year’s end:

It’s about essential standards — really living in those standards and figuring out what students need to do, the learning targets within those big, ginormous standards, so that by the end of the year, they’re proficient. Honestly, it’s kind of embarrassing to admit that it’s taken me 27 years in education to really get it. But I finally understand standards — I finally get it.

Educators in one district described a parallel evolution, articulating a planning discipline aligned with standards-based mastery: beginning with the standard, unpacking what it requires, identifying the end goal, and building assessments and lessons backward from that target. In this framing, the shift also signals growing teacher expertise, because using curriculum as a flexible resource—rather than a script—requires educators to make high-skill, real-time instructional decisions. Teachers demonstrate this nimble expertise when they can interpret the demands of a standard, identify the component skills and concepts students must master, and then choose—or adapt—materials that best advance that learning.

Instead of defaulting to “what’s next in the book,” they diagnose where students are, anticipate likely misunderstandings, and adjust tasks, pacing, and supports while staying anchored to the same learning goal. The standard remains the fixed point, reflecting teachers’ developing professional judgment to navigate multiple resources, calibrate rigor, and maintain instructional coherence even as circumstances (student needs, time, or materials) change:

Before, everything was curriculum-driven. You had to start on this page and finish on that page. Now we’re looking at the curriculum through a completely different lens. The curriculum is a resource, but the standard is what we focus on. We start with the standard, unpack it, talk about what it really requires, and then decide which of our resources will actually help us teach toward that standard. We also think about the end goal—what do we want students to be able to do by the end of the unit? Creating our assessments first helps us break that down. As we plan lessons, we know exactly what we need to teach for students to reach each step. It becomes, “I need to teach this so they can get here, and this so they can get there.” We’re giving them those small pieces along the way to make the final assessment both successful and attainable. Ultimately, the goal is for students to learn the standard—and this approach really supports that.

One teacher described the focus on essential standards as “very liberating for my brain, and I know for our teachers too... it’s helped us really prioritize things that are going to help kids in the next grade level.” At the same time, teachers described the standards work as a genuinely difficult professional learning curve—one that required them to unlearn familiar routines and develop a new kind of instructional expertise. Early on, “deconstructing the standard” was not common language or practice for many teams.

There was anxiety because everything was unfamiliar. We didn’t yet understand why we were doing certain things. It felt like we were being asked to rebuild everything we thought we knew about planning and instruction. But once we actually understood the process—once we learned how to deconstruct standards, create meaningful learning targets, pace units intentionally, and use assessments the right way—everything changed. Our anxiety went down. Our confidence grew. And seeing the results with our students made a huge difference. So the hardest part was really the beginning: breaking old habits, adjusting team dynamics, getting everyone on board, and learning a completely new system while still teaching full-time. But now that we’re all aligned, everything is so much easier. We got over the hump, and now we can really move forward.

Developing a strong grasp of essential standards—and translating them into day-to-day teaching practice—is a gradual process that unfolds over years, not something that happens

immediately: “It’s not like we mastered it in the first year—honestly, it took a couple of years before I felt confident about what I was doing with the standards.” This teacher went on to say:

Dissecting the standards helped us really pinpoint what students need to master and which skills they need to develop to get there. That clarity made a big difference. The way the unit plans were structured also helped a lot. The deconstruction of the standards helped us focus on what the standard is actually asking—without getting distracted by other skills that aren’t part of it. Just staying centered on the standard itself.

Teachers described essential standards as the anchor that helps teams distinguish what must be consistent across classrooms and what can remain flexible. Once standards are clear, it becomes easier to determine which elements of instruction need to be “tight” to support equity and coherence (learning goals, targets, and expected outcomes) and which can remain “loose” to preserve teacher expertise and responsiveness (materials, strategies, and instructional style):

Alignment doesn’t mean every classroom looks exactly the same. Walking into one class and into another class shouldn’t feel identical. Alignment means we’re all sticking to the standard—we’re all moving toward the same learning goal. But each teacher still has room for their own style, their own resources, their own strategies. There’s a “tight” part—the standard, the learning target, the outcome—and then there’s a “loose” part, which is how you teach it. Sharing has become a huge part of that. We link everything to the unit plan so everyone has access. We have shared resource folders, and at the beginning of a unit, we each pick a strategy we’re going to use and share it with the team. Anyone can use anything that’s posted. That wasn’t the case in the beginning, and we’re still building that muscle, but with practice, we’re adding more into the planner and building stronger alignment—while still keeping the freedom to teach in the way that works best for us.

Understanding standards “created the clarity that made alignment possible—not as uniformity, but as shared direction.” That shared clarity improved coordination and made collaboration more seamless without forcing every classroom to look identical. Because the standard defines the tight core, teachers can share and adapt resources more freely without losing coherence. The teacher described how shared unit plans and resource folders support this approach: teachers contribute strategies, others adopt them as needed, and the team builds a collective repertoire over time.

Designing common assessments

A consistent technical shift across districts was the routine design and regular use of common assessments at the grade level, which became a major lever for strengthening instruction. Teachers described a move from treating assessment as something appended at the end of a unit to using it as the starting point for planning. This backward-design approach helped teams clarify intended outcomes, stay anchored to essential standards, and make instruction more intentional.

In one district, a teacher explained that teams are increasingly building the assessment before lesson planning, using the standard to define the outcome and the learning targets that must be taught along the way. Rather than “going and then trying to make an assessment at the end,” the team begins by identifying the standard, naming the targets, and creating an assessment that reflects what students should be able to do, then planning lessons that deliberately build toward that end goal:

I feel like more teachers are getting into the habit of creating assessments, at least in our grade level, creating assessments before even lesson planning for what that unit might look like. So I think that's very essential, aside from just looking at at our standards, is finding what is the outcome we want from this, creating this an assessment for that, and then getting it done right and so then that allow allows us to work towards that end goal, yeah, rather than just kind of going and then trying to make an assessment at the end, right? So I think finding the standard, identifying targets and making an assessment were like vital and important to this process. I think that creating the assessments before, while we're planning, I think it helps drive the lessons, and it makes the teaching a lot more intentional, and we just stick to what the standard says.

Teachers also described a more mature understanding of what makes an assessment meaningful. In Sacramento, one educator noted becoming more discerning about rigor, alignment, and the essential skills being measured. Before PLCs, assessments were often inherited from curriculum materials with limited teacher input; now, teachers felt empowered to craft assessments that reflect what they most want students to learn—and then use the results to guide interventions:

I think that I am more critical in the way that I use assessment. I want the assessment to be high rigor and also connect to essential skills and also making sure that it's common core aligned. I think prior to adopting the PLC model, most of the assessment that we used was from whatever the district adopted curriculum was, we don't really have the autonomy, or we don't feel liberated to create our own assessment. Now we have the

opportunity where we get to make assessments that really drive the essential student learning that we want to see within the classroom. So whatever that we're teaching is tied to the essential skill, then we can use the results from the assessment to gauge how do we use that to inform our interventions? So it provides a lot of opportunity for us to use assessment in multiple ways.

Coaching and shared practice helped teachers develop assessment design as a professional skill—not simply generating more questions, but crafting tighter items that reveal more about student thinking. As one teacher reflected, learning that “less is more” improved their ability to write targeted prompts that surface specific misconceptions and guide next instructional steps: “We can create one specific question that can tease out a lot of evidence of student learning or what students are lacking... The coaching we got on creating assessments was very beneficial. And I think that’s one skill set that most teachers probably lack and could benefit from.”

Teachers also described a mindset shift in how assessment data are interpreted and used. Instead of viewing tests as a judgment of students, they increasingly treated results as feedback on instruction—information that helps identify gaps and adjust approaches. This framing supported student confidence and persistence by positioning assessment as a tool for learning rather than a label: “Now I really see tests differently. They’re not a reflection of the student—they’re a reflection of what I need to work on with them and what gaps I need to help close.”

Finally, common assessments served as the glue that made collaboration coherent across classrooms. Teachers emphasized that when teams teach toward the same essential standards and use the same assessments, they can meaningfully compare results, share strategies, and maintain alignment in pacing and expectations. One educator contrasted earlier efforts—where teams drifted apart even with a pacing guide—with the coherence created by shared standards, CFAs, and structured collaboration: “This time, we’re all teaching toward the same essential standards, using the same assessments, and sharing ideas. Before, even with a pacing guide, teams were scattered—some far behind, some far ahead—so CFAs weren’t aligned and everything felt inconsistent. Now, the structure keeps us together.”

Together, these reflections illustrate how designing common assessments—especially when planned before lessons—became both a technical skill and a cultural shift. It strengthened alignment to essential standards, increased instructional intentionality, improved the use of data to guide intervention, and helped teachers view assessment as a tool for continuous improvement rather than compliance.

Shared system for reviewing student learning

In Year 3, a third technical shift took hold across sites: collaboration became anchored in a shared system for reviewing student learning. Earlier efforts often created time for teams to meet, but meetings varied widely in focus and quality. By Year 3, educators described a shift from “meeting to meet” collaboration to a predictable routine in which student evidence is consistently brought forward, examined using common prompts, and used to determine concrete instructional responses. This section describes how that shared “operating system,” often organized around the four PLC questions and supported by clear roles, agendas, and aligned tools, turned collaborative time into a more reliable mechanism for instructional improvement.

What made the difference was not simply having collaborative time, but establishing a predictable process in which student evidence is consistently brought to the table and discussed in a structured way. Teachers described a shift from loosely defined collaboration—often focused on school business, venting, or “compliance documents”—to the reliable execution of a shared process with a clear purpose, common prompts, and a predictable flow. One educator emphasized that the weekly practice of putting student work in front of the team is what transforms collaboration into meaningful instructional improvement: “Every week, there’s student work on the table. Everyone can literally put their eyes over it and ask, What does it look like for the kids? What are they doing? That’s the secret sauce in this work.” This routine keeps the focus anchored in what students are producing and learning—not in intentions, plans, or adult perceptions. Educators repeatedly pointed to the four PLC questions as the mechanism that makes collaborative work coherent, actionable, and sustained. One school leader described using the four questions as the clearest indicator of whether a school is truly “doing the work”:

I would look for those four essential questions. What do you want all students to know? How will we know if they have learned it? What do you do when they haven’t learned it? What if they have? If those four questions are being addressed, then I would know that they’re doing the right work.

Another school leader contrasted prior “collaborative time” with current PLCs by naming the specific elements that changed:

Prior to PLCs, we had collaborative time, but the kinds of work that was happening in those collaborative times was not the same kind of work that is happening now. The main difference is the intentionality, the eye on data, and the focus on only about those four questions. Teachers have shared that before PLCs, they would complete some sort of compliancy type of document to show what they had been doing in their collaborative

time. But I don't mandate anything that happens in those meetings other than you're only focused on those four questions, right? And they have agendas and I have access to all of those. And we communicate on the agendas in Google Docs. So if they have any questions for me, they'll note me in the document, and I can pull it up and give my two cents, or go and visit a meeting if I need to. But it's just really that intentionality and all within the parameters of those four questions. So as long as you are using those as your map, every meeting is going to be intentional and meaningful if you're focused on those four questions.

This approach did not require heavy-handed micromanagement. The leader described setting a simple non-negotiable—stay anchored to the four questions—while giving teams autonomy to run their meetings within that structure. Shared agendas and transparent documentation supported light-touch coaching and quick support when needed.

One teacher describes how the way they use their PLC time has changed to focus on student learning:

We've gone from one end of the spectrum to the other. In the beginning, our PLC meetings were all over the place. We spent half the time talking about recess duty or field trips, or we'd end up venting and going off-topic. When I look back at those old agendas, it's almost funny compared to where we are now. Today our agendas are completely focused. We're on top of things. We look at our data. We talk about what's working. We ask, "How are we going to move these kids?" We share strategies. We manage our time much better. And all three members of our fourth-grade team have stepped up—we hold each other accountable. We allow maybe two minutes of venting, and then it's, "Okay, let's move on." We're excited to share what we've tried to help our struggling students. We were never like that before.

Teachers described a key shift in their collaborative culture: developing a shared, consistent process for bringing evidence of student learning into the room—and using it to align work across classrooms. Rather than treating PLC time as informal planning, teams began to rely on common tools and routine data review to stay focused on what students needed and to reduce confusion and disagreement about what to teach and when. As one teacher explained, "Our meetings are more focused. We know where we're going because we have unit planners, pacing maps, and proficiency maps that really guide us. We use data constantly, and that keeps us centered on what students need."

Technical change requires new skills. Learning to collaborate through a disciplined process to review student evidence meant teachers had to build capacities many had never been explicitly

taught: interpreting evidence together, asking the right questions, identifying patterns, and translating findings into instructional adjustments. As a result, technical change depends on supports that build real competence—clear models and exemplars of what “good” looks like, agreed-upon norms and roles for collaboration, scaffolded protocols that make the work repeatable, opportunities to practice across repeated cycles, and coaching and feedback that help teams refine quality rather than simply complete steps. It also depends on enabling conditions, such as protected time and usable data systems, so these skills can be applied consistently in regular meetings. Without these supports, technical reforms can slide into surface compliance without strengthening teachers’ capabilities to collaborate in this way. As one teacher explained,

The collaboration—the way we work together now—is dramatically different. Our grade level honestly didn’t know how to collaborate before. And to be fair, we were never really taught... What helped us learn to collaborate was [our coach] being in our meetings and showing us what real collaboration looks like. When we first started PLCs, our mindset was totally different. We felt like, “This is taking too much time. I could be working in my classroom.” Now we collaborate even outside PLC time—after school, through group texts, in the mornings. I think a big part of it was just being shown how to do it. We were pretty closed off at first, but once someone modeled the structure and purpose for us, we opened up. Now our conversations are completely focused on student learning. We’ve been given guiding questions about our data—Which students struggled? Which need enrichment? What do we need to adjust or reteach?—and we use those every time. We get 110 minutes to collaborate, and we divide that time intentionally based on the data we’re reviewing. We also take roles now. I’ll say, “Okay, who’s the timekeeper? Who’s the note taker?” Before, we never held each other accountable like that. Now we have a structure, and we stick to it. We come prepared. We know the questions by heart. We get right into: “These students need help. These students are ready to go deeper.” We already have our data pulled, and we use the entire time efficiently. The difference is night and day.

As this teacher described, the main lever for collaboration was not a new mandate, but modeling and coaching. The teacher attributed their growth to having someone demonstrate what productive collaboration looks like in practice. Once the structure and purpose were modeled, teachers became more open, shifted their mindset about the value of PLC time, and began collaborating even beyond scheduled meetings. Building strong PLCs requires more than protecting time on the calendar. Engaging productively in collaborative processes to review student evidence is a technical skill that must be taught and coached explicitly so that

collaborative time consistently translates into improved instruction and stronger student learning.

In addition to building collaborative skill, it was equally important to develop the infrastructure needed to review student data consistently and reliably. Across districts, teachers emphasized that meaningful collaboration did not emerge simply because time was scheduled; it emerged because teams adopted a clear structure, a shared process, and defined roles that kept meetings focused on student learning. Without that architecture, “collaboration” often defaulted to conversations that did not reliably change instruction.

At one site, teachers described how Wednesday meetings—which previously functioned largely as logistical coordination—became a fundamentally different kind of meeting, with predictable routines and clear roles: “Each grade level follows the same structure: we start and end at the same time, we have facilitators and note-takers, and we spend our time looking at data and student test scores.” That shift was reinforced by coherent assessment practices across grades: “From kindergarten through sixth grade, we all use the same common formative assessments and summatives... if we’re not looking at data, we’re creating or revising those assessments.” In other words, structure was not only about meeting norms; it was reinforced by shared tools that made collaboration possible and aligned.

Finally, the most consistent message across sites about what underpins successful, data-based collaboration was the need for adequate time. Protected time is a non-negotiable foundation on which teachers build disciplined routines for examining evidence, planning responses, and coordinating interventions. As one principal put it, the bedrock of the work is having collaboration time that is truly protected:

You need to have a built in time that is almost biblical in a sense that it does not move this can't be infringed upon for whatever reason. I don't care if, if the school was burning down. You still collaboration time, so it just can't be infringed upon it. I don't know how any other way to put it, how serious I am about that time.

Strong PLCs depend on more than goodwill and a positive intention to collaborate. They require an explicit operating system—consistent meeting structures, shared guiding questions, defined roles, aligned assessments, and adequate time—that enables teachers to use collaborative time to diagnose learning needs, coordinate responses, and steadily improve instruction.

As teachers strengthened the technical side of the work—prioritizing and mastering essential standards, routinely reviewing evidence of student learning to guide instructional adjustments, and co-designing common assessments and lessons—they also began to encounter a different kind of change. Implementing new tools and routines required teachers to surface and rethink

underlying beliefs about what students can learn, redefine roles and expectations for collective responsibility, and establish norms for using data as a source of inquiry rather than judgment. The next section focuses on these adaptive shifts: changes in mindsets, roles, and relational practices that help technical improvements take hold and translate into sustained improvement.

Commitment to all students learning

By Year 3, the most foundational adaptive shift was a deepened commitment to all students learning. Across sites, teachers and leaders described a move away from “teaching to the middle” toward embracing responsibility for students across the full performance spectrum—especially those who have been historically underserved. This was not simply a new set of expectations; it was a mindset reset grounded in moral purpose, reconnecting educators to what students deserve, what drew many teachers to the profession, and what educators can accomplish when they work together. As this commitment took hold, teachers became more willing to challenge deficit beliefs, persist through frustration, and view student success as a shared responsibility—shifting from “my kids” to “our kids”—which infused the technical PLC routines with greater purpose and meaning.

Teachers and leaders described this shift as an increased willingness to plan intentionally for students on the margins, not only those who are most engaged or easiest to teach. For example, one school leader observed, “I’ve seen culturally, us begin to embrace all students, as opposed to just the middle of the road students that we typically teach to.” Another noted that a sense of “those kids just can’t get it” used to be more common, “but it isn’t anymore. And now, teachers are more willing to call out and challenge those limiting mindsets.”

As teams expanded their focus beyond the “middle-of-the-road” student, teachers described a culture that positioned improvement as a shared commitment to what every student is entitled to: “It’s not about finger-pointing or mandates, it’s about asking, ‘What do our kids deserve?’ It reminds us that we’re educators, we’re capable, and we can make the changes we need when we have clear expectations and the flexibility to do what’s best for our students.” Broadening the focus beyond the middle also included deliberate plans for students who were performing below and above grade level. For higher-performing students, this sometimes meant access to learning experiences in other grade levels. For example, as a kindergarten teacher explained:

We have a few kindergarteners who can already read this year, and when we look at what students need, we also have children who need enrichment and extension. This is our second year sending a kindergartener to small group time in second grade... Overall, we’re trying to adjust supports based on student needs, including providing extra

extension for students who have mastered certain skills so they stay challenged and engaged.

Adaptive changes to the commitment to all students' learning deepened when schools treated the work as more than a set of new routines and instead started by reconnecting teachers to the moral purpose of teaching. Leaders described intentionally slowing down at the outset, creating space for staff to revisit mission and vision, and reflect on identity. As one school leader recollected:

In the first year and a half, we really worked on our mission and our vision, like who we are, what do we want to be, what do we want to do? So we just worked on ourselves with changing our mindset. I think that made a difference. Instead of just jumping right into work, you know, instead of just saying, Here's your assignment, create a CFA for this essential standard. So we didn't do that. We just started off with our mindsets.

This kind of collective reflection served as a cultural reset: it acknowledged the reality of burnout and frustration and helped staff name what they still cared about even when day-to-day conditions felt exhausting. By grounding improvement efforts in shared purpose, schools reframed PLC work from “one more assignment” to a values-driven commitment to students. Rather than beginning with technical deliverables—such as immediately building CFAs for an essential standard—teachers first developed a shared understanding of what they were trying to accomplish and why it mattered.

