

Designing Effective Curriculum-Based Professional Learning

Insights from a Cross-Study Portfolio

Stacey Alicea
Executive Director
Research Partnership for Professional Learning

Arielle Boguslav
Research Scientist
Annenberg Institute at Brown University
Research Partnership for Professional Learning

Camea Davis
Director of Equitable Research Practice Partnerships
Research Partnership for Professional Learning

John Papay
Director of Annenberg Institute at Brown University
Research Partnership for Professional Learning

Lorna Porter
Senior Research Associate
Annenberg Institute at Brown University
Research Partnership for Professional Learning

Poorvaja Sundar
Project Manager
Research Partnership for Professional Learning

Nathaniel Schwartz
Director of Applied Research for Annenberg Institute
at Brown University Research Partnership for
Professional Learning

Nitika Tolani
Managing Director of Learning & Engagement
Research Partnership for Professional Learning

Table of Contents

01

4 – EXECUTIVE SUMMARY



02

7 – STUDY SUMMARIES

9 – Anticipating Common Misconceptions: Enhancing Equitable Math Instruction Through Guided Adaptation of HQIM

13 – Exploring the Impact of Thinking Routines and Professional Learning Delivery Models on Teacher and Student Outcomes in Mathematics Education

17 – Unlocking the Potential of EL Education's Math Approach to Create Equitable, Capable, and Courageous Math Communities for All

21 – Formative Feedback Study: Enhancing Observation-Based Feedback for Coaching Conversations



03

25 – CROSS-CUTTING EMERGING INSIGHTS AND DESIGN IMPLICATIONS

31 – What We're Still Learning

32 – Endnotes



Introduction

Purpose

This report synthesizes emerging and actionable insights across four studies funded by the Research Partnership for Professional Learning (RPPL) to identify which professional learning (PL) designs and conditions improve teaching practice and student outcomes.

Overview of RPPL's Request for Proposals (RFP)

RPPL's request for proposals for the studies aimed to advance the evidence base on curriculum-based professional learning (CBPL) in equitable math instruction supporting every student, including Black, Latine, Indigenous, AAPI, multilingual learners, students with disabilities, and those experiencing poverty. These studies were expected to investigate:

- Existing PL practices to examine how specific design features—particularly those aligned to High Quality Instructional Materials (HQIM)—influenced effective teacher practice and student learning
- Contextual factors that influenced the successful use of HQIM in school districts.

In addition, the studies were planned in partnership with [RPPL's network](#) of PL organizations, researchers, funders, and system leaders, aligned with RPPL's [Learning Agenda](#), and included an element from [Building Better PL: How to Strengthen Teacher Learning](#).

Acknowledgment Statement

We are grateful to the PL providers, researchers, and educators whose participation and partnership made this work possible. We also thank the RFP reviewers for their thoughtful feedback and insights, which helped strengthen and refine this synthesis.

This work was supported by the Gates Foundation. The findings and conclusions presented in this report are those of the authors and do not necessarily reflect the views or policies of the Foundation.

Executive Summary

S.01

Executive Summary



In 2022, the Research Partnership for Professional Learning (RPPL) released Building Better PL: How to Strengthen Teacher Learning, a brief that summarized available research evidence about promising features of professional learning (PL) design for teachers that influence classroom instruction and student outcomes. This report follows up on the next steps needed in research, outlined in the previous brief, to deepen our collective knowledge about curriculum-based professional learning (CBPL) in equitable math instruction.

We synthesize emerging findings from four studies focused on the implementation of high-quality instructional materials (HQIM) in mathematics, selected for their related findings and implications across multiple approaches.

The PL designs studied were developed and delivered by a series of PL provider organizations in partnership with school district leadership staff. The findings of this brief are most relevant for PL providers, including PL provider organizations and school districts that design and deliver their own PL for educators. We note what findings may apply differently to district-led versus externally provided PL.

Collectively, these studies suggest the efficacy of: (1) peer collaboration that brings teachers together to make concrete, instructional decisions; (2) coaching that is anchored in common instructional frameworks, data, and classroom artifacts to support PL consistency and effectiveness; and (3) adaptation of HQIM grounded in guided analysis of student learning data.