This shift helped move participation from compliance to ownership. The work was not about being monitored or proving engagement in collaboration; it was about honoring what students deserve and what educators are capable of when they have clarity, support, and flexibility. In this frame, the technical tasks (standards, assessments, data routines) became tools in service of a mission—making it easier for teachers to persist through the hard parts, stay engaged when results were uneven, and sustain collective effort over time. For some teachers, a shared mission quickly strengthened their commitment to every student's learning. Others needed more time for that commitment to take hold, but as they engaged in the work and saw results for students, they came to embrace it as well. One teacher shared a story about a colleague who was initially resistant to the PLC approach:

She had her emergent bilinguals, several of them reading zero words per minute at the beginning of the year, and now they're at 120 and it's May, and it blew her mind. She said, “Oh, my God, it's because they are in small groups in this rotation.” So there's the buy in when you see that these kids that we assumed would get it someday down the road, can get it today. That's why we're here. I mean, no matter what, what attitude you

come in with, most teachers are there for the kids. And when you see that your work is actually changing their lives, it's impactful.

Beyond strengthening commitment to students across the full performance spectrum—from those who are struggling most to those ready for enrichment—one of the most significant adaptive shifts was a change in who teachers felt responsible for. Teachers increasingly described student learning as a collective obligation: not only “my students in my classroom,” but all students in the grade level and, ultimately, across the school. This mindset shift reframed collaboration from sharing ideas when convenient to working as a team with shared ownership for outcomes—so that supporting any student became a shared responsibility, not the work of an individual teacher alone. As one teacher put it, “As the process went along, and even after the first year, we were like, ‘Hey, we are in it for the grade level, not just our class. We’re in it for the grade level and essentially the school.’”

Importantly, the visibility of growth became one of the strongest incentives to sustain the work. The same leader emphasized that stipends may help, but what truly kept teachers engaged was seeing student learning change over time:

There are the four questions, and if you do these things, you're going to see growth and student results. And that's really like the greatest incentive to all of this. It is not the per diem, even though that's nice, it's the growth our kids, who were not reading at all, have made in a school year. It's the growth that our students are making between every CFA. Teachers are really incentivized by student growth.

By building in regular cycles of evidence and response, PLCs allow teachers to experience progress as real and measurable, especially for students who previously struggled, reinforcing the belief among teachers that the work is worth the effort.

Culture of adult collaboration

A second adaptive shift was the development of a more interdependent adult culture—teachers moved from working in isolation to relying on one another as a normal, expected part of solving instructional problems. This section highlights how shifts in vulnerability, shared responsibility, and leadership roles helped collaboration become embedded as “how we do things here,” sustained less by top-down direction and more by peer norms and collective ownership.

Teachers described a marked shift from isolation to interdependence over the course of the three years. They moved from a culture where collaboration—and even social connection—was limited to one in which relying on one another to solve instructional problems became the

norm. One teacher recalled that when they arrived, “everybody was doing their own thing,” and the separation felt so entrenched that “nobody wasn’t even eating lunch together... everybody was in their silo.” This starting point underscores that the change was not only procedural; it required new relational norms and a different level of trust among adults.

Over time, teachers described collaboration as a space where vulnerability became acceptable—and even expected—as part of improving practice. As one teacher explained, the shift also included changing how they talked about the time:

I think one of the biggest changes has been around collaboration—especially the language we use. Sometimes we don’t even say, “Oh, we have PLC today.” Instead, we say, “We have collaboration time,” and that small shift really changes the mindset. It becomes a time when we can talk with each other, be vulnerable, and say, “I’m struggling with this,” and then create opportunities to learn from one another. It’s not always perfect, but we try to use that time to share instructional strategies and support each other.

Another teacher emphasized that this kind of openness is not automatic—especially for newer or quieter teachers—but that the culture evolved so the team became “comfortable enough with each other to say, ‘Hey, I need help.’” In the same reflection, they captured the norm shift succinctly: “We’re a team. No one is doing this alone. We’re collaborating, supporting each other, and learning together.”

Over time, teachers described collaboration shifting from simply meeting together to operating as an interdependent team—where expertise is shared, support is expected, and responsibility for students is collective. As one teacher reflected on how this culture developed over multiple years: “Over these three years, we’ve learned how to use each other’s strengths, lean on each other’s knowledge, and truly function as a team because, again, they’re our kids.” These accounts suggest an adaptive shift in what collaboration means: not just meeting together, but building a climate in which asking for support is seen as professional, admitting difficulty is safe, and learning is collective because the team shares the same goals for the same students. In that environment, collaborative time becomes more honest and growth-oriented, with teachers drawing on one another’s strengths to better meet students’ needs.

One indicator that collaboration had become “how we do things here” was that participation was no longer driven primarily by top-down expectations; it was sustained by peer norms and shared ownership. As one staff member explained, the shift felt less like compliance and more like a culture that reinforces itself:

It works, because it's not like the principal is saying, 'You have to do this.' It's more like, everyone else is doing it, and if you're the only one not doing it, you kind of want to. It's that indirect peer pressure, I guess. It's becoming the [School] way—baked into the culture.

Teachers emphasized that the cultural shift was not confined to individual PLC teams—it spread across the school, creating a shared sense of purpose, common routines, and even a common language for talking about student needs. As collaboration became more consistent, staff began to see that “this is what we do” across the entire school, so that learning from other grade levels and discussing data felt normal and shared. As one teacher described:

What's interesting is that the unity goes beyond individual teams. As a whole school, we've gone through the same struggles—being frustrated, confused, stressed, skeptical—and now we're on the same page. In the beginning, it felt like, “Here we go, another new thing,” and people didn't take it seriously. But now everyone knows: This is what we do. And hearing other grade levels share their ideas for intervention or strategies—we're learning from each other. During staff meetings, when [our principal] brings up data or intervention groups, we all immediately know what she's referring to. We're speaking the same language.

One of the most marked and consistent adaptive changes that emerged across sites was a shift in teachers' sense of responsibility—from “my students in my classroom” to “our students across the team, grade level, and school.” Interviewees described this as more than a semantic change; it reflected a deeper reorientation of beliefs, roles, and norms about what it means to teach in a collaborative improvement model. In many settings, teachers entered the work with an implicit view that student learning was primarily an individual responsibility, shaped by one's own classroom decisions and one's own class list. Over time, PLC routines—especially shared analysis of common assessment evidence and joint planning for instruction and intervention—expanded teachers' sense of accountability for students. As one teacher reflected,

I think, if anything, what I noticed is that we work even closer together, how to come up with some assessment and thinking about, I'm working with your student. You're working with my student, they are “our” student, and seeing the grade team as collectively “our students”, and making sure that we help each other to develop that great learning experiences for each student. So I think that component has been enhanced.

Another teachers' perspective shared the same sentiment:

This is not just one teacher's student. This is our third-grade student. Whether it's me or another teacher on the team, they're vested. We know every student in our grade level. So I think that is the best part, you know, not only for us, because it takes some of that workload off of us, for sure—the planning time and things like that—and then for the students as well, because they know. They come into my class for deployment and they say, "Oh, we're doing that in my class too." So it's like, nothing's hidden. It's all there. I think that's the best, yeah.

Another driver of the shift in adult collaboration was the guiding coalition, which signaled movement toward a distributed leadership model in which leadership was grounded not in positional authority, but in collective professional responsibility. Early on, many schools relied on principals or coaches to set agendas, enforce expectations, and push the work forward. By Year 3, participants described the guiding coalition as a mechanism for teacher-led implementation—a group that not only represented the work, but actively drove it. Rather than waiting for an administrator to prompt next steps, coalition members initiated core instructional-improvement tasks, as one principal described:

Let's just look at our guiding coalition. They lead the work. We're having a guiding coalition meeting, and one of my team members said, "you're not supposed to lead this meeting. It should be the role of the facilitator." So they're holding us accountable to our team roles and norms. Now in the moment, I probably thought that came across a little cutting, but they're not afraid of speaking up. Even when it comes to developing a CFA, there are discussions around that without me having to be the one to initiate it, or examining data in order to form the groups. Again, I don't have to lead that charge. So it's been a beauty to see them take the lead on this work. Amazing. But again, I think it's because of the intentionality and because of the commitment level.

This shift matters because it changes both ownership and sustainability. When the work is principal-driven, collaboration can feel like compliance. When leadership becomes teacher-led, the work is more likely to persist through leadership turnover, shifting priorities, or day-to-day disruptions. The guiding coalition can function as a stabilizing structure that carries norms, protects the integrity of the process, and keeps the focus on student learning even when administrators are not in the room.

One site began the IAM with one principal and then transitioned to a new principal after the first year; staff described that leadership change as challenging for both relationships and the viability of the work: "We had a lot of administrative changes, and so relying on the guiding coalition is important because we had a lot of mistrust." A Solution Tree coach similarly emphasized the coalition's role in sustaining implementation beyond external support: "Once I

step aside, and all the other consultants step aside, the guiding coalition will be key to leading this work... and I believe we will continue to move the work forward.”

Taken together, these shifts suggest that the IAM helped schools build the relational trust, shared norms, and collective responsibility required for disciplined collaborative improvement. By Year 3, collaboration was increasingly embedded as a cultural expectation, enabling teams to coordinate instruction and intervention around shared evidence and a shared commitment to “our” students.

Mindsets about data use for inquiry

Another key adaptive shift was a reorientation in how educators made meaning of evidence. Rather than experiencing data as a judgment of teacher competence—and something to avoid, defend against, or comply with—teams increasingly treated it as shared information for learning: a neutral prompt to surface hard truths, ask better questions about students’ experiences, and decide together what to do next. This mindset shift—making data safe, actionable, and collective—enabled more honest analysis and quicker movement from results to coordinated instructional responses.

Across interviews, participants described the shift to “data as inquiry” as fundamentally a change in mindset: moving away from treating numbers as a verdict on teacher competence and toward using evidence as a prompt for collective learning and next steps. One leader described how this shift in stance toward data began with mindset:

I think it has to start with mindset. You can't move anything if you don't change the mindset about not using data as punishment...don't use data to condemn, use it as information to make your next move.

In that framing, what matters most isn’t simply having student data, but using it to identify what students need and what the team should do differently. One leader described how, in the past, data wasn’t used to shape instruction, it was treated as paperwork rather than a tool for learning:

You'd go in the room and the teacher is just plowing through a lesson and asking and answering their own questions. No one's learning, nor do they care. And data doesn't influence anything. Data is something they put in grade books only.

Leaders described that one of the earliest hurdles was simply having data that was accessible and organized in a way that made it usable for PLC inquiry. But once teams could actually see

timely, relevant information together, it shifted the conversation, revealing gaps that were easy to miss and helping redirect attention to what students needed most:

One of the things that I know was the initial barrier was really just getting data that could be used for PLC conversations. And so one that's amazing that you now have access to information that didn't used to be there at all. Those conversations couldn't happen... you realize, dang, we're not at grade level.

Access to usable data helped enable this belief shift, but it also required changing the emotional and social meaning of data in adult culture. Several participants noted that data can easily become a source of shame or defensiveness, especially early in implementation. Leaders described the importance of normalizing data use for improvement: when students are not making progress, the appropriate response is not blame, but adjustment.

Leaders also described shifting from data as judgment to data as inquiry by using evidence to surface “hard truths” and ask better questions about what students were actually experiencing—identifying mismatches in the system and deciding what needed to change. As one principal explained, looking across multiple indicators revealed a stark contradiction that prompted deeper inquiry about rigor and expectations:

Let's look at our reality when we have 70 percent of our school that are three or more grade levels below in math and ELA as it relates to our state and local indicators, but we have 70 percent of our students who are have a 3.5 GPA or higher, like those two things ...so we had the discussion around grade inflation, around what's happening, around the rigor in our classrooms, and what we're going to do.

These accounts suggest that “data use” is not merely a technical practice (collecting scores, updating trackers), but a cultural shift in meaning-making. The turning point is when educators experience data as something that helps them decide the next move for students, and when the organization signals that seeing a gap is not a failure, but the starting point of improvement.

When data is treated as inquiry rather than judgment, teachers described a clear change in what happens *next*: instead of defending individual practice or explaining results away, teams move more quickly into shared problem-solving. Data becomes a common reference point for prioritizing what matters most and deciding together how to respond—shifting the conversation from “whose fault is this?” to “what do we do now?” One teacher described how the team’s direction was explicitly derived from evidence:

Our vision for our grade level came from our data. I think it's looking at our data and said, “Hey, these are our weaknesses in here. Let's look—what can we do to help support

our students in this way?” And we came up with the plan. I kind of had, like, a rough outline: “This is where our weaknesses are. Let’s see where we can do intervention groups to help them with that.” And they took it. And because we know each other so well now—three years—we can do that, and it came out.

Teams also shifted their approach, treating data as a tool and a starting point for planning concrete, collective responses. This inquiry stance reshaped social norms around accountability. Rather than accountability meaning pressure or exposure, it became mutual support and shared follow-through. Teachers described checking in on one another’s data not to police, but to offer help and keep the work moving. As one teacher described, when she notices a teammate hasn’t uploaded their data:

...even just showing up in their classroom: “Hey, do you need me to help you put in that data?” Or, you know, “What do you need?” And so just in a positive way, I guess. And sometimes they have so much stuff going on—they do need help putting that data in. Or so-and-so’s sick in their family, like, “Hey, do you need me to help you with that?”

In this way, data practices become relational: teams maintain momentum by making it safe to ask for help and normal to receive it.

Leaders also emphasized that treating data as inquiry requires a disciplined improvement orientation: evidence should drive action, not excuses. In one district, a principal reinforced the expectation that results should prompt a solution-focused response: “If it hasn’t improved doing what you’ve always done, then the solution is to try something different.”

Across these examples, the adaptive shift is that data moves from being an outcome used to judge individuals to a shared tool teams use to coordinate action. Data helps teams name priorities, design targeted supports, and sustain follow-through through mutual support. Under this mindset, the organizing question is no longer “Who did poorly?” but “What do students need, and what will we do—together—to meet that need?”

Adjustments to interventions and instruction based on student need

Thus far, this report has traced the technical and adaptive changes that emerged over three years, such as new routines for collaborative inquiry, shared systems for examining student learning, and deeper shifts in mindset, trust, and collective responsibility. In this section, we turn to a proximal outcome of those changes, the extent to which teams began making real-time adjustments to instruction and interventions based on student need. Given the relatively short span of the program and the evaluation window, it is reasonable to expect the

earliest measurable effects to appear in educator behaviors. These include how teachers organize support, respond to evidence, and differentiate instruction. We examine those practice-level adjustments here, and then investigate changes in student learning outcomes in the subsequent section.

As teams developed stronger mastery of essential standards, used evidence of student learning, and collaborated more consistently, teachers described shifting from *collecting* data to *using* data as a practical tool for instructional decision-making. Instead of treating a low score as a general signal that “students didn’t get it,” teams used the evidence to diagnose the specific breakdown. They then adjusted instruction accordingly by reteaching, revising explanations, adding scaffolds, or trying new strategies. Importantly, this work happened collaboratively. Teams compared interpretations, shared what worked in their classrooms, and brought in coaches or administrators when needed. As one teacher described:

If we see that a lot of students are struggling with the same essential standard, the first thing we do is talk through it together. We’ll look closely at the CFA results and ask ourselves, “What exactly did they struggle with? Was it the vocabulary? Was it the multi-step part? Was it the way we taught it?” We’ve gotten much better at breaking the skill apart and figuring out the specific part they didn’t get. Then we start brainstorming strategies. Sometimes we share things we tried in our own classrooms—maybe a graphic organizer worked really well for one of us, or someone found a way to model the problem that finally clicked for their students. We’re a lot more open now, and because we trust each other more, those conversations feel productive instead of intimidating. And if we’ve really tried everything we can think of, that’s when we ask for help. We’ll go to our coaches or to our principal and say, “Okay, here’s what we’ve done. Here’s where we’re stuck. What else can we try?” They’ll suggest scaffolds, strategies, or resources we hadn’t thought of. The nice thing is that now we don’t wait until the end of the unit to do that—we’re responding right away, because we’re looking at data as it comes in. So it’s a cycle: look at the data, identify the gap, try strategies, share what worked, and bring in support if we need it. And over time, we’ve learned that even if it feels like we’re at the end of our rope, someone on the team—or one of our coaches—always has an idea that moves us forward.

Across sites, another observable change in practice was the move toward flexible, intentional grouping of students based on evidence of learning, which many teachers called “deployment” when students were regrouped across classrooms. Teachers described this as a meaningful departure from prior intervention approaches that were either inconsistent or left to an individual teacher to differentiate instruction for their own students. By Year 3, several teams began treating grouping as a shared, grade-level design task: using common evidence to place

students into targeted groups based on evidence of student learning and coordinating who would teach which students during a protected block. Groupings were revisited for different learning targets based on demonstrated student learning needs:

Deployment is when we group students across classrooms. For example, if one teacher has really high results, we ask, “Would you like to extend with the kids who already mastered the material, or would you like to support intervention and share your strategies?” That process—making time for intervention and aligning it across a team—has been a complete game changer. Before, interventions were happening randomly, at different times, with no alignment. Now we have a protected intervention block in our schedule, and we’re really intentional about not letting anything take that time away. Sometimes deployment happens across classrooms, and kids get to learn from another teacher, which is really powerful. Other times, it happens in our own rooms as Tier 2 or Tier 3 support. But whether it’s deployment or interventions within the classroom, the mindset shift is the same: these are our kids, not just “my kids.”

A defining feature of this approach was that students were grouped by learning needs and intentionally matched to teacher strengths. This approach to intervention elevated teacher expertise and made it a resource shared across the grade level. As teams worked together more consistently, they developed a clearer sense of which teachers had particular strengths in specific content areas. They began leveraging those strengths often not for the most advanced students, but for those who needed the most support. In this model, students benefit from access to strong instructional practice. Teachers also benefit from one another’s expertise as strategies circulate across classrooms. As one teacher explained:

At one site, a teacher described a deliberate grade-level decision to group students needing the most intensive support with the strongest math teacher, rather than assigning that teacher to the advanced group:

We believe that we should take the teacher who is the most comfortable teaching math and utilize their expertise in that specific subject area, and that teacher most likely should not be teaching the advanced group. The reasoning behind that is that we believe we could really utilize that teacher's specific skills to target students who really need intense intervention. So a teacher like that, most likely they will take the below grade level group or the students who are maybe not quite as skilled.. we had a conversation among the team, and we decided that that would be the better model that we use.

Beyond academic support, educators also described social and cultural benefits of the intervention structures—especially when groups were intentionally mixed across classrooms. By

reorganizing students into different learning groups, schools created new opportunities for connection and belonging, reducing isolation and strengthening relationships that carried beyond the intervention period. One educator explained:

Because all of our intervention groups are mixed classes, you're getting kids from different classrooms together, and it's generating friendships across the classrooms that may not have happened... So that changes the entire culture of the school. They're not in isolation anymore...I think, it feels a lot more like a family than they did before.

Teachers also described how changes to their instructional approach through the IAM influenced how students approached their own learning. As instruction, intervention, and assessment became more coherent and routine, students began to understand why they were learning specific skills, how practice connected to assessment, and how evidence was used to guide additional support rather than label them as “failing.” In this way, the system helped strengthen students’ sense of purpose and a growth-oriented mindset about improvement. As one teacher explained:

I think students can tell that their learning has a purpose. We're not just teaching something because “it's school.” When I teach a skill, it's intentional, and I tell them that. I point out what's new, what we've learned before, and how we're building onto it. And I ask them, “After we learn this, what do we do next?” They'll say, “We take an assessment.” They know the pattern—we learn, we practice, then we assess. So they're aware that their learning leads somewhere. They know what they're learning now is getting them ready for the next grade, and they understand the connection between instruction and assessment. I think that helps them see that we're not just learning for the sake of learning—there's a purpose behind everything we do. I think it's also important for students to understand that taking a test isn't about failing or passing—it's about helping us, as teachers, figure out how to support them. The assessment shows us what skill they need help with so we can bring them back into intervention. It's really about shifting that mindset. Instead of thinking, “I didn't pass, so that's it,” we help them see that there's more beyond that. We can work on the skill together, they can improve, and they can retake the test once they understand the concept. And then there's the power of “yet.” You didn't master this yet, but we're going to revisit it, and you will. And giving them feedback along the way helps reinforce that.

And as students experienced repeated cycles of goal-setting, practice, assessment, and targeted support, teachers noted a shift in student ownership of their own learning, paying attention to their own data and viewing growth as something they could influence. In some cases, even very young students started asking explicit questions about whether they were improving, signaling

that progress monitoring had become part of the learning culture rather than something done only *to* them. As one teacher at a different site described:

The kids are more interested in tracking their progress now, which they never, ever, ever even considered before. And that's, that's a big deal. Because if you've got first graders wanting to know, did I improve? Am I caught up? Things like that. It's a big deal. It is a big deal.

One educator attributes the change in student engagement in their own learning to the increased clarity and coherence of the learning cycle—clear goals, transparent expectations, and a visible connection between instruction, assessment, and next steps:

It's very different from before. Previously, they'd take a test from a workbook or a blackline master, get a score, and that was it. They didn't really understand the purpose, or what the expectations behind it were. Now there's much more clarity for them. They may not consciously realize how different it is, but I can see it. They're more aware of their learning, they understand the goals, and they know what the road ahead looks like.

These examples of adjustments to teaching and learning suggest that the technical routines and adaptive shifts described earlier translated into concrete changes in classroom practice. By Year 3, teams were using shared evidence to diagnose learning needs more nimbly, coordinate reteaching and supports, and organize intervention more intentionally. This included flexible staffing and the strategic use of teacher strengths. These adjustments are an important proximal indicator that collaborative inquiry was changing teaching and learning in IAM schools.

Quantitative Results

The qualitative findings above document how the IAM supported technical and adaptive shifts in collaborative routines, instructional practice, and shared responsibility for student learning. To complement that implementation evidence, this section examines whether IAM participation was associated with measurable changes in student outcomes over time.

Using publicly available, school-level data on math and ELA achievement and attendance, we estimate IAM's association with outcomes across participating schools and compare patterns across post-implementation years. Given the small number of IAM schools and the loss of two baseline testing years during COVID, we expect effects to be small and noisy, and statistical power is limited. Even so, we look for consistent directional patterns, particularly whether outcomes appear more favorable by Year 3, which would align with the implementation maturation story suggested by the qualitative findings. Although the estimates are not

statistically significant, we report results transparently and emphasize robustness checks and sensitivity analyses to assess whether the observed patterns are stable across specifications.

The following section presents results from the impact analysis comparing IAM schools to non-participating campuses within the pilot districts. Additionally, we provide a descriptive portrait of IAM schools performance on the outcome measures included in the impact analysis to better understand these schools performance trends absolutely, and relative to comparison schools. We present sensitivity analyses in Appendix C that expand the comparison pool beyond the pilot districts. We examine two dimensions of student outcomes: math and reading/ELA performance, and school-day attendance. This section is organized by finding. Within each outcome domain, we present results from the TWFE and event-study specifications.

Before presenting the quantitative results, we note that several IAM schools have received external recognition from Solution Tree and AllThingsPLC for implementing the PLC at Work framework and documenting student learning gains. One school was designated a Model PLC School, and three schools were recognized as Promising Practices Schools. These recognitions are based on school applications that include narratives describing PLC structures and consecutive years of improvement in student achievement data, often drawing on multiple assessments and subgroup-specific evidence.

In contrast, our quantitative analysis estimates IAM's incremental association with publicly available SBAC and attendance outcomes relative to comparison schools in the same districts during a period of uneven post-pandemic recovery. For this reason, it is not contradictory for a school to demonstrate meaningful within-school improvement, and to earn recognition, while the average difference-in-differences estimates remain small and statistically indistinguishable from zero.

Descriptive Findings

First, we summarize the performance trajectories of IAM schools, non-participating schools within IAM districts that were selected for the comparison pool used in the impact study. We calculate performance trajectories for two counterfactual conditions: non-IAM schools with no adjustments to account for differences in baseline characteristics, and using statistical weights to rebalance the composition of the comparison schools to resemble IAM schools during the pre-implementation period. These findings illustrate the range of outcomes observed among IAM participants, and also provide important contextual information to support the impact results. These trajectories reflect each school's progress over time, and how these trajectories compared to statistically comparable schools in implementing districts. These descriptive

portraits should not, however, be interpreted as providing evidence of the **impact** of IAM implementation on participating schools, since

Finding 1. Several “Spotlight Schools” demonstrated large, consistent improvement in state standardized assessment performance.

The figures below depict the performance trajectory of IAM schools between the year prior to implementation (2021/22) and the last post-adoption year for which data are available (2024/25). In each figure, we also superimpose three additional data series: the unweighted control group average, the weighted control group average, and the IAM school average. Exceptional improvements, or declines—which we define as being higher or lower than the median change among IAM schools between 2023/24 and 2024/25—between 2023/24 and 2024/25 are green and red, respectively.

Several patterns are evident in the figures. First, performance levels and trajectories differ considerably across schools, and some of these patterns are consistent with IAM stakeholders' field reports. For instance, Harkness Elementary exhibited above-average ELA and mathematics performance improvements between the last two years, and their performance by the third of implementation was *higher* than the year immediately prior to implementation. Fillmore's largest improvements were limited to ELA tests.

Second, in the post-adoption period, the patterns of change are not consistent across IAM schools. The largest improvements occurred between the second and third year of implementation; this descriptive pattern is a precursor for the impact findings, which find consistently larger, and positive effects, during the third year of implementation. Highland, Harkness, and Nishimoto exhibited the largest improvements in mathematics proficiency rates during this time frame. Last, attendance rates improved across all IAM schools. However, these improvements **were nearly identical** to those experienced by similar comparison schools.

Figure 3. ELA proficiency rate trajectories, by IAM participation status, 2021/22 to 2024/25

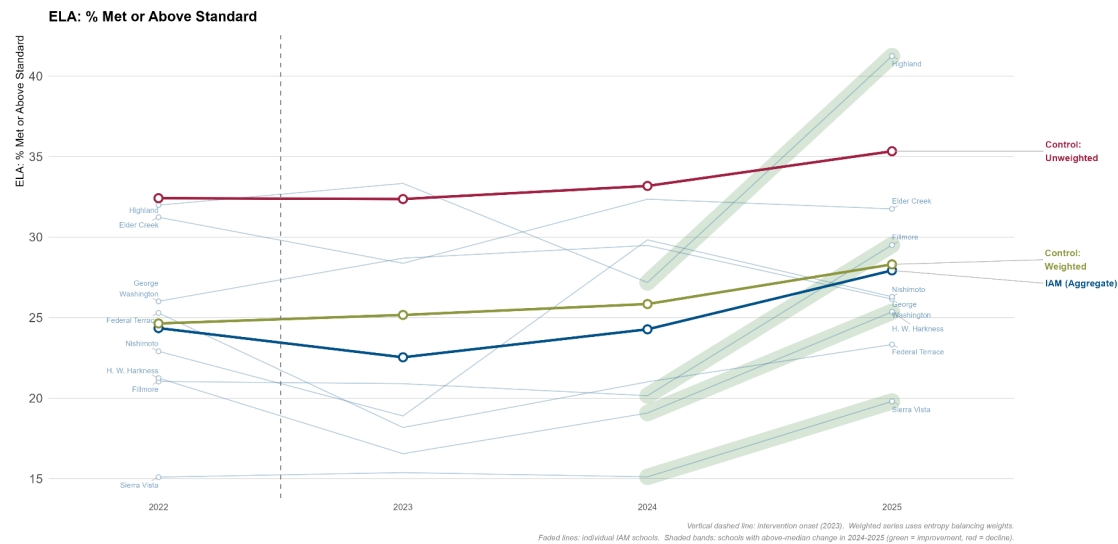
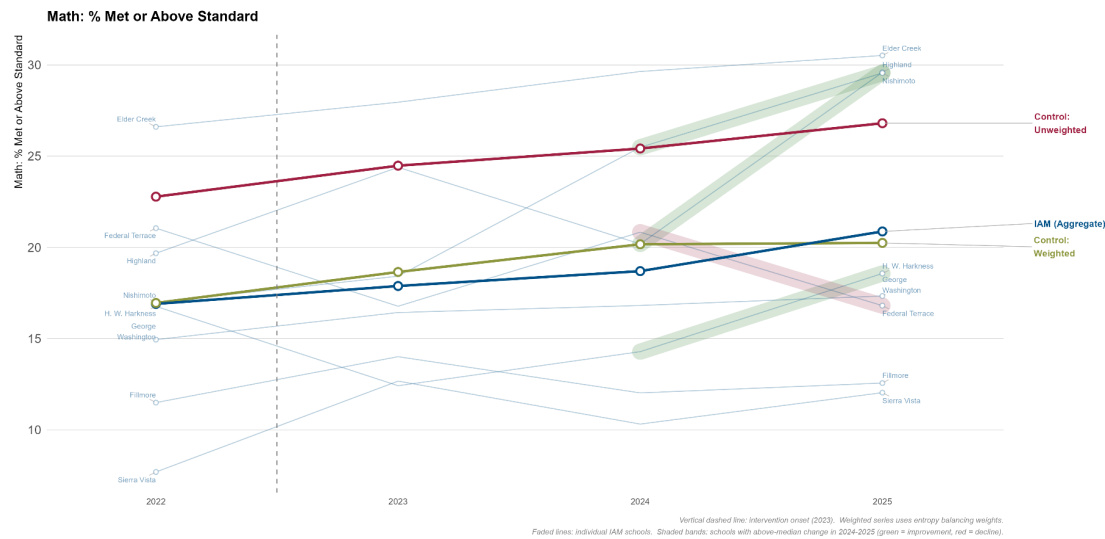


Figure 4. Math proficiency rate trajectories, by IAM participation status, 2021/22 to 2024/25

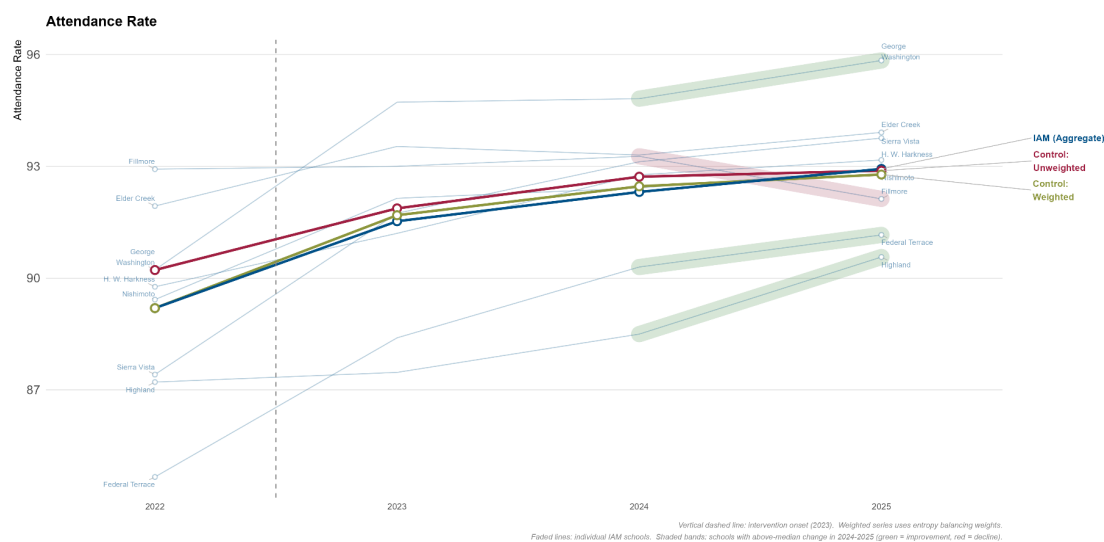


Finding 2. School-level attendance rates increased sharply and consistently across nearly all IAM schools

Figure 5 reproduces the presentation of school-level performance trends presented in the figures above, but for student attendance rates, attendance rates increased from 2021/22 to 2024/25 for both the IAM schools, on average, and the comparison groups. The patterns exhibit substantial heterogeneity across IAM schools. Some schools appear to achieve large improvements from 2023/24 to 2024/25, while others flatten or decline slightly. That variation matters because the aggregate IAM trend—and the analyses presented in the impact results section that follows—masks meaningful differences in school-specific trajectories.

At the same time, these patterns should be interpreted cautiously. The figure is descriptive, not causal, and visual convergence or divergence across lines should not be taken as evidence that IAM caused changes in attendance. The post-2021/22 improvement in IAM schools could reflect the intervention, but it could also reflect broader recovery trends that occurred in California and nationwide in the aftermath of the COVID-19 pandemic. A cautious interpretation is that IAM schools appear to have improved attendance over time and generally followed the same patterns of comparison schools—both statistically unweighted and weighted.

Figure 5. School day attendance rate trajectories, by IAM participation status, 2021/22 to 2024/25



Impact Findings

Finding 1. IAM adoption was associated with a medium-sized, positive effect on standardized mathematics achievement by the third year of implementation.

The event-study figures below show estimated year-by-year effects of IAM on math and ELA scale scores. Each point represents the estimated difference between IAM and comparison schools in that year, relative to the year before IAM launched (2021–22). Vertical bars are 95 percent confidence intervals. The figures communicate two important pieces of information about the dynamic, time-varying effect of IAM on student outcomes. For instance, during the pre-launch phase, the statistically insignificant effects near zero lend tentative support for the parallel trends assumption that underpins the validity of the post-adoption effects shown starting in the first year of implementation (2022-23).

Across all four testing outcomes, event study estimates are statistically indistinguishable from zero. However, while math scale scores decline modestly in the first two post-launch years, the third-year impact is medium-sized and positive (though not statistically significant). ELA estimates are modestly negative in the first and second implementation years (2022-23 and 2023-24) (approximately -0.10 SD). The ELA proficiency rate shows a larger negative estimate in the first adoption year (2022-23, approximately -3 percentage points) but is not statistically significant. Because the first implementation year is also the second post-COVID year of statewide testing, differential COVID recovery between IAM and comparison schools cannot be ruled out as a contributor to these patterns.

Pre-launch estimates are close to zero across all academic outcomes, lending support to the parallel trends assumption that is critical for establishing the validity of these estimates; however, the two-year COVID gap in the pre-period may conceal differential trajectories in student performance between 2018-19 and 2021-22. For instance, if the COVID epidemic asymmetrically affected IAM and non-IAM schools in difficult to measure ways, the estimates of the pre-adoption period for the available years may not accurately capture their underlying trajectory, though the direction of the bias this may introduce is indeterminate.

Figure 6. Event-study estimates: math standardized scale score

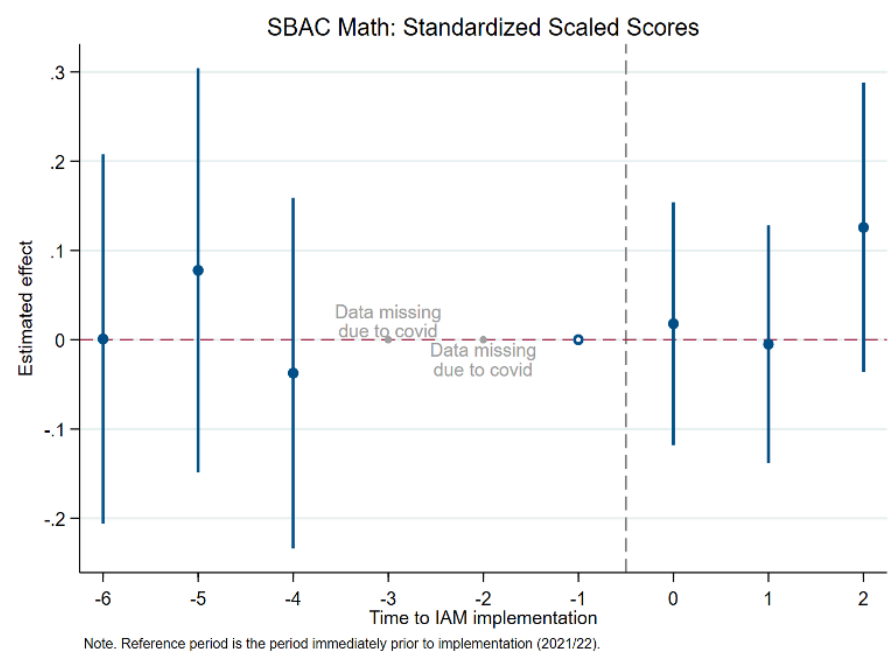


Figure 7. Event-study estimates: math proficiency rate

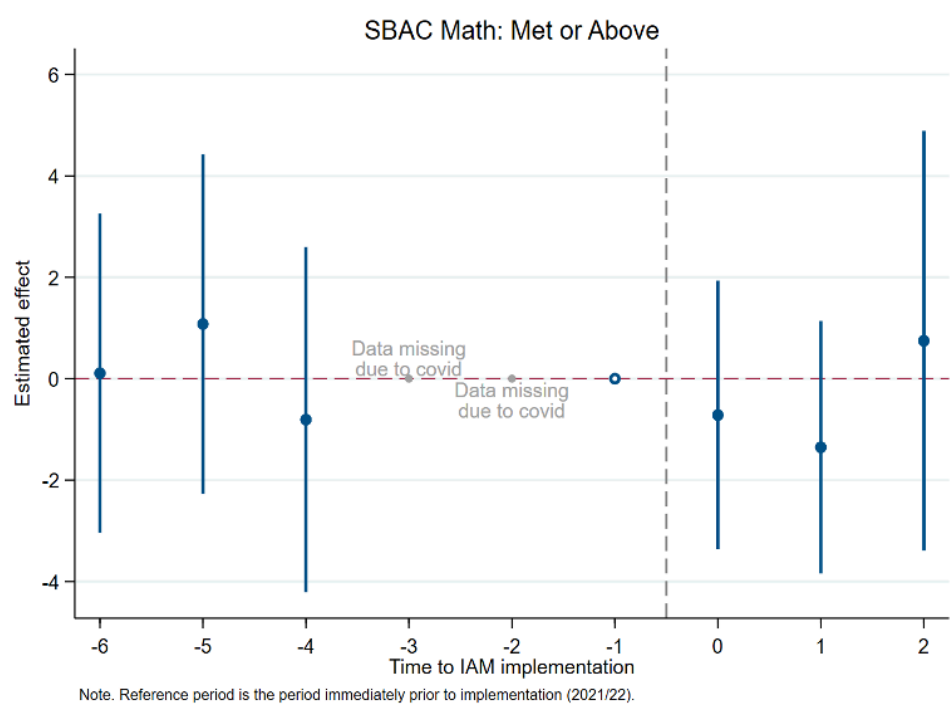
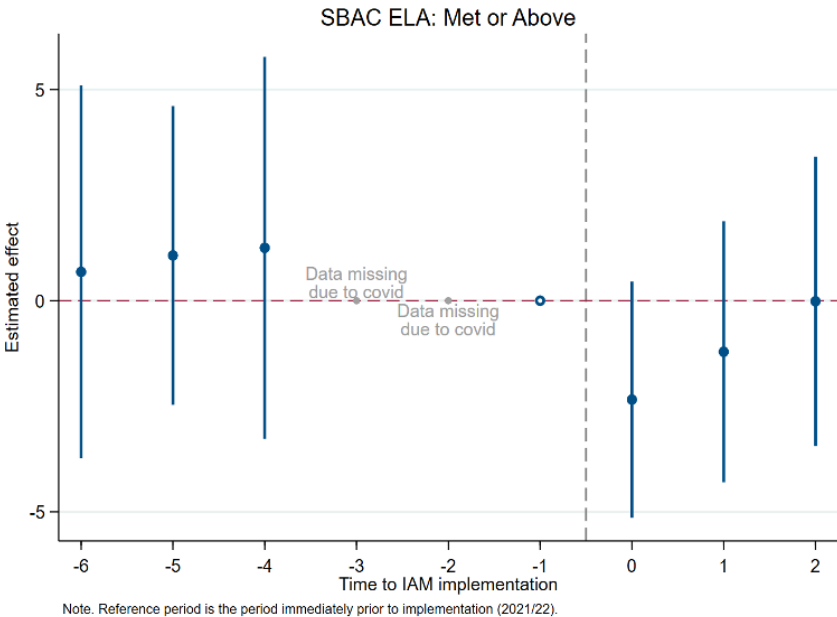


Figure 8. Event-study estimates: ELA standardized scale score



Figure 9. Event-study estimates: ELA proficiency rate



For the SBAC ELA, the pre-to-post sign reversal, from small positive pre-period gaps to a notably negative effect in the first adoption year, is the clearest anomalous pattern. Two interpretations warrant consideration. First, IAM implementation may have produced a short-term drop in ELA proficiency rates in the first implementation year that attenuated by year 2. Second, comparison schools may have rebounded more quickly from COVID-era disruptions on this outcome, creating an apparent post-period divergence unrelated to IAM.