Peer collaboration around concrete instructional decisions

- **What the Research Tells Us:** For peer collaboration to translate into real shifts in practice, teachers need collaborative opportunities that foster joint work around specific shared practices that lead to aligned instructional decisions.
- **Design Implications:** PL providers should offer opportunities for collaborative decision-making around specific elements of instruction. They should also thoughtfully group teachers using the same curriculum, objectives, and grade bands so they can engage in joint instructional decision-making.

Coaching anchored in common frameworks and artifacts

- **What the Research Tells Us:** Effective coaches facilitate teacher reflection and offer feedback in ways that result in shared goals for improvement. Instructional frameworks, data, and classroom artifacts can help facilitate this alignment by providing a common vision of effective instruction.
- **Design Implications:** PL providers should consider which instructional frameworks, data, and classroom artifacts will best communicate clear guidance for instructional improvement, noting when it may be helpful to use multiple artifacts together.

HQIM adaptation driven by data on student misconceptions

- **What the Research Tells Us:** Data that is specific, student-centered, and embedded within instructional routines offers targeted information about curriculum adaptations that is most likely to prove meaningful to both teachers and students.
- **Design Implications:** PL providers should consider what kind of student data (e.g., student work samples and patterns in student thinking) will be most useful for guiding adaptation decisions and evaluating the impacts of those decisions.

Across the varying PL models, the most promising approaches did not simply provide teachers with knowledge or strategies. They created structures that helped educators make disciplined instructional decisions—together, with evidence, in ways that remained aligned to the goals of HQIM.

Continue reading for more in depth analysis of and findings from each study.

Study Summaries

S.02 Studies At a Glance

STUDY TITLE AND ORGANIZATION	Anticipating Common Misconceptions: Enhancing Equitable Math Instruction Through Guided Adaptation of HQIM Achievement Network (ANet)	Exploring the Impact of Thinking Routines and PL Delivery Models on Teacher and Student Outcomes in Mathematics Education Throughline Learning	Unlocking the Potential of EL Education's Math Approach to Create Equitable, Capable, and Courageous Math Communities for All EL Education, Math Culture Lab, and National Implementation Research Network (NIRN)	Enhancing Observation-Based Feedback for Coaching Conversations University of Virginia, Annenberg Institute at Brown University, and Teach For America
FOCUS	Identify impact of PL on guided adaptation of HQIM to address student math misconceptions on teacher and student outcomes.	Identify impact of PL on specific cognitive routines to bridge gaps in student problem-solving within HQIM on teacher and student.	Build an understanding of how PL on building relationships with students and equitable teaching practices to amplify HQIM shapes school-wide math instructional practices.	Impacts of different formats for communicating classroom observation feedback to teachers on teacher perceptions, coaching conversations, and instructional improvement.
STUDY DESIGN	School-level RCT: Schools were randomly assigned to the treatment or control groups, with teachers then receiving the intervention.	School-level RCT: Schools were randomly assigned to the treatment or control groups, with teachers then receiving the intervention.	Mixed-methods case study: Case study of program implementation and impacts in one school.	Coach-level RCT: Multi-treatment design with coaches randomized to provide teachers feedback through one of three conditions.
TREATMENT AND CONTROL PL DESIGN	Control: Teachers received group PL sessions on student misconceptions and strategies to address them. Treatment: Teachers received group PL similar to the control group, but with an additional focus on guided adaptation of HQIM to address student misconceptions	Control: Business as usual. Treatment: Teachers receive PLCs & 1:1 coaching sessions focused on the 7-step thinking routines protocol.	Control: No control group. Treatment: Multi-pronged PL model, including whole-school PL and targeted support for Math Lead focused on building their capacity to lead PL group sessions and coaching for teachers on student relationships and math identity.	Control: Business-as-usual descriptive feedback. Treatment 1: Coach provides feedback using numeric score reports. Treatment 2: Coach provides feedback using numeric score reports with comparative data.
TOPLINE FINDINGS	PL that incorporated guided adaptation of HQIM—grounded in student misconception data—suggests improvements in student outcomes more than PL focused on misconceptions alone.	PL delivered via PLCs & 1:1 coaching sessions and focused on the 7-step thinking routines protocol offers an example of improving teacher confidence and students' problem-solving skills in comparison with teachers who did not receive the PL.	PL that addressed system-level reform and classroom practice was not associated with sustained changes in math culture, nor growth in student outcomes.	Feedback grounded in observation data was associated with high teacher and coach satisfaction.
EVIDENCE CLASSIFICATION	Tier 3: Promising Evidence	Tier 4: Demonstrates a Rationale	This mixed-methods case study was designed to provide rich insight into implementation and school-level instructional change efforts rather than causal impact estimation.	Tier 2: Moderate Evidence (Based on preliminary mid-year data)