Finding 2. The combined effect of AM adoption was small, but positive for standardized mathematics scaled scores, and small and negative for ELA test performance.

Using two measures of math and ELA performance, the estimated effect of the IAM model on pilot schools was small and slightly negative—save for the math standardized scale score—but the effects were not statistically significant. For instance, relative to the differences between treatment and comparison schools, the percentage of students in tested grades at the school was nearly 2 percentage points less likely than non-participating campuses to earn a score that met or exceeded proficiency, and one percentage point less likely for Math. The results for the standardized scale scores exhibit a similar pattern.

Figure 10. Standardized scale scores

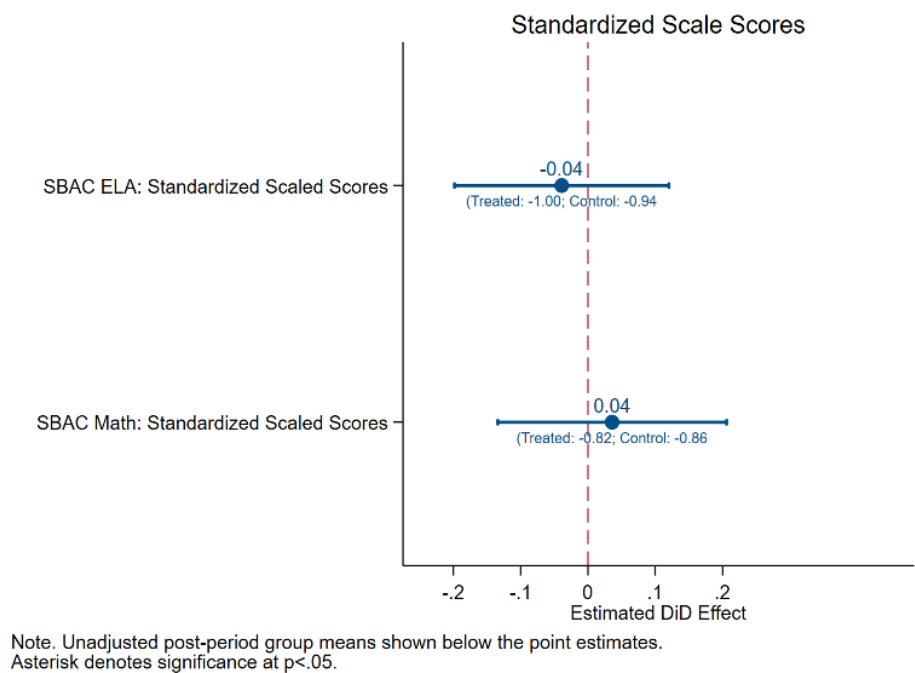
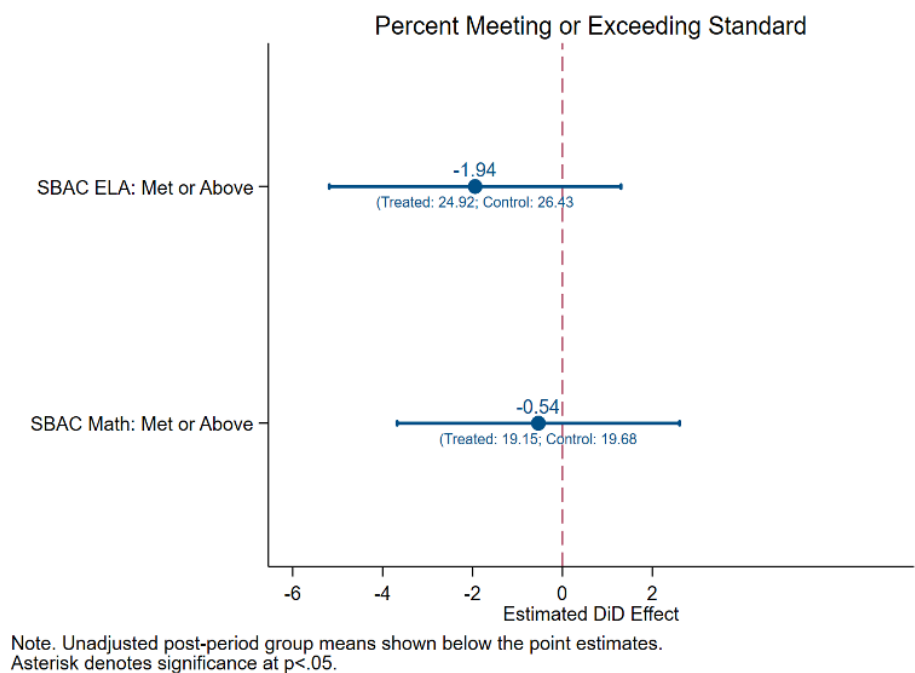


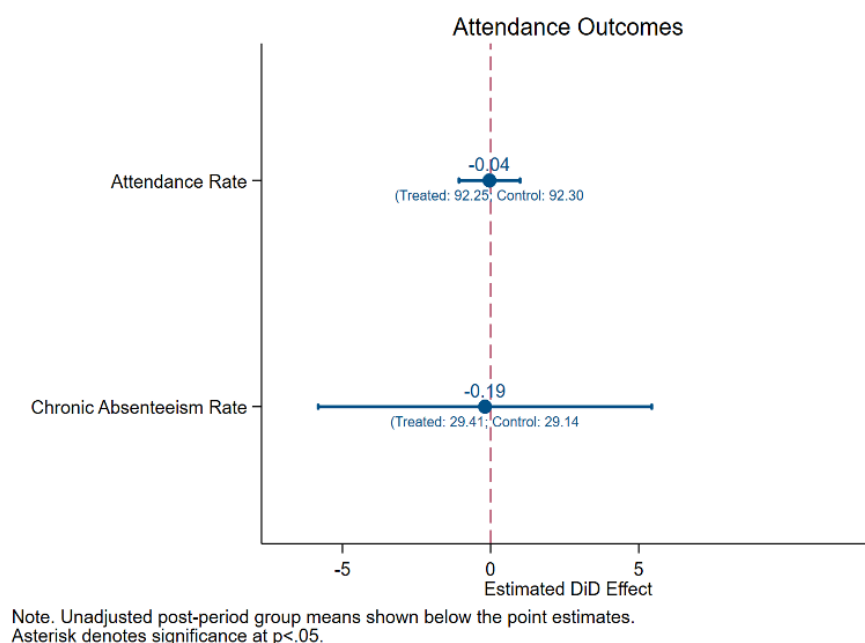
Figure 11. Percentage of students who met or exceeded the proficiency standard



Finding 3. Differences in pre- and post-adoption school-day attendance rates between IAM and non-IAM schools were not statistically distinguishable from zero

There is no evidence that IAM adoption substantively affected school-level attendance or chronic absenteeism rates. For instance, the estimated effect on overall attendance rate is -0.04 percentage points, with very similar post-period means (treated 92.3 percent). For chronic absenteeism, the effect is similarly small, negative (0.19 percentage points, with practically identical post-adoption means), and not statistically significant.

Figure 12. School-day attendance



Finding 4. IAM was not associated with any detectable changes in school-day attendance or chronic absenteeism in any post-implementation year.

The event-study figures below display year-by-year estimated differences in attendance rate and chronic absenteeism between IAM and comparison schools, relative to the year before IAM launched (2021–22). For both outcomes, pre-launch estimates are flat and roughly zero, providing the cleanest pre-trend pattern in the analysis and lending support to the parallel trends assumption.

In the post-launch period, neither measure of student attendance exhibits a statistically significant treatment effect in any year. Attendance rate estimates in the first and second year

are small and slightly negative (approximately -0.2 to -0.3 percentage points). Chronic absenteeism post-launch estimates are positive and near zero (approximately $+0.5$ to $+1.7$ percentage points), but not statistically significant.

Figure 13. Event-study estimates: attendance rate

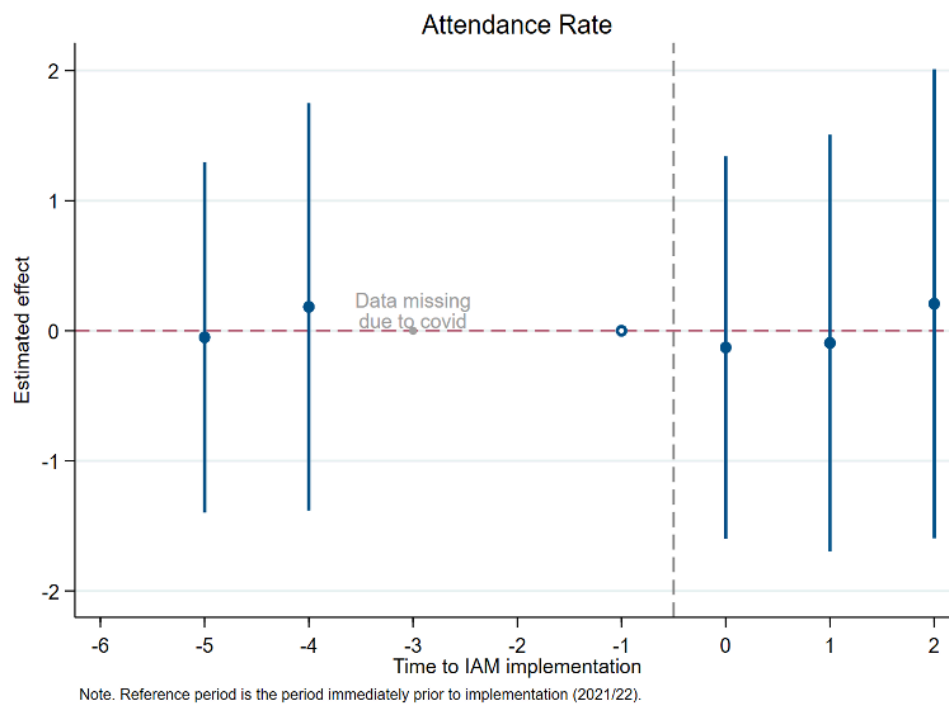
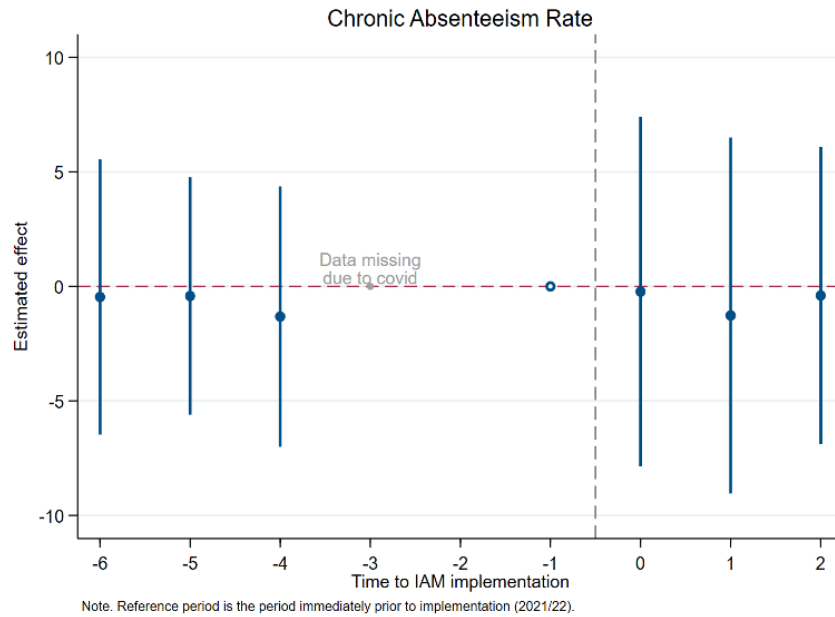


Figure 14. Event-study estimates: chronic absenteeism rate



Effects of IAM Adoption by Tested Grade Levels

IAM schools serve diverse student populations in high-need California communities. Understanding whether program effects are shared broadly or concentrated in particular grades or student groups is both a substantive question and an equity consideration. This section presents results separately by grade level (grades 3–6) cohorts.

Testing Outcomes

Finding 5. Grade-level effects of IAM on testing outcomes are highly heterogeneous, with Grade 6 showing positive and growing effects and Grade 5 showing consistent negative effects across all four outcomes.

The figures below show event-study estimates for math and ELA scale scores and proficiency rates by grade level (Grades 3–6). Each point represents the estimated difference between IAM and comparison schools in that year for a given grade, relative to the pre-implementation reference year (2021-22). Asterisks denote estimates that are statistically significant at the $p < .05$ level.

The grade-level design also permits insight into how effects may evolve as students accumulate exposure to IAM-supported instruction. Because the same grade is observed in each implementation year, the change in a grade's event-study estimate from the first year of

implementation (2022-23) to the third year (2024-25) reflects, in part, how the gap between IAM and comparison schools shifts as the cohort of tested students changes. Students in Grade 3 in 2022-23 had experienced roughly one year of IAM, while Grade 3 students tested in 2024-25 who enrolled in kindergarten or first grade at an IAM school would have been enrolled for up to three years. If cumulative exposure or program maturation is associated with improvement, estimates for a given grade would be expected to strengthen across the post-period. (A number of assumptions underpin this interpretation. First, it requires that student mobility rates are low and not systematically different between IAM and comparison schools; selective enrollment or attrition would shift the composition of tested students in ways unrelated to program exposure. Second, school-level grade means aggregate all students tested in a given grade and year, including those who transferred in mid-program, so the estimates reflect a mix of exposure durations rather than a clean dosage comparison. Third, this logic assumes IAM was implemented with consistent fidelity across all grade levels from the start of the program; if the model was phased in or adopted more gradually in lower grades, actual student exposure would be smaller than the number of years of school enrollment implies. Finally, differential post-COVID recovery trajectories between IAM and comparison schools could introduce grade-specific time trends that confound changes in post-period estimates independently of program effects.)

Across all four testing outcomes, the grade-level figures reveal substantial heterogeneity that is obscured in the school-level overall estimates presented above. The most consistent pattern is a divergence between Grade 6—which shows positive and growing post-period effects—and Grade 5, which shows persistently negative effects throughout the implementation period.

Grade 6 shows positive post-period effects that grow over time and reach statistical significance by the third year of implementation. For standardized scale scores, the year-3 effect is approximately +0.47 SD in math and +0.65 SD in ELA. Proficiency rate estimates at year 3 are approximately +9 percentage points in math and +12 percentage points in ELA. This finding should be interpreted cautiously and with the pre-implementation patterns in mind. That is, grade 6 also exhibits elevated pre-period estimates across outcomes—some reaching statistical significance—indicating pre-existing differences in trajectory between Grade 6 IAM and comparison schools, which could bias the post-adoption effects.

Grade 5 shows the opposite pattern, with consistently negative and statistically significant post-period effects across most outcomes. ELA proficiency estimates for Grade 5 range from approximately -6 to -11 percentage points across the three post-implementation years; math estimates are similarly negative (approximately -7 to -8 percentage points). Scale scores for Grade 5 also trend negative in both subjects. Notably, Grade 5 also shows negative pre-period estimates in most outcomes, suggesting that IAM and comparison schools in this grade may have differed in underlying trajectory before the program launched.

Grades 3 and 4 present more mixed patterns. Grade 4 trends toward modest positive effects in the second and third implementation years—particularly in math scale scores ($\sim +0.37$ SD at year 3) and math proficiency ($\sim +6$ percentage points)—but does not reach conventional significance thresholds in any outcome. Grade 3 begins the post-period near zero but trends increasingly negative by year 3, with statistically significant proficiency rate estimates in both ELA (approximately -11 percentage points) and math (approximately -9 percentage points).

We note that these grade-level patterns should be interpreted with caution. In a sample of eight schools with mixed grade configurations, any single grade's estimates may reflect the experience of only one or two schools, producing imprecise estimates with wide margins of uncertainty. Further, the pre-adoption coefficients suggest the parallel trends assumption is likely violated. Thus, the grade-level results are best understood as descriptive of heterogeneity rather than as precise causal estimates.

Figure 15. Event-study estimates: math proficiency rate by grade

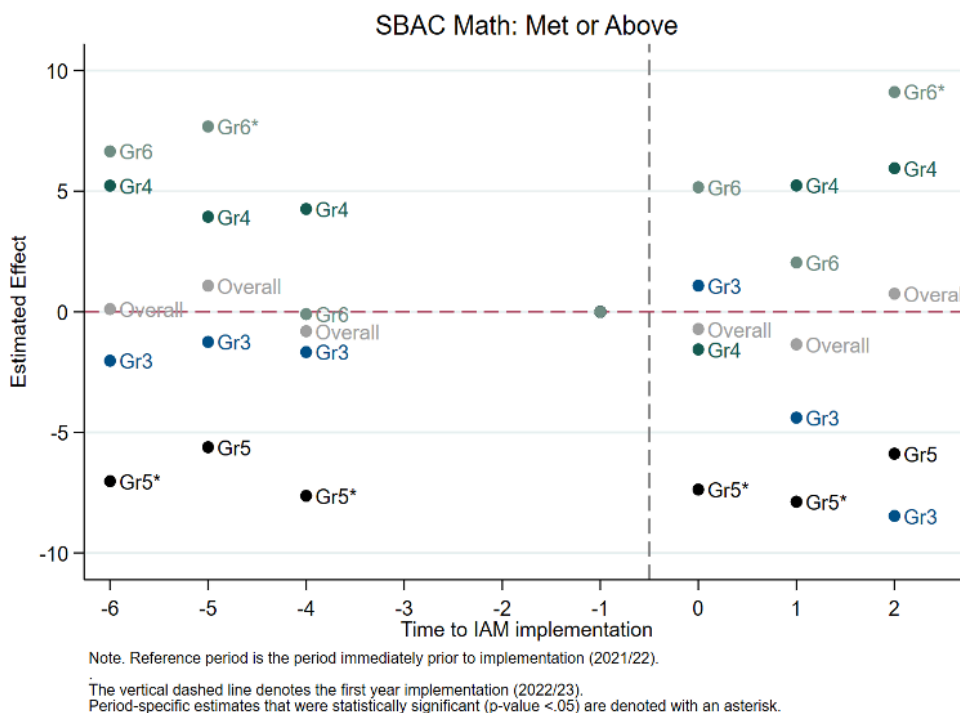


Figure 16. Event-study estimates: math standardized scale score by grade

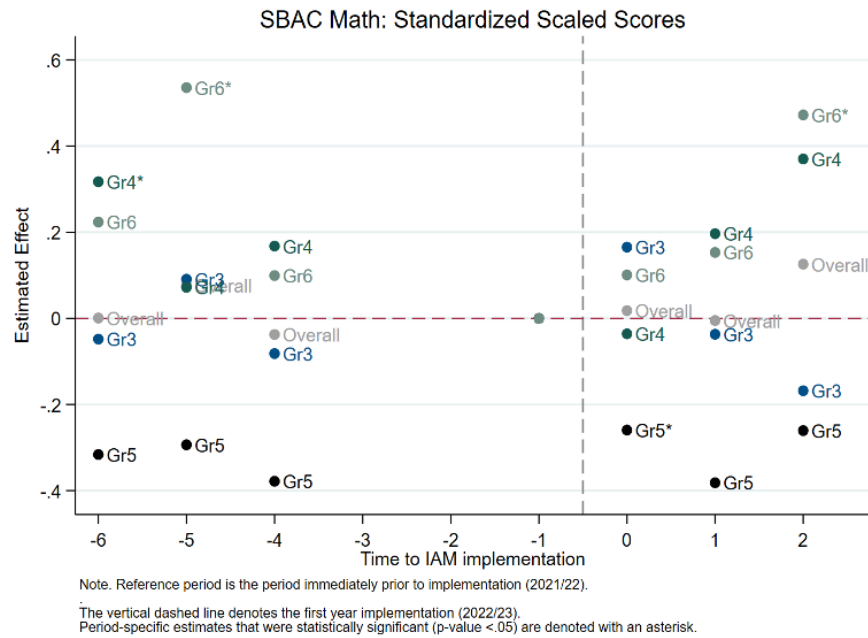


Figure 17. Event-study estimates: ELA proficiency rate by grade

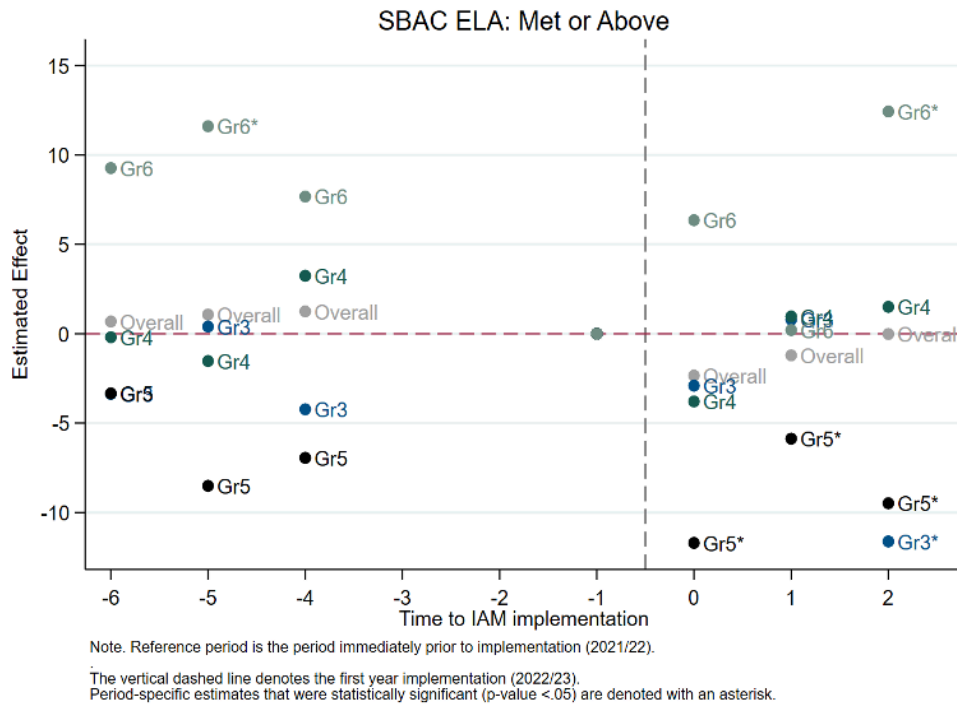


Figure 18. Event-study estimates: ELA standardized scale score by grade



Discussion of Quantitative Results

The findings from this analysis provide limited evidence that participation in IAM improved student outcomes in the first three years of implementation. Across the six outcomes examined — standardized math and ELA scale scores, proficiency rates in math and ELA, school-day attendance rates, and chronic absenteeism — estimated effects are consistently small and statistically indistinguishable from zero at the school level, suggesting that IAM's model has not yet produced detectable average gains across the five pilot districts during the observation window.

The event-study estimates are informative about the timing and trajectory of effects. Across all academic outcomes, post-period estimates cluster near zero throughout the three post-implementation years with no consistent pattern of improvement over time. Math scale scores are modestly negative in the first two post-launch years but turn slightly positive by the third year; ELA scale score estimates are modestly negative in years one and two (approximately -0.10 SD) before attenuating. For attendance, pre- and post-launch estimates are flat and near zero — the cleanest pre-trend pattern in the analysis, providing clear evidence of parallel pre-period trends for attendance outcomes. A notable caveat is the COVID-era assessment gap (2019-20 to 2020-21): the absence of test data during these years limits the ability to rule out differential recovery trajectories between IAM and comparison schools, and the first implementation year (2022-23) coincided with the second year of statewide post-pandemic

testing, which may have introduced outcome-specific confounds that our design cannot fully absorb. The pooled difference-in-differences estimates, which average effects across all post-launch years, are consistent with the event-study findings.

Heterogeneity by grade reveals sharply divergent patterns that are obscured in the school-level pooled estimates. Grade 6 shows positive and growing effects across all four academic outcomes, reaching statistical significance by the third implementation year: approximately +0.47 SD in math and +0.65 SD in ELA scale scores, and approximately +9 and +12 percentage points in math and ELA proficiency rates, respectively. Grade 5, by contrast, shows consistently negative and statistically significant effects across most outcomes throughout the post-period, with proficiency rate estimates ranging from approximately –7 to –11 percentage points. Grades 3 and 4 show more mixed patterns: Grade 4 trends modestly positive in later years (not statistically significant), while Grade 3 trends increasingly negative and reaches statistical significance in proficiency rates by the third year. These grade-level differences should be interpreted with care. Pre-existing differences in outcome trajectories between IAM and comparison schools are evident for several grades — particularly Grade 5 and Grade 6 — which weakens the parallel trends assumption at the grade level and complicates causal attribution. With only eight schools across five districts, grade-specific estimates also reflect the experience of a very small number of schools, and the wide confidence intervals reflect corresponding imprecision.

Limitations of Quantitative Results and Considerations for Interpretation

The estimates reported in this study rest on a set of identifying assumptions and data constraints that readers should consider when interpreting the findings. We discuss each in turn below.

Unsurprising Impact on Outcomes Over a Short Time Span. Significant improvement in student outcomes is difficult to achieve under strong district conditions, and are even more so in districts that are in the highest tier of state-level support. A similar study on the impact of intensive Solution Tree support in Arkansas found no significant impact on student outcomes across 90 project schools, some of which had received five years of support (Barnes & McKenzie, (2024). Expecting results from the most struggling districts in California over a shorter time frame was unlikely and should not discredit the changes to adult practice that occurred over that same time frame.

Conditional Parallel Trends. The most important assumption underlying the difference-in-differences design is that, absent IAM, the outcome trajectories of IAM schools

and comparison schools would have evolved similarly. This assumption is demanding and, particularly in evaluations of educational interventions, challenging to confidently demonstrate and is, directly, untestable in the post-treatment period. Nonetheless, we provide suggestive, albeit indirect, evidence to buttress this assumption by examining the magnitude of pre-adoption event-study coefficients to ensure they are not statistically different from zero (Appendix E).

A particular concern is that IAM schools may have been differentially affected by post-COVID recovery dynamics. Because IAM launched in 2022–23 — a period of substantial heterogeneity in the pace of pandemic-era recovery across California schools — it is possible that some portion of estimated gains reflects differential recovery trajectories rather than program effects. Our design partially addresses this by restricting comparison schools to the same districts as IAM schools (ensuring shared district-level recovery conditions and policies), and by including multi-year pre-period baseline covariates in the primary model with entropy weights. Nevertheless, school-level differences in recovery pace that are not fully captured by our baseline measures remain a potential source of bias.

Selection on Unobservables. Statistical weighting methods, like those used in this study, balance the comparison group on observed baseline characteristics, but they cannot account for unmeasured differences between IAM and non-IAM schools. Schools that opted into IAM may differ from non-participants in ways that are difficult to capture in administrative data. For example, principal leadership quality, staff cohesion, community engagement, district or school leadership receptivity to external support, or other organizational factors may independently affect academic improvement. If these unobserved characteristics are positively correlated with both IAM participation and outcome growth, our estimates will overstate the causal effect of the program. Our statistical models attempt to account for time-invariant school- and district-level factors that are unavailable in the state administrative data, and we provide a battery of robustness checks to assess the sensitivity of the primary findings to this threat.

School-Level Data and Aggregation. The analysis relies on school-level aggregate measures of student outcomes rather than individual student records. Aggregation introduces two limitations. First, the analysis cannot follow individual students over time; schools' aggregate scores can change because of compositional changes in enrollment (e.g., if higher-achieving families enroll in or leave IAM schools after launch) rather than because individual students improved. We partially address this by including school fixed effects, which absorb time-invariant school characteristics, but if enrollment composition changes *within* a school over the study window in ways correlated with IAM participation, the estimates could be confounded. Second, school-level averages mask heterogeneity within schools. A program that benefits students in the middle of the achievement distribution but has no effect on the

highest- or lowest-performing students would produce an average treatment effect of limited policy relevance that is not distinguishable from a more uniform effect using the available data.

An additional limitation is that the scope of the analysis of achievement data is limited to tested grades (i.e., 3–8).

Short Post-Intervention Window. The current analysis covers at most three years of post-launch data ($k = 0, +1, +2$). Many school-improvement interventions require several years to develop and consolidate, and short-run estimates may understate long-run impacts if program effects grow over time. Conversely, some interventions produce early gains that fade as novelty effects dissipate or as comparison schools adopt similar practices. The event-study estimates illuminate the trajectory of effects within the observed window, but the analysis cannot characterize what happens beyond the third intervention year, or assess the long-run sustainability of any observed gains.

Implementation Fidelity and Dosage. The estimates in this study represent *intent-to-treat* effects — the average effect of being an IAM school — rather than the effect of any specific level of program implementation. Administrative data do not include information on whether or how fully IAM was implemented at each school, the degree of teacher participation in professional development, or the fidelity with which program components were delivered. If implementation varied substantially across schools, the estimated average effect will obscure meaningful heterogeneity: high-fidelity implementers may have experienced larger benefits, while low-fidelity schools may have experienced no effect.

External Validity. The five districts in this study share certain characteristics: they are predominantly urban and suburban, serve high proportions of socioeconomically disadvantaged and racially and ethnically diverse students. Findings may not generalize to schools in more rural settings, states with different accountability systems, or districts with different resource levels. The relatively small number of treated schools (8 IAM schools across five districts) also compromises the statistical power of the study, particularly for subgroups. Results should be interpreted as suggestive evidence about IAM's effects in contexts similar to the five participating districts.

Potential Spillovers and the Stable Unit Treatment Value Assumption. The study team winnowed the analytic sample to schools in the five pilot districts for the primary analysis. This inclusion rule can improve the internal validity of the design, by ensuring both IAM and non-IAM schools are exposed to a similar organizational setting that may influence student performance. However, the credibility of the treatment effect estimates relies on the assumption that the intervention did not introduce equilibrium effects for other non-IAM

schools in the district. This stable unit treatment value (SUTVA) assumption (Rubin, 1980) assumes, for instance, that IAM adoption did not lead to an exodus of high-performing teachers to non-participating schools (or, alternatively, a migration of high-performing teachers from other schools to IAM campuses), and it assumes that, say the coaching or other external support provided through IAM is targeted exclusively at pilot schools and not comparison campuses.

Challenges to Scale

A primary goal of the IAM was to leverage the work of the pilot schools to scale the work to other schools across the district. The PLC model is a school based model that can be adapted to support entire districts (Cottingham & Myung, 2025). To do so, districts need to develop plans to scale the work that matches their context and resources while also developing their capacity alongside their schools. This differs from the school-level PLC model which is much more prescriptive and in the instance of the IAM, heavily supported with coaching. While we saw evidence of practices changing at individual schools we heard little evidence of how these changes would be scaled to other sites. The one district that voluntarily joined the IAM had initial scale strategies in place, but interviewees noted none of those had started by the end of year three. None of the districts in DTA were prepared to scale at the end of the pilot. In this section we describe five of the primary challenges of district-level scale that were shared by interviewees.

One of the most immediate structural barriers to implementation identified across the Intensive Assistance Model (IAM) pilot was the significant financial and logistical strain of securing substitute coverage for teacher training and collaborative planning. The intensive nature of the model—requiring upwards of 50 days of on-site coaching annually—necessitated frequent teacher pull-outs that many districts struggled to manage amidst widespread staffing shortages. Critically, the substantial cost of these substitutes was not included in the primary pilot fees, which already totaled nearly \$940,000 per school over three years. For participating districts, many of which were already in fiscal receivership or facing significant budgetary constraints, this added expense created a prohibitive financial hurdle. In contexts where school days lacked embedded collaborative structures, administrators were forced to divert limited local resources to ensure teachers had the time together to collaborate, creating a precarious implementation environment where the model's success was often dictated by a district's ability to absorb these unallocated costs.

A secondary barrier to sustainability involves a significant gap in district-level expertise, as many central office staff lack the specific capacity required to lead or support a comprehensive

"Model PLC at Work" transformation. While some district personnel possessed prior experience with informal professional learning communities, few had successfully navigated a school through the systemic cultural and structural shifts required for this specific intensive assistance model. This challenge was exacerbated by a lack of coordination; district staff were frequently excluded from school-level trainings unless explicitly directed by senior leadership or invited by a principal, leaving them as observers rather than active facilitators or collaborators in the work. Without a clear mechanism for developing this expertise through the IAM coaching allocation, district staff expressed significant concerns regarding their ability to provide the "supportive accountability" and technical guidance necessary to sustain the model once the external assistance ends.

In many participating districts, an initial concern was the potential for labor tensions, specifically regarding whether collectively bargained planning time could be repurposed for the highly structured collaboration required by the IAM. These concerns were particularly acute in districts already navigating intense labor negotiations or the destabilizing process of entering DTA. While some districts experienced resistance from a small number of educators reluctant to shift from individual planning to collective accountability, this tension generally proved less disruptive than anticipated. Evidence from the IAM suggests that in the majority of pilot schools, teachers were encouraged by the collaborative framework and eventually became advocates for the instructional shifts. Critically, districts that proactively included union representation on school-site "guiding coalitions" and in initial trainings experienced significantly smoother implementation. By integrating labor representation into the leadership of the model from the outset, districts were able to establish a shared understanding of common expectations and prevent any tensions experienced by staff from blocking systemic progress.

A significant barrier to systemic impact was the absence of a concrete, actionable plan for scaling the model beyond the initial pilot schools. While various strategies were proposed in interviews – such as developing "lab schools" where principals could serve as peer mentors or cultivating a cadre of county-level coaches to replicate the intensive training – we did not collect evidence of any of these scaling strategies in action. In several districts, Curriculum, Instruction, and Assessment staff participated in early trainings, but by the end of the pilot, it was clear that their involvement lacked a strategic trajectory. Additionally, several districts expanded their Solution Tree support to include working with the district cabinet but again processes for scaling to other schools was unclear. Without a clear vision for how these district-level leaders would facilitate the model at other sites, the work remained constrained to pilot schools. The lack of an initial plan for scale at the outset of the IAM meant that most districts were still developing a scale strategy or had no such plan by the end of the pilot.

The final barrier to successful implementation was the pervasive issue of initiative overload, where schools were forced to navigate a fragmented landscape of competing priorities. As noted earlier in the brief, several pilot sites experienced a "push-in" model where as many as 19 different coaches from various district-mandated initiatives were operating simultaneously within a single school. This lack of coherence placed an immense burden on schools already struggling with staffing shortages and student needs, forcing site leaders to actively negotiate with the district office to opt out of certain programs just to preserve the "minimum viable" time needed for the IAM. This tension highlights a significant disconnect in district-level understanding of what was required of schools to participate in and effectively implement the model.

Policy Recommendations

The IAM pilot underscores a core lesson for California's improvement agenda. Shifting teaching and learning at scale requires more than setting new expectations. It requires both the technical capacity to implement new practices and the adaptive capacity to build and sustain norms of collaboration, shared responsibility, and evidence-based decision-making. These shifts are unlikely to take hold without aligned support.

The policy recommendations that follow translate these findings into actionable steps for state and district leaders. They focus on building an infrastructure for capacity development, strengthening accountability by emphasizing the processes that produce outcomes, securing protected time for collaboration and intervention, and investing in districts as the essential unit of change for coordinating improvement across schools.

Policy Recommendation 1: Establish a robust state and regional infrastructure of instructional coaching and leadership support to ensure that school-level autonomy is paired with the technical and adaptive capacity required for high-quality implementation and improvement of teaching and learning. While this model trusts local leaders to make the best decisions for their students, the evidence indicates that implementation fails when schools are left to navigate complex shifts without ongoing professional development and adult capacity building. This support must originate from the district or state level, or through specialized external providers as demonstrated in the IAM pilot. Other states, such as Arkansas, have leveraged similar pilots to cultivate internal leadership, developing regional and state-level coaching structures focused on PLC processes to sustain capacity building for teachers and principals. As one Arkansas state-level policy director described, the commitment was to build internal capacity so that districts "are not going to always depend on [external providers]," ensuring that ongoing coaching is facilitated by "a local practitioner." The development of a similar cadre of

instructional leaders is essential for California, as success in high-performing districts is consistently linked to sustained guidance that allows teams to achieve implementation fidelity over time. However, such support is currently an outlier within the California educational ecosystem. To bridge this gap, the state must transition from a compliance-oriented posture to one that actively equips leaders through a multi-year investment in high-capacity practitioners and external providers who can offer the specialized expertise necessary for systemic change.

Policy Recommendation 2: Evolve state accountability for districts under Direct Technical Assistance (DTA) by limiting dashboard indicators to core academic outcomes while integrating process-oriented metrics to guide intensive support. Evidence from the IAM pilot suggests that high-fidelity implementation is frequently compromised by initiative overload, requiring school leaders to negotiate with district offices to deprioritize competing mandates in favor of deep engagement with instructional change. By narrowing the accountability focus to ELA and math performance alongside implementation fidelity, the state can empower struggling systems to make the technical and adaptive shifts necessary to impact the instructional core. While the state is already reducing dashboard indicators, we recommend further streamlining these metrics for DTA districts and introducing process indicators—modeled after the PLC at Work Continuum or the California Standards for the Teaching Profession—to evaluate changes in adult collaborative practices. This shift requires the state to move beyond compliance-based monitoring toward providing the specialized coaching and regional infrastructure defined in recommendation one. Limiting priorities to a defined set of criteria for improvement allows the most challenged districts to develop the deep expertise required to move teaching and learning.

Policy Recommendation 3: To sustain high-fidelity implementation, California districts must prioritize restructuring elementary school schedules to embed "specials"—such as PE, music, and the arts—within the instructional day as a primary mechanism for securing essential teacher collaboration and student intervention time. Evidence from the IAM pilot reveals that the lack of dedicated time was the most immediate barrier to success. Districts must guarantee a minimum of one hour of weekly collaboration for educators to analyze shared assessments and refine collective practice. Our evaluation found that the schools able to implement the program with the greatest fidelity were able to do so in large part due to district decisions to integrate these specialized subjects back into the core school day. In so doing, districts create a dual-benefit system: 1) Students gain access to diverse academic and social experiences that foster holistic growth, while 2) teachers gain the structural "release time" necessary to collaboratively analyze evidence of student learning and provide immediate, data-driven interventions for high-needs students. Providing access to the resource of time benefits both students and teachers and must be accompanied by the support of policy recommendations

one and two to ensure that time is used to improve teaching and learning through collaborative adult learning, planning, and intervention.