Learn more about each study's design, sample, and measures [here](#).

RPPL is adapting the ESSA Tiers of Evidence framework to classify the strength of evidence from each study. This adaptation involves RPPL's assessment of study rigor and sample size, as well as the significance of the outcomes.

Anticipating Common Misconceptions: Enhancing Equitable Math Instruction Through Guided Adaptation of HQIM

Achievement Network (ANet)

Brandi Phillips | Director Math Content Learning, Achievement Network

Patrick Poehailos | Senior Analyst, Achievement Network

Nadia Salman | Director of Development, Achievement Network

Tier 3: Promising Evidence

Study Overview

This study tests the hypothesis that providing PL focused on teachers' guided adaptation of math HQIM in the context of student misconceptions is a more effective PL design than PL focused solely on student misconceptions.

This study compared Achievement Network's (ANet) PL intervention, which included two versions of group PL focused on understanding and responding to student misconceptions. The first approach (control group) focused on analyzing students' misconceptions and learning strategies to address them. The second approach (treatment group) similarly centered on analyzing and responding to student misconceptions, but also included an additional focus on guided curricular adaptation of the district's HQIM (enVision Florida B.E.S.T) to address misconceptions.

Curricular adaptation is defined as a balanced approach to instructional decision-making in which teachers implement the core components and instructional strategies of their curriculum with integrity while making necessary adjustments to meet the diverse needs of their students.¹ Teachers frequently adapt or supplement their HQIM. On the one hand, research suggests that these adaptations can be implemented in ways that reduce rigor, a problem that may disproportionately affect students from marginalized groups.² On the other hand, when teachers are provided with opportunities to develop adaptations with support, there is evidence that guided adaptation can lead to improved student outcomes.^{3 4 5}

Study Methods and Research Questions

For the study, 27 middle schools were randomly assigned to either the treatment or the control condition, with 36 sixth grade teachers in the treatment group and 35 in the control group. Across conditions, teachers taught a total of 3,314 students. Researchers integrated quantitative analysis of student math benchmark assessment performance with teacher survey and focus group data to answer the following research questions:

1. To what extent does providing teachers with PL that supports them to make guided adaptations of their HQIM, based on their student misconception data, improve teachers' perception of their HQIM compared to teachers who receive PL based only on understanding student misconceptions?

2. To what extent do teachers who receive the PL on guided adaptations of HQIM make more adjustments to HQIM that enhance learning opportunities for students than those who do not?
3. To what extent do students of teachers who receive the PL on guided adaptations of HQIM perform better on the targeted math content than the students of teachers who only received PL about understanding student misconceptions?

[More information on study methods is available here.](#)

Inside the PL

ANet's PL intervention engaged teachers in a structured process for anticipating student misconceptions and planning instruction aligned to their HQIM. As part of the PL, all teachers (control and treatment) first closely analyzed the grade-level math benchmark standards document to clarify what students were expected to know and do. Then, the PL scope and sequence offered both treatment and control teachers the chance to:

- Experience the math as learners: Solve grade-level tasks to unpack strategies, representations, and mathematical language required by the standard
- Analyze student misconceptions: Examine assessment data and common student misconceptions to identify where students struggle

- Model instructional responses: Learn targeted questioning strategies and instructional language to address misconceptions

The treatment group then engaged in a guided process to plan adaptations for upcoming lessons. Treatment teachers were supported to plan specific, rigorous adaptations to HQIM (e.g., conceptual language, adjusted questioning) grounded in student needs in response to the identified misconceptions. This focus on student needs included support from the PL team in considering students' cultural funds of knowledge and aligning adaptations to be responsive to them.