Policy Recommendation 4: California must prioritize the district as the essential unit of change, providing the structured technical assistance necessary to coordinate supports coherently across the system for both adults and students. Scaling high-quality instruction requires moving beyond isolated school-based improvements toward a model where district leaders—the only actors positioned to ensure consistency across sites—take responsibility for systemic alignment. Evidence from this pilot and other examinations of scaling district improvement suggests that relying on individual school spread leads to uneven quality and fragile implementation (Gallagher et al, 2026). Instead, district leaders must use their authority to embed instructional priorities into required routines, such as professional learning time and monitoring processes. To avoid a "patchwork" of disjointed efforts, the state should help districts move away from monitoring that focuses only on approving initial plans toward the creation of feedback loops that treat monitoring as a tool for learning and continuous adjustment. Ultimately, for improvement to last, senior leadership must commit to long-term, visible involvement that repurposes existing systems to ensure that high-quality teaching and collaborative adult processes become "the way we do things" throughout the entire district.

Conclusion

The IAM pilot, as a structured intervention to improve teaching and learning in high-need California districts, centered its efforts on building coherent, instructional-focused systems through the PLC at Work framework. The evaluation confirms that the core mechanism of the IAM is a process of scaffolded adult collaboration, driving change across two critical dimensions: technical shifts (e.g., mastery of essential standards, developing common assessments and systems for reviewing student learning) and adaptive shifts (e.g., changes to commitments to students' learning, adult collaboration, and mindsets about data use). A central finding of this analysis is that for the technical improvements to achieve fidelity of implementation and ultimately impact, schools must successfully engage in the underlying adaptive work, thereby moving beyond superficial implementation to fundamentally transform instructional decision-making and professional culture.

While the three-year evaluation timeframe provides a foundational analysis of implementation, the findings serve as critical proximal signals of whether the IAM is taking root in ways that plausibly influence longer-term student outcomes. Based on evidence from the pilot, the model's technical and adaptive shifts are unlikely to translate into statistically significant gains in standardized measures of student learning within a few years in districts facing persistent

barriers to academic improvement. However, the pilot suggests the model can meaningfully strengthen teacher practice; over a longer time horizon, these shifts hold potential to translate into statistically significant improvements in student learning.

For policy leaders, this evidence underscores that effective large-scale educational reform requires deliberate investment in the technical, organizational supports for teaching and learning as well as the relational work that constitutes adaptive change, developing and protecting the space for collaborative inquiry and shared leadership. The pilot suggests that schools can establish the necessary technical and adaptive changes in three years but high-needs schools are unlikely to impact student outcomes over the same time frame. Future research must now trace whether these promising shifts in professional practice—including increased consistency in collaborative routines and more coherent instructional adjustments—translate into sustained, significant gains in student achievement and attendance, thereby validating the model's viability for expansion to additional districts facing chronic underperformance.

References

- Barnes, K., & McKenzie, S. C. (2024). *Effects of PLC at Work in Arkansas on student academic outcomes*. Office for Education Policy, University of Arkansas.
https://arkansasadvocate.com/wp-content/uploads/2024/05/Effects_of_PLC_at_Work_in_AR.pdf
- Cottingham, B. W., Hough, H. J., & Myung, J. (2023, December). *What does it take to accelerate the learning of every child? Early insights from a CCEE school-improvement pilot* [Report]. Policy Analysis for California Education.
<https://edpolicyinca.org/publications/what-does-it-take-accelerate-learning-every-child>
- Cottingham, B. W., & Myung, J. (2025, August). *Improving the instructional core at scale: A case study of Chino Valley Unified School District* [Report]. Policy Analysis for California Education.
<https://edpolicyinca.org/publications/improving-instructional-core-scale>
- DuFour, R., DuFour, R., Eaker, R., Many, T. W., & Mattos, M. (2016). *Learning by doing: A handbook for Professional Learning Communities at Work* (3rd ed.). Solution Tree Press.
- DuFour, R., DuFour, R., Eaker, R., Many, T. W., Mattos, M., & Muhammad, A. (2024). *Learning by doing: A handbook for Professional Learning Communities at Work* (4th ed.). Solution Tree.
- Gallagher, H. A., Faw, L., & Cottingham, B. W. (2026, February). *How districts scale instructional improvement that lasts* [Practice brief]. Policy Analysis for California Education.
<https://edpolicyinca.org/publications/how-districts-scale-instructional-improvement-lasts>
- Hahnel, C., & Melnicoe, H. (2019, November). *The implications of Sacramento city unified's ongoing budgetary challenges for local and state policy* [Report]. Policy Analysis for California Education.
<https://edpolicyinca.org/publications/implications-sacramento-city-unifieds-ongoing-budgetary-challenges-local-and-state>
- Heifetz, R. A., & Laurie, D. L. (1997). The work of leadership. *Harvard Business Review*, 75(1), 124–134. <https://pubmed.ncbi.nlm.nih.gov/10174450/>
- Rubin, D. B. (1980). Comment on randomization analysis of experimental data: The Fisher randomization test. *Journal of the American Statistical Association*, 75(371), 591–593.
<https://doi.org/10.2307/2287653>

Talbert, J. E., & David, J. L. (2019). *Sanger Unified School District: Positive outliers case study*. Learning Policy Institute.

<https://learningpolicyinstitute.org/sanger-unified-school-district-positive-outliers-case-study>

Torres, K., Roccograndi, A., Holmgren, M., & Taylor, S. (2019). *On the road to Professional Learning Communities (PLC) at Work® in Wyoming: Year 1 evaluation*. Education Northwest.

Appendix A
Individual School Change Continuum Data

Continuum Progress: Sierra Vista Elementary School

Indicator	Year 1			Year 2			Year 3		
	Beginning	Middle	End	Beginning	Middle	End	Beginning	Middle	End
<i>Clear Purpose</i>	Implementing	Developing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Communicate purpose</i>	Implementing	Developing	Developing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining
<i>Shared Mission</i>	Implementing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Shared Vision</i>	Initiating	Implementing	Developing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining
<i>Collective Commitments</i>	Developing	Developing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Common School Goals</i>	Initiating	Developing	Developing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining
<i>Interdependent teams</i>	Implementing	Developing	Developing	Developing	Developing	Developing	Developing	Sustaining	Sustaining
<i>Norms</i>	Implementing	Developing	Developing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining
<i>Smart Goals</i>	Initiating	Implementing	Implementing	Implementing	Developing	Developing	Developing	Developing	Sustaining
<i>Determine essential learning</i>	Implementing	Developing	Developing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining
<i>Analyze data</i>	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	Sustaining	Sustaining
<i>Clarify proficiency</i>	Initiating	Implementing	Developing	Developing	Developing	Developing	Developing	Developing	Sustaining
<i>CFA's</i>	Implementing	Implementing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>RTI</i>	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	Developing	Developing
Rating Scale									
		Pre-Initiating	Initiating	Implementing	Developing	Sustaining			

Continuum Progress: Nishimoto Elementary School

Indicator	Year 1			Year 2			Year 3		
	Beginning	Middle	End	Beginning	Middle	End	Beginning	Middle	End
<i>Clear Purpose</i>	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	--	Sustaining
<i>Communicate purpose</i>	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	--	Sustaining
<i>Shared Mission</i>	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	--	Sustaining
<i>Shared Vision</i>	Implementing	Implementing	Implementing	Implementing	Implementing	Developing	Developing	--	Sustaining
<i>Collective Commitments</i>	Implementing	Implementing	Implementing	Implementing	Implementing	Developing	Developing	--	Sustaining
<i>Common School Goals</i>	Initiating	Implementing	Implementing	Implementing	Developing	Developing	Developing	--	Developing
<i>Interdependent teams</i>	Implementing	Developing	Developing	Implementing	Developing	Developing	Developing	--	Sustaining
<i>Norms</i>	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	--	Sustaining
<i>Smart Goals</i>	Pre-Initiating	Initiating	Initiating	Implementing	Developing	Developing	Developing	--	Sustaining
<i>Determine essential learning</i>	Initiating	Implementing	Implementing	Implementing	Developing	Sustaining	Sustaining	--	Sustaining
<i>Analyze data</i>	Initiating	Implementing	Implementing	Implementing	Developing	Developing	Developing	--	Sustaining
<i>Clarify proficiency</i>	Pre-Initiating	Implementing	Implementing	Implementing	Developing	Sustaining	Sustaining	--	Sustaining
<i>CFA's</i>	Pre-Initiating	Implementing	Implementing	Implementing	Developing	Developing	Developing	--	Sustaining
<i>RTI</i>	Initiating	Implementing	Implementing	Implementing	Developing	Developing	Developing	--	Developing
Rating Scale									
		Pre-Initiating	Initiating	Implementing	Developing	Sustaining			

Continuum Progress: George Washington Elementary School

Indicator	Year 1			Year 2			Year 3		
	Beginning	Middle	End	Beginning	Middle	End	Beginning	Middle	End
<i>Clear Purpose</i>	Initiating	Implementing	Developing	Developing	Developing	Developing	Developing	Developing	Sustaining
<i>Communicate purpose</i>	Implementing	Implementing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Shared Mission</i>	Pre-Initiating	Initiating	Implementing	Implementing	Implementing	Developing	Developing	Developing	Sustaining
<i>Shared Vision</i>	Pre-Initiating	Initiating	Implementing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Collective Commitments</i>	Initiating	Initiating	Implementing	Implementing	Implementing	Developing	Developing	Developing	Developing
<i>Common School Goals</i>	Initiating	Initiating	Initiating	Implementing	Implementing	Developing	Developing	Developing	Developing
<i>Interdependent teams</i>	Implementing	Developing	Developing	Developing	Developing	Developing	Developing	Developing	Developing
<i>Norms</i>	Implementing	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	Developing
<i>Smart Goals</i>	Pre-Initiating	Initiating	Initiating	Implementing	Implementing	Developing	Developing	Developing	Developing
<i>Determine essential learning</i>	Initiating	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	Developing
<i>Analyze data</i>	Initiating	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	Sustaining
<i>Clarify proficiency</i>	Pre-Initiating	Implementing	Implementing	Implementing	Developing	Developing	Developing	Developing	Developing
<i>CFA's</i>	Pre-Initiating	Implementing	Implementing	Implementing	Implementing	Developing	Developing	Developing	Developing
<i>RTI</i>	Initiating	Implementing	Implementing	Implementing	Implementing	Implementing	Implementing	Implementing	Developing
Rating Scale									
		Pre-Initiating	Initiating	Implementing	Developing	Sustaining			

Continuum Progress: Fillmore Middle School

Indicator	Year 1 2022-23			Year 2 2023-24			Year 3 2024-25		
	Beginning	Middle	End	Beginning	Middle	End	Beginning	Middle	End
<i>Clear Purpose</i>	Initiating	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Communicate purpose</i>	Implementing	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Shared Mission</i>	Pre-Initiating	Implementing	Developing	Implementing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Shared Vision</i>	Pre-Initiating	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Collective Commitments</i>	Pre-Initiating	Implementing	Developing	Developing	Developing	Developing	Developing	Developing	Sustaining
<i>Common School Goals</i>	Initiating	Developing	Developing	Pre-Initiating	Developing	Developing	Developing	Developing	Sustaining
<i>Interdependent teams</i>	Implementing	Developing	Sustaining	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Norms</i>	Pre-Initiating	Developing	Developing	Developing	Developing	Developing	Developing	Sustaining	Sustaining
<i>Smart Goals</i>	Implementing	Developing	Developing	Pre-Initiating	Developing	Developing	Developing	Developing	Developing
<i>Determine essential learning</i>	Implementing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Analyze data</i>	Initiating	Implementing	Implementing	Implementing	Developing	Developing	Developing	Developing	Sustaining
<i>Clarify proficiency</i>	Implementing	Implementing	Developing	Developing	Implementing	Developing	Developing	Sustaining	Sustaining
<i>CFA's</i>	Implementing	Developing	Developing	Developing	Developing	Developing	Developing	Developing	Sustaining
<i>RTI</i>	Pre-Initiating	Pre-Initiating	Implementing	Developing	Developing	Developing	Developing	Developing	Developing
Rating Scale									
		Pre-Initiating	Initiating	Implementing	Developing	Sustaining			

Continuum Progress: HW Harkness Elementary School

Indicator	Year 1			Year 2			Year 3		
	Beginning	Middle	End	Beginning	Middle	End	Beginning	Middle	End
<i>Clear Purpose</i>	Pre-Initiating	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Communicate purpose</i>	Pre-Initiating	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Shared Mission</i>	Initiating	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Shared Vision</i>	Initiating	Developing	Sustaining	Sustaining	Pre-Initiating	Pre-Initiating	Developing	Developing	Developing
<i>Collective Commitments</i>	Pre-Initiating	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Common School Goals</i>	Pre-Initiating	Implementing	Implementing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Interdependent teams</i>	Implementing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Norms</i>	Pre-Initiating	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Smart Goals</i>	Pre-Initiating	Implementing	Implementing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Determine essential learning</i>	Pre-Initiating	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Analyze data</i>	Initiating	Initiating	Implementing	Developing	Sustaining	Sustaining	Sustaining	Sustaining	Sustaining
<i>Clarify proficiency</i>	Pre-Initiating	Pre-Initiating	Pre-Initiating	Implementing	Developing	Developing	Developing	Sustaining	Sustaining
<i>CFA's</i>	Initiating	Initiating	Implementing	Developing	Developing	Developing	Developing	Sustaining	Sustaining
<i>RTI</i>	Pre-Initiating	Pre-Initiating	Developing	Developing	Developing	Developing	Developing	Developing	Developing
Rating Scale									
		Pre-Initiating	Initiating	Implementing	Developing	Sustaining			

Continuum Progress: Elder Creek Elementary School

Indicator	Year 1			Year 2			Year 3		
	Beginning	Middle	End	Beginning	Middle	End	Beginning	Middle	End
<i>Clear Purpose</i>	Initiating	Developing	Developing	Developing	Developing	Developing	Developing	Developing	Sustaining
<i>Communicate purpose</i>	Initiating	Initiating	Implementing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Shared Mission</i>	Initiating	Developing	Developing	Developing	Developing	Developing	Developing	Developing	Sustaining
<i>Shared Vision</i>	Initiating	Implementing	Developing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Collective Commitments</i>	Pre-Initiating	Initiating	Implementing	Initiating	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Common School Goals</i>	Pre-Initiating	Initiating	Initiating	Initiating	Implementing	Developing	Developing	Developing	Developing
<i>Interdependent teams</i>	Pre-Initiating	Implementing	Implementing	Developing	Developing	Developing	Developing	Developing	Sustaining
<i>Norms</i>	Pre-Initiating	Pre-Initiating	Initiating	Implementing	Developing	Developing	Developing	Developing	Sustaining
<i>Smart Goals</i>	Pre-Initiating	Pre-Initiating	Pre-Initiating	Pre-Initiating	Implementing	Implementing	Implementing	Implementing	Developing
<i>Determine essential learning</i>	Pre-Initiating	Implementing	Implementing	Developing	Developing	Sustaining	Sustaining	Sustaining	Sustaining
<i>Analyze data</i>	Initiating	Initiating	Initiating	Developing	Developing	Developing	Developing	Developing	Developing
<i>Clarify proficiency</i>	Pre-Initiating	Pre-Initiating	Pre-Initiating	Implementing	Implementing	Implementing	Implementing	Implementing	Developing
<i>CFA's</i>	Pre-Initiating	Pre-Initiating	Pre-Initiating	Implementing	Implementing	Developing	Developing	Developing	Developing
<i>RTI</i>	Pre-Initiating	Pre-Initiating	Pre-Initiating	Implementing	Developing	Developing	Developing	Developing	Developing
Rating Scale									
		Pre-Initiating	Initiating	Implementing	Developing	Sustaining			

Continuum Progress: Highland Elementary School

Indicator	Year 1			Year 2			Year 3		
	Beginning	Middle	End	Beginning	Middle	End	Beginning	Middle	End
<i>Clear Purpose</i>	Initiating	Implementing	Implementing	Implementing	Implementing	Developing	Developing	N/A	Developing
<i>Communicate purpose</i>	Initiating	Implementing	Developing	Implementing	Developing	Sustaining	Sustaining	N/A	Sustaining
<i>Shared Mission</i>	Pre-Initiating	Initiating	Initiating	Initiating	Implementing	Developing	Developing	N/A	Developing
<i>Shared Vision</i>	Initiating	Implementing	Developing	Implementing	Developing	Developing	Developing	N/A	Developing
<i>Collective Commitments</i>	Pre-Initiating	Pre-Initiating	Pre-Initiating	Pre-Initiating	Initiating	Implementing	Implementing	N/A	Developing
<i>Common School Goals</i>	Initiating	Initiating	Initiating	Initiating	Initiating	Initiating	Initiating	N/A	Developing
<i>Interdependent teams</i>	Initiating	Initiating	Initiating	Developing	Developing	Developing	Developing	N/A	Developing
<i>Norms</i>	Pre-Initiating	Initiating	Initiating	Initiating	Implementing	Implementing	Implementing	N/A	Developing
<i>Smart Goals</i>	Pre-Initiating	Pre-Initiating	Pre-Initiating	Pre-Initiating	Initiating	Implementing	Implementing	N/A	Implementing
<i>Determine essential learning</i>	Pre-Initiating	Initiating	Initiating	Implementing	Developing	Developing	Developing	N/A	Developing
<i>Analyze data</i>	Pre-Initiating	Pre-Initiating	Pre-Initiating	Pre-Initiating	Initiating	Developing	Developing	N/A	Developing
<i>Clarify proficiency</i>	Pre-Initiating	Initiating	Initiating	Initiating	Implementing	Developing	Developing	N/A	Developing
<i>CFA's</i>	Pre-Initiating	Initiating	Initiating	Initiating	Implementing	Developing	Developing	N/A	Developing
<i>RTI</i>	Pre-Initiating	Pre-Initiating	Pre-Initiating	Pre-Initiating	Pre-Initiating	Implementing	Implementing	N/A	Developing
Rating Scale									
		Pre-Initiating	Initiating	Implementing	Developing	Sustaining			

Appendix B Description of Source Data

All data are from the California Department of Education (CDE) and are publicly available. Schools are identified across files using the 14-digit County–District–School (CDS) code. The 2019–20 school year is excluded from all analyses.

CAASPP assessment results. California Assessment of Student Performance and Progress. School-level files provide mean scale scores and proficiency rates in math and ELA for grades 3–8. We retain school-aggregate records and reshape to wide format by subject. Proficiency rates are expressed as proportions (0–1).

Attendance. CDE student attendance data files provide school-level average days absent per student. Attendance rate is calculated as $1 - (\text{average days absent} \div 180)$.

Chronic absenteeism. CDE chronic absenteeism files provide school-level rates and counts of students missing 10% or more of instructional days. The rate is expressed as a proportion (0–1).

Enrollment. CDE cumulative enrollment files (``cenrollYYYY.txt``) provide total school enrollment. We retain school-level totals (``ReportingCategory == "TA"``).

Treatment assignment. IAM schools are those designated as IAM campuses beginning in the 2022–23 school year in the five participating districts. Schools are uniquely identified, and linked to state administrative data files, using the 14-digit county-district-school (CDS) code.

Data Availability by Outcome and School Year

The table below summarizes whether each outcome variable is available for each school year in the study period. The 2019–20 school year is absent from all outcome measures due to COVID-19 school closures and the suspension of statewide testing and attendance reporting. Covariate data (enrollment, demographics, FRPM eligibility) are available for all years including 2019–20, but that year is excluded from all analyses regardless.

Table B1. Data availability by outcome and school year

Outcome	2016–17	2017–18	2018–19	2019–20	2020–21	2021–22	2022–23	2023–24	2024–25
Math scale score	✓	✓	✓	—	—	✓	✓	✓	✓
ELA scale score	✓	✓	✓	—	—	✓	✓	✓	✓
% Proficient, math	✓	✓	✓	—	—	✓	✓	✓	✓
% Proficient, ELA	✓	✓	✓	—	—	✓	✓	✓	✓

Outcome	2016–17	2017–18	2018–19	2019–20	2020–21	2021–22	2022–23	2023–24	2024–25
Attendance rate	—	✓	✓	—	✓	✓	✓	✓	✓
Chronic absenteeism rate	✓	✓	✓	—	✓	✓	✓	✓	✓
<i>Note: ✓ = data available; — = data not available (COVID-19 disruption).</i>									

Appendix C

Description of the Weighting Method

Entropy Weights

To improve the comparability of IAM and comparison schools on observed baseline characteristics, we construct comparison group weights using entropy balancing (Hainmueller, 2012), implemented in Stata via the `ebalance` package (Hainmueller and Xu, 2013).

Entropy balancing assigns a weight w_i to each comparison school $i \in \mathcal{C}$ by solving the following constrained optimization problem:

$$\min_{\{w_i\}} \sum_{i \in \mathcal{C}} w_i \log \left(\frac{w_i}{q_i} \right)$$

subject to:

$$\sum_{i \in \mathcal{C}} w_i X_{ij} = \bar{X}_j^{\text{IAM}} \quad \forall j, \quad \sum_{i \in \mathcal{C}} w_i = 1, \quad w_i \geq 0$$

where $q_i = 1 / |\mathcal{C}|$ is the base (uniform) weight, X_{ij} is the value of covariate j for comparison school i , and $\bar{X}_j^{\text{IAM}}$ is the corresponding mean among IAM schools.

The objective minimizes the Kullback–Leibler divergence of the solution weights from the uniform base weights — that is, weights deviate from uniform only as much as necessary to satisfy the balance constraints. The first-moment constraints guarantee that the weighted comparison group means *exactly* match IAM school means for every covariate included in the algorithm, rather than achieving only approximate balance as in propensity score approaches. This exact balancing is evident in the balance figure embedded below, where all standardized mean differences (SMDs) are zero. Zhao and Percival (2017) show that entropy balancing estimators are doubly robust: the treatment effect estimate is consistent if either the weighting model or the outcome regression model is correctly specified.

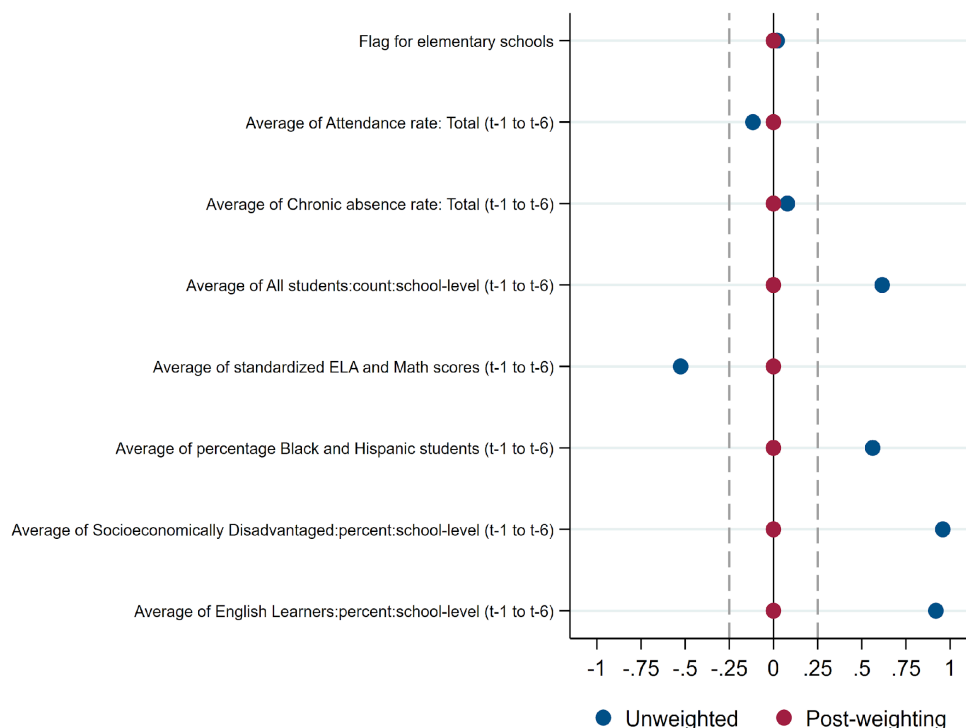
The following covariates were included as balancing constraints:

- Prior math and ELA scale scores (rolling averages over up to 6 pre-period years)
- Prior attendance and chronic absenteeism rates (rolling averages over up to 6 pre-period years)
- Student enrollment counts
- Share of students: economically disadvantaged, English learners
- Racial/ethnic composition: percentage of Black and Hispanic students
- School configuration (elementary vs. middle)

Balance Diagnostics

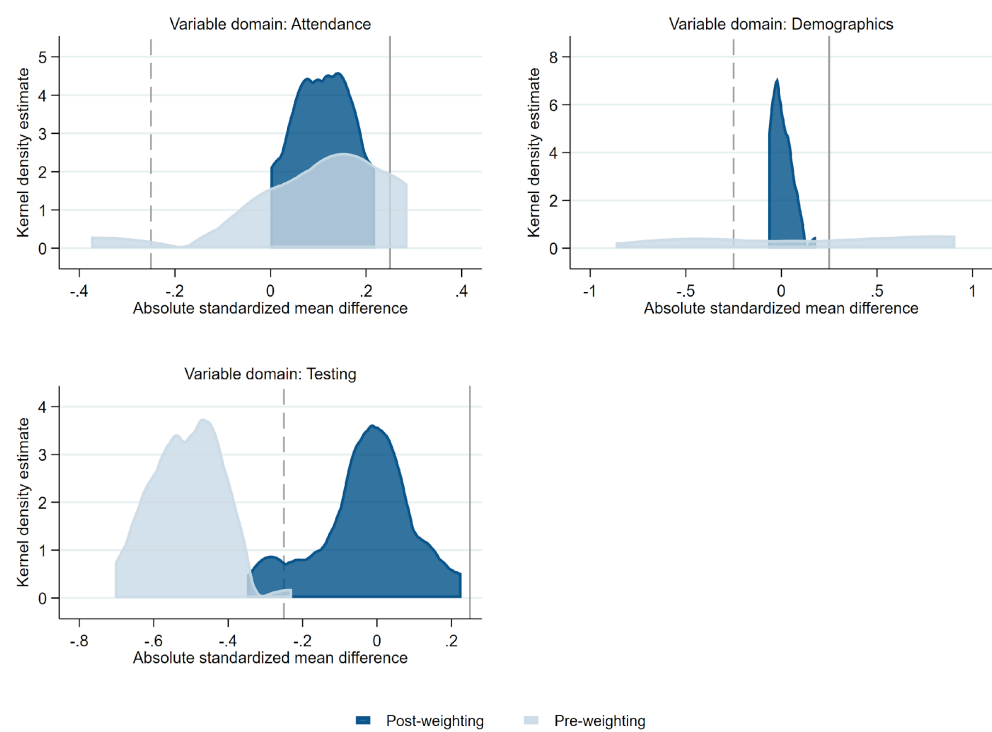
After applying entropy weights, IAM schools and their comparison counterparts are closely comparable on baseline characteristics. The table below reports standardized mean differences (SMDs) for the weighted sample. The pre- and post-weighting differences between IAM and non-IAM schools illustrate the efficacy of the entropy balancing procedure.

Figure C1: Love plot, weighting covariates



Entropy balancing mechanically produces weights that provide exact balance between groups for the covariates included in the algorithm. However, the weighting procedure may yield weights that have unintended, deleterious effects on the balance between groups for covariates **not included in the model used to construct the weights**. To explore this potential threat to equivalence between IAM and non-IAM schools on other covariates that were not included in the algorithm, we plot the distribution of nearly thirty covariates (including up to six lags for each demographic and outcome measure) using kernel density estimation. The figure below provides the distribution of SMDs across all available covariates.

Figure 15. Kernel density plot, all covariates



Appendix D

Description of the Estimation Method

Pooled Difference-in-Differences

The primary estimator regresses school-by-year outcomes on a treatment–post interaction alongside school (α_i) and year fixed effects (λ_t) (1). The model was fit separately by outcome and, for testing outcomes, by subject.

$$Y_{it} = \alpha_i + \lambda_t + \beta \cdot \text{Treat}_i \times \text{Post}_t + \varepsilon_{it}$$

where Y_{it} represents a school-level outcome in school i at time t . Testing outcomes were standardized by subject, grade, and year. β is the average treatment effect on the treated (ATT) — the average difference in outcomes between IAM and comparison schools in post-launch years, after accounting for baseline school differences through weighting. Standard errors are clustered at the school level to account for serial correlation within schools over time.

Event Study (Dynamic Effects)

The specification in equation (1) assumes that the effect of the IAM intervention is constant across all post-intervention periods and ignores potential threats to the identification of an unbiased impact estimate introduced by different pre-intervention outcome trends between IAM and non-IAM schools and students. Put differently, a key assumption underpinning model (1) is that in the absence of IAM implementation, post-intervention outcomes would be the same between IAM and comparison schools and, thus, any deviations from the underlying performance trajectory that occurs is due to the introduction of IAM. This is a strong assumption that, in an observational setting cannot be assumed based on randomization. However, with multiple years of baseline information, we can examine the presence/absence of pre-trends that may misattribute differential changes in the outcome before and after implementation, between IAM and non-IAM schools, to differences in secular trends between the two groups. To address both shortcomings, we estimated dynamic effects using event-study methods described in the functional form formalized in equation (2):

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \gamma_k \cdot \text{Treat}_i \times 1[t - T_0 = k] + \varepsilon_{it}$$

k indexes years relative to IAM's launch ($T_0 = 2022\text{--}23$), ranging from $k = -6$ to $k = +2$. The year immediately before launch ($k = -1$) is the omitted reference period; all coefficients are interpreted relative to that baseline. Pre-launch coefficients ($k < 0$) near zero support parallel trends; post-launch coefficients ($k \geq 0$) trace the dynamic treatment path.

The advantage of this specification lies in the ability to estimate dynamic treatment effects before and after IAM implementation. The trajectory of the baseline treatment effects evaluates the validity of the assumption of parallel trends before the introduction of the

program — these effects should be null — and the evolution of the post-implementation effects relaxes the assumption of constant effects of exposure to IAM, allowing the research team to estimate whether the impacts of the program grow or decay over time.

Appendix E

Evidence of Parallel Trends

A key requirement of the evaluation's identification strategy is that IAM and comparison schools would have followed parallel outcome trajectories in the absence of the program. The figures below plot school-year means for both groups from 2016-17 through 2024-25; the vertical line marks the 2022-23 launch of IAM.

Figure E1: Percentage at or above proficiency, math

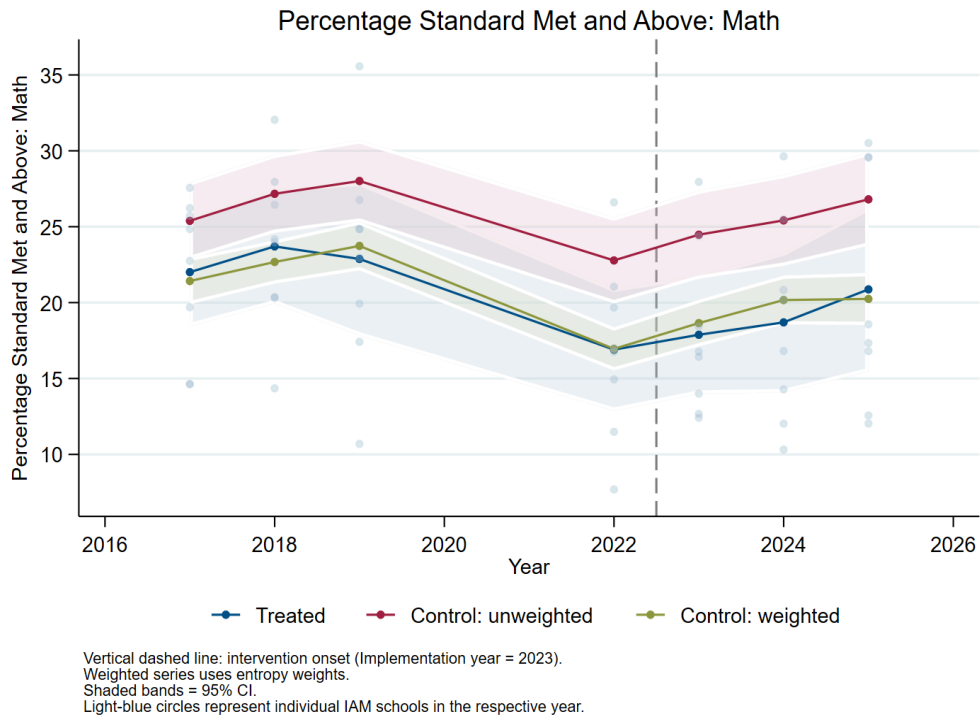


Figure E2. Standardized scale scores, math

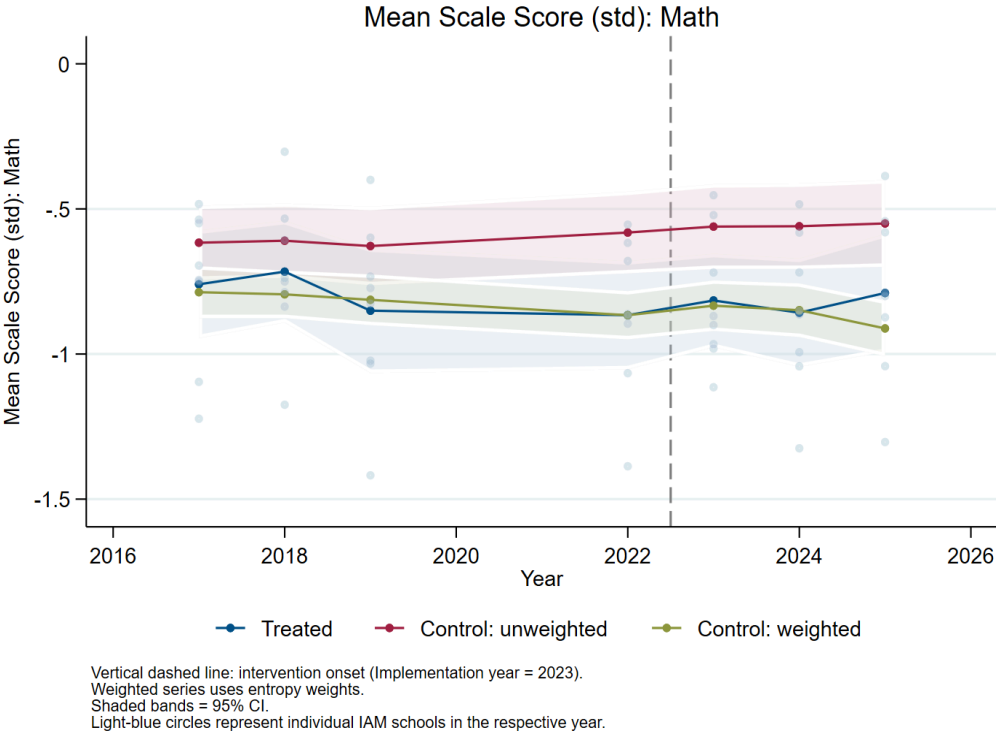


Figure E3. Percentage at or above proficiency, ELA

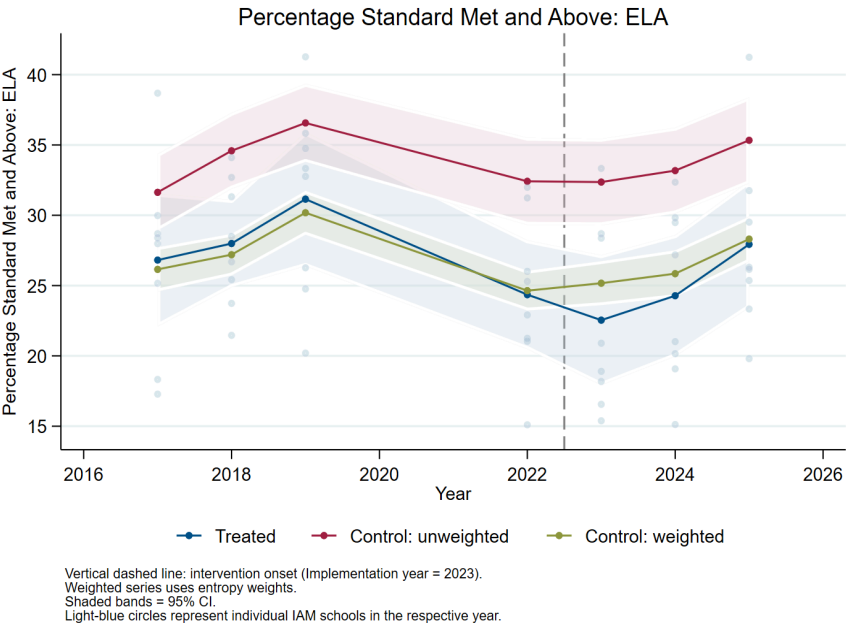


Figure E4. Standardized scale scores, ELA

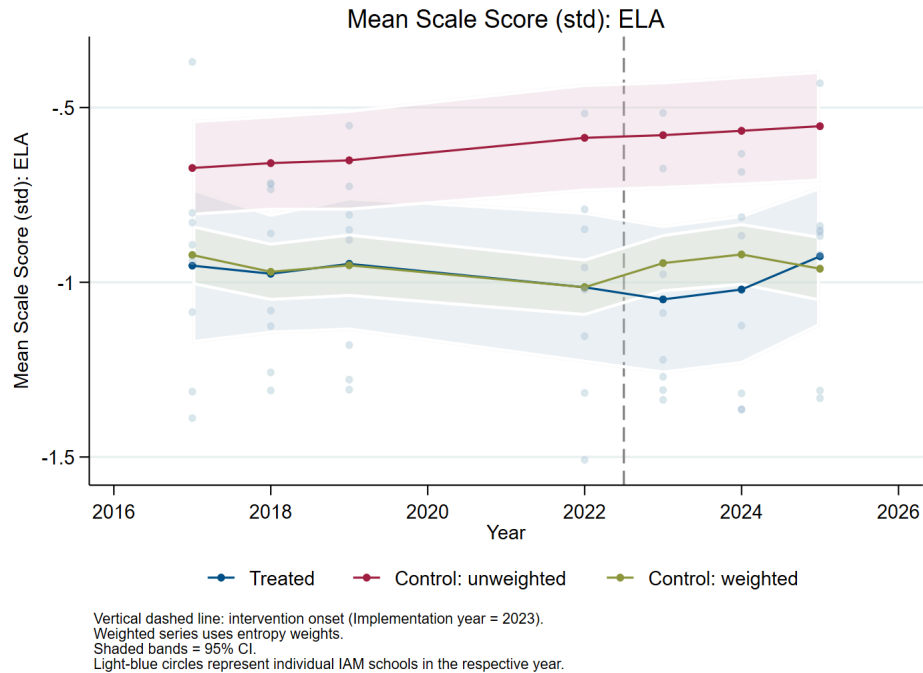


Figure E5: Chronic absenteeism

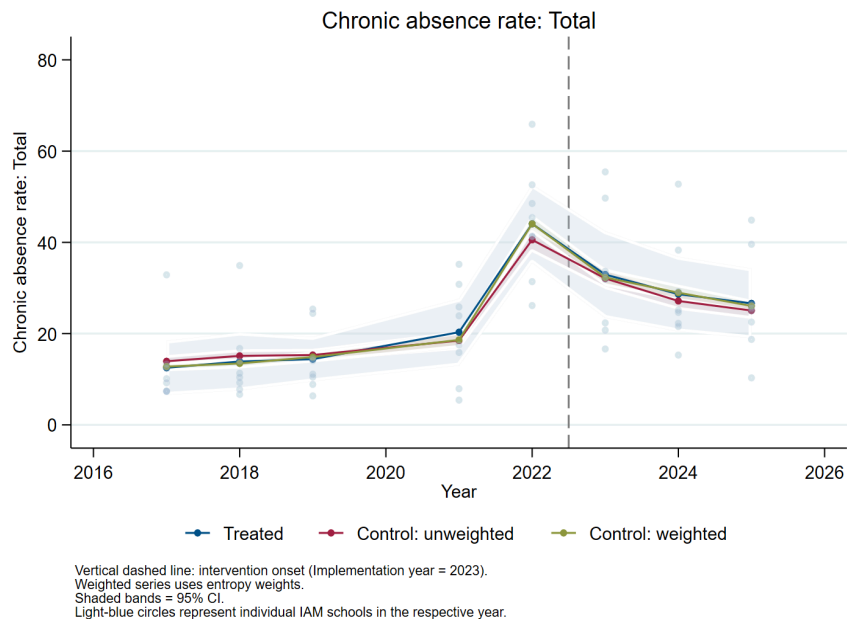
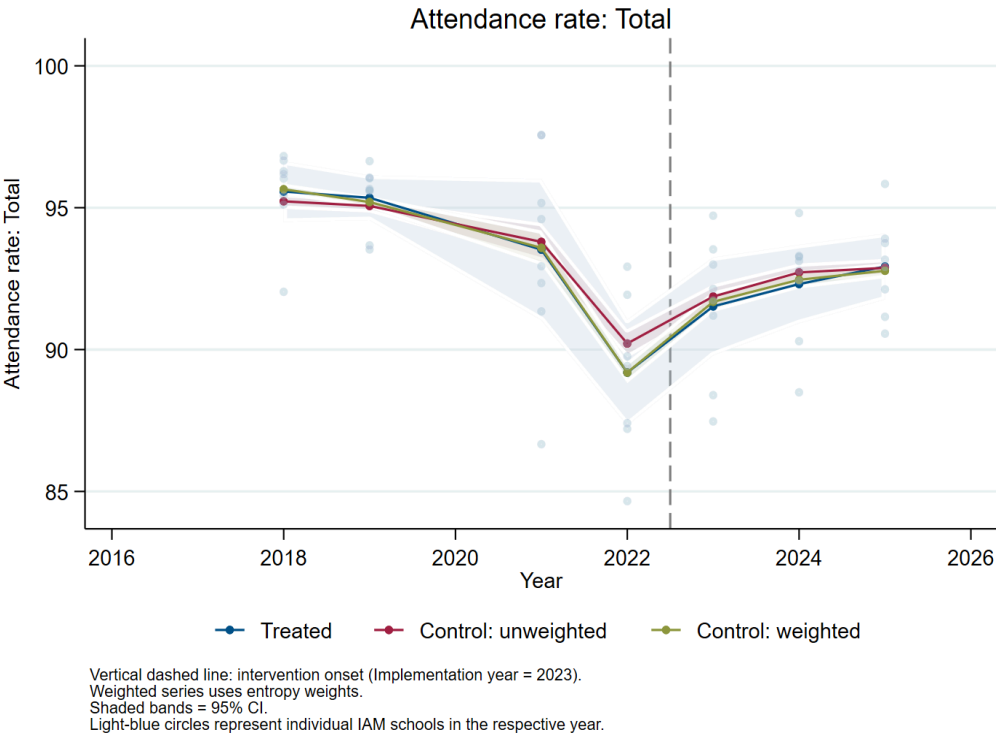


Figure E6. Attendance rates



Appendix F

Sensitivity Analyses Under Alternative Specifications

The figure below plots post-period event-study coefficients for all six outcomes across three implementation years (2022/23 through 2024/25) under six specifications:

- **Primary:** The main analysis — entropy-balanced comparison sample restricted to schools in IAM pilot districts, event-study estimates.
- **Full sample:** All eligible public elementary and middle schools in California, entropy-balanced
- **PS-matched:** Propensity-score-matched comparison group in place of entropy balancing
- **Time-varying controls:** The primary specification augmented with time-varying school-level controls (enrollment, demographic shares, lagged outcomes)
- **CITS:** A comparative interrupted time series model that adds a treated-school-specific linear time trend to the primary specification, allowing for differential pre-existing trends
- **NN:** Nearest-neighbor Mahalanobis-matched comparison group in place of entropy balancing, selecting the ten-nearest neighbors for each IAM school

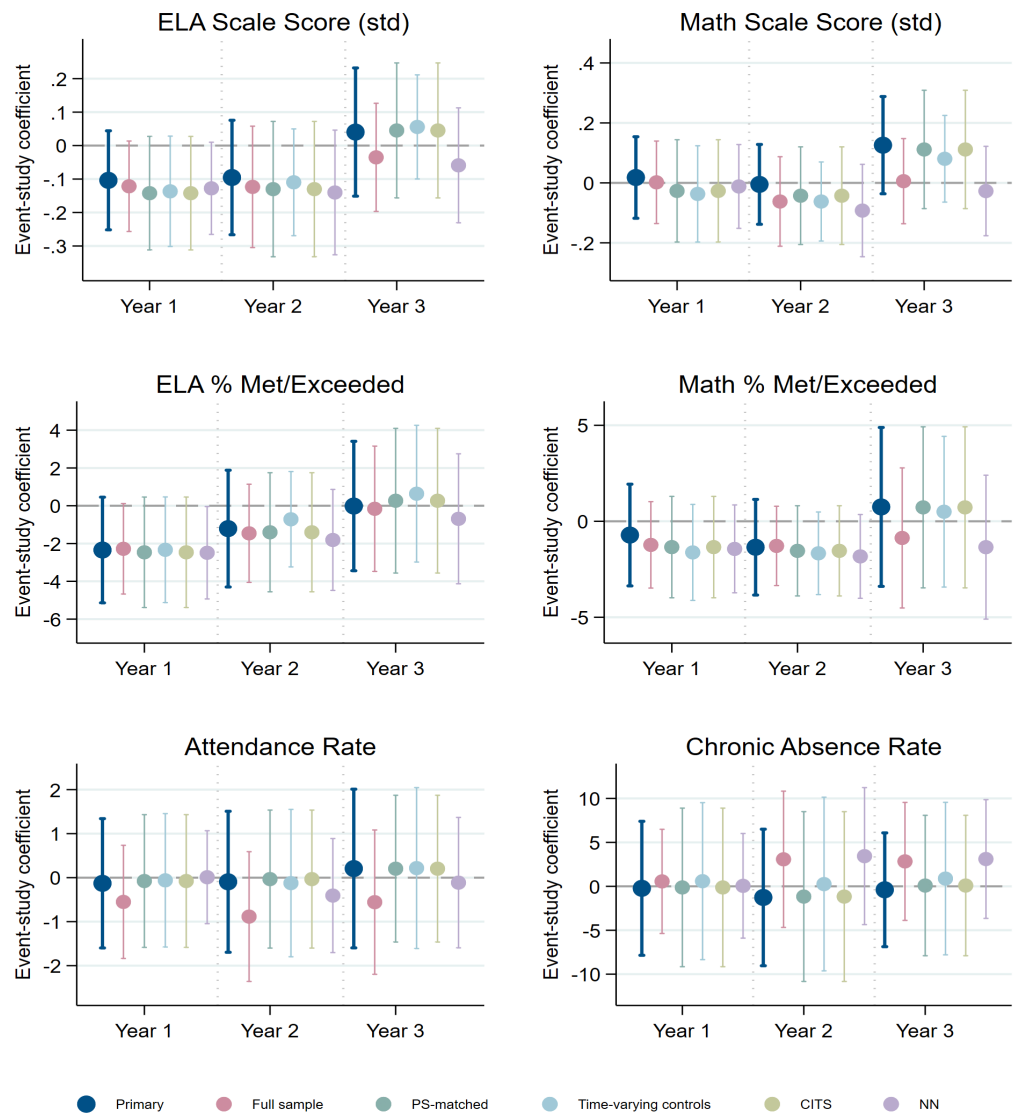
Several patterns emerge from the specification curve. For **math outcomes**, the cross-specification pattern is relatively consistent: estimates across most specifications are near zero or modestly negative in the first two implementation years, then turn positive by Year 3. This trajectory holds across the primary specification and most analytical variants; the alternative comparison-group specifications (full sample, PS-matched, and nearest-neighbor) diverge most in Year 3, widening the curve at that horizon. No specification produces statistically significant effects in any year, but the directional consistency of the negative-to-positive trajectory across the majority of specifications is noteworthy.

For **ELA outcomes**, the curve is wider and less directionally consistent across years, with estimates more sensitive to comparison-group construction and the counterfactual trend assumption.

For **attendance outcomes**, the picture is mixed across implementation years. In Year 1, all specifications produce estimates clustered tightly near zero — the narrowest and most consistent pattern in the figure. Cross-specification divergence increases in Years 2 and 3, particularly for chronic absenteeism, where some specifications produce meaningfully positive estimates while others remain near zero and confidence intervals widen substantially. The attendance null result is most clearly supported in the first implementation year; conclusions about Years 2 and 3 are less stable across the specification curve.

Across all outcomes and specifications, no estimate is statistically distinguishable from zero at conventional thresholds.

Figure F1. Specification curve: post-period event-study estimates across six analytical specifications

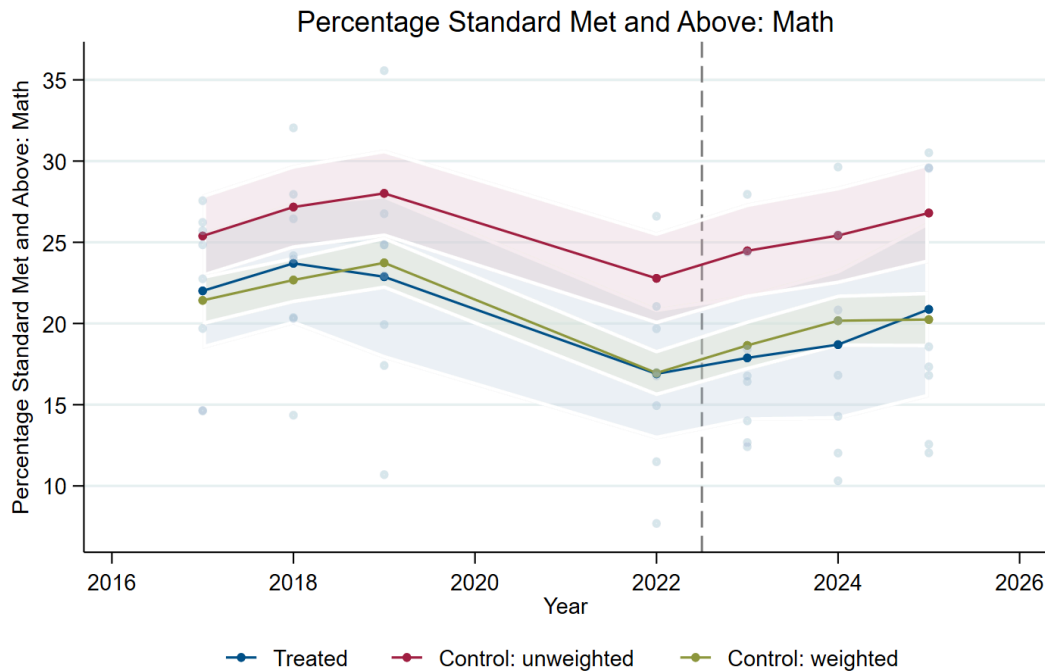


Appendix G Pre-Implementation Outcome Trajectories

Each figure below plots the school-year mean outcome for IAM schools (treated), the unweighted comparison pool, and the entropy-weighted comparison pool from 2016/17 through 2024/25. The vertical dashed line marks IAM's launch in 2022/23. Shaded bands are 95% confidence intervals; light blue dots represent individual IAM schools. The figures serve two purposes: they illustrate how entropy weighting improves comparability between groups on the outcome itself prior to IAM, and they provide a visual basis for assessing whether pre-implementation trends were roughly parallel between treated and weighted comparison schools. Outcomes are grouped by domain below.

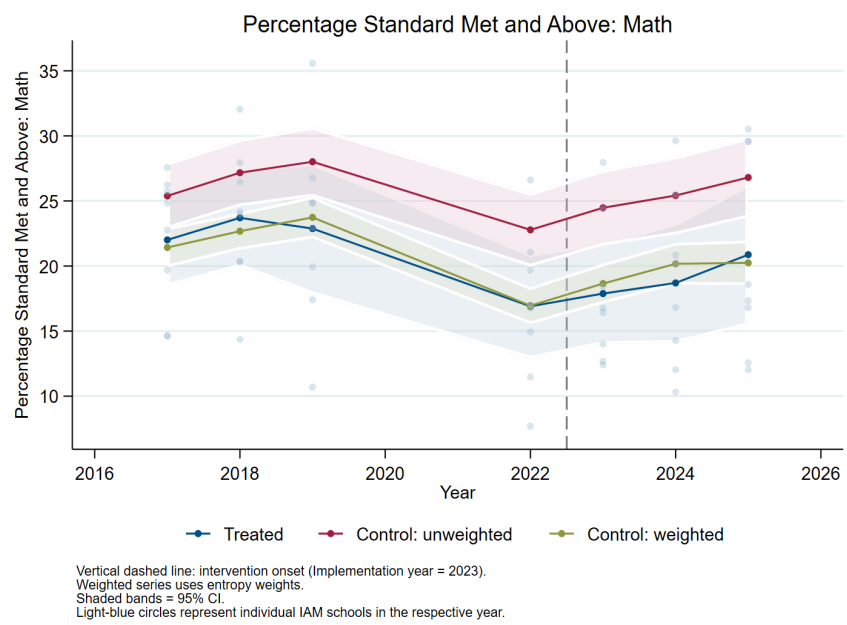
Academic Achievement: Math

Figure G1. Math standardized scale score



Vertical dashed line: intervention onset (Implementation year = 2023).
 Weighted series uses entropy weights.
 Shaded bands = 95% CI.
 Light-blue circles represent individual IAM schools in the respective year.

Figure G2. Math proficiency rate (% Met or Exceeded)



Academic Achievement: ELA

Figure G3. ELA standardized scale score

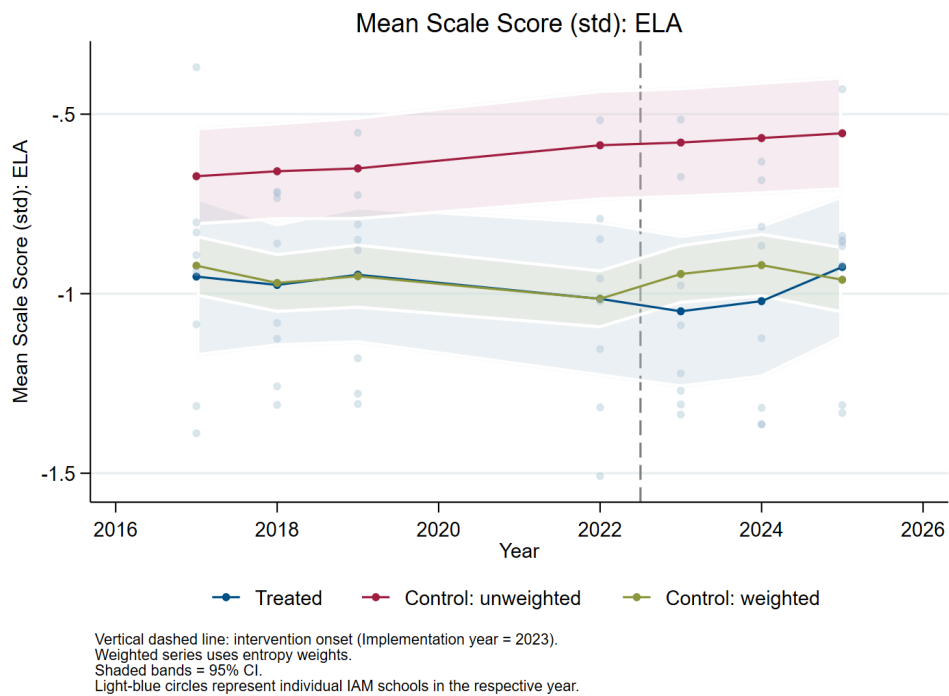
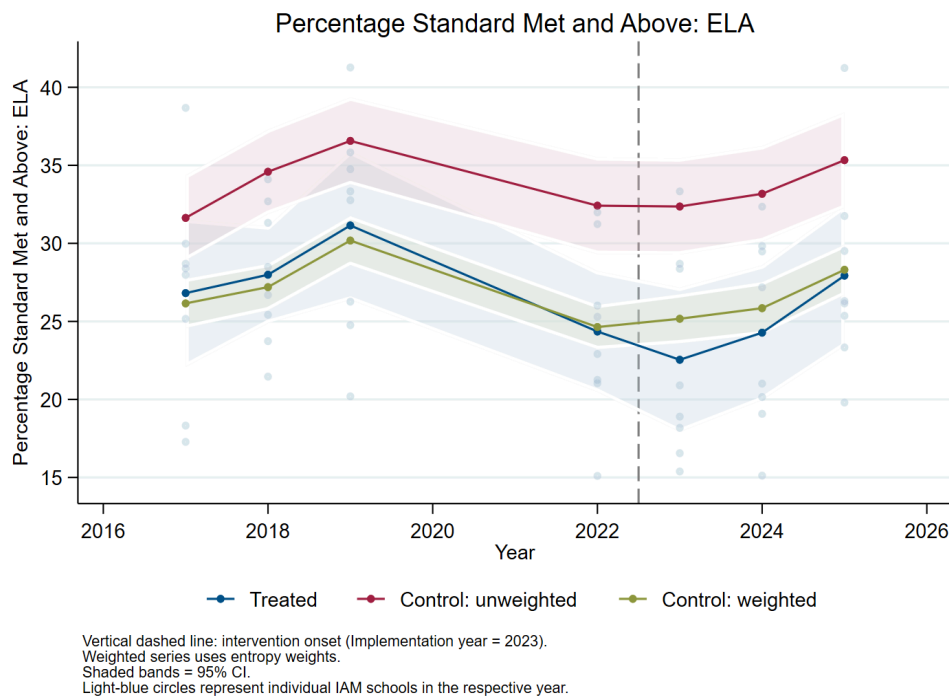


Figure G4. ELA proficiency rate (% Met or Exceeded)



Attendance

Figure G5. Attendance rate

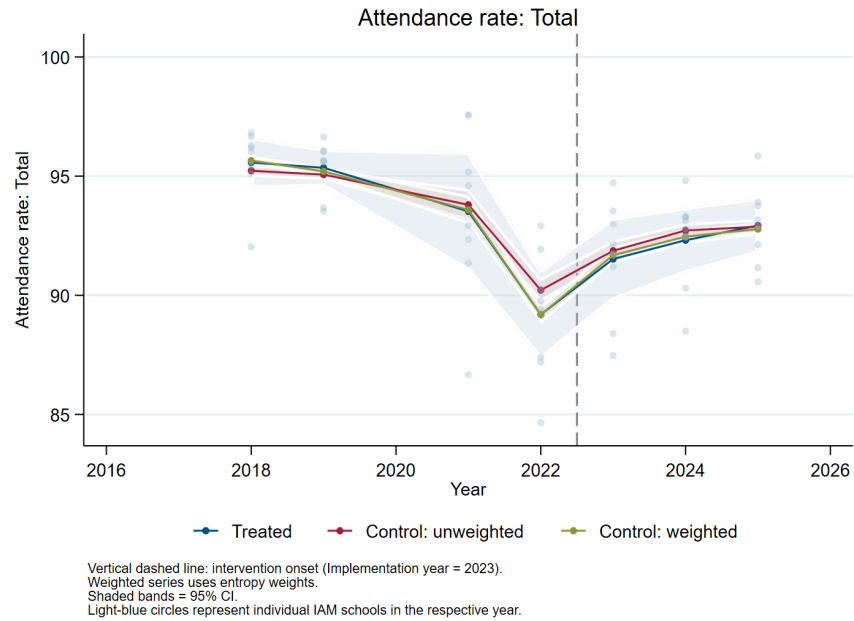


Figure G6. Chronic absenteeism rate