Practitioner Insight

Anchor Adaptation in Student Data

“

This work reinforced that adaptation is most powerful when it is intentional—not improvisational. Teachers benefited from support that helped them

anticipate where students might struggle and plan responsive instructional moves in advance.”

In response to the study's findings, ANet translated this learning into practice with their district partner through a four-step planning process designed to support thoughtful adaptation within HQIM. The process included: (1) close reading the standard, (2) doing the math, (3) understanding common misconceptions, and (4) adapting instruction.

Working with 6th-grade teachers on representing relationships between quantities, facilitators paired deep analysis of the instructional goal with evidence about how students commonly misunderstand the content. Teachers then pre-planned targeted responses—including specific teacher moves, sentence frames, and student watch lists—based on prior performance data.

This approach created a structured pathway for adapting instruction while remaining grounded in the curriculum's intended learning goals.

— ADAPTED FROM REFLECTIONS SHARED BY
Nadia Salman and the ANet team



Findings

01 Teachers who received PL on guided adaptations of HQIM reported more positive perceptions of HQIM than those who did not receive the guided adaptation PL; however, the difference was not statistically significant when modeled.

Teachers in the treatment group rated their curriculum more favorably than comparison teachers across feasibility, appropriateness, differentiation, cultural relevance, and support for multilingual learners. However, these differences were not statistically significant in all models. Small sample sizes and limited statistical precision make it difficult to definitively determine whether these differences represent true effects of the PL.

02 The students of teachers who received PL on guided adaptations of HQIM based on their understanding of student misconceptions performed better on the targeted math content than students whose teachers received PL only on understanding student misconceptions.

Student performance was assessed on four assessment items aligned with the math standard that the PL was focused on. On average, scores on the standards-aligned items were higher for students whose teachers received the treatment PL. Students whose teachers participated in the control PL scored 2.2 out of 4 points on average, while students whose teachers participated in the treatment PL scored 2.5 points on average, a small statistically significant effect of 0.085 SD.

03 The teachers who received PL on guided adaptations of HQIM made adjustments to HQIM, and 37% of students reported that the changes enhanced their learning compared to 27% of students in the control group.

The study explored how teachers adapted HQIM using focus groups with treatment and control teachers. Researchers qualitatively coded this data to identify the number of adaptations made and how strongly they expected those adaptations to enhance learning. Across all teachers, those in the treatment group reported 19 curriculum modifications, while those in the control group reported 11. Among the reported curriculum modifications, 37% of those in the treatment group were coded as learning-enhancing, compared with 27% in the control group. These raw differences suggest that teachers in the treatment group made more adaptations in their work, and that a higher percentage of their adaptations were learning-enhancing. However, these differences were not statistically significant when compared quantitatively after controlling for teachers' years of experience with the curriculum and overall PL satisfaction. As with the perception findings above, sample size limitations apply here as well.

Making Sense of the Results

What This Means for Teachers and Students



This study provides promising evidence that PL focused on anticipating misconceptions and guiding teachers to adapt HQIM can improve student learning outcomes. While changes in teacher perceptions and reported instructional practices were directionally positive but not statistically significant, student performance improved following the intervention. The evidence suggests that even a single PL session focused on data-informed curriculum adaptation can have impacts on classroom instruction and student outcomes.

The following takeaways can inform PL design:

1. Consider incorporating opportunities for guided adaptation of HQIM within CBPL. This study suggests that PL providers can consider how guided adaptation can move from theory to practice within math HQIM implementation. PL providers could consider providing specific protocols for anticipating where the HQIM will fall short of students' needs based on student assessment data and for designing support scaffolds before the lesson is taught.
2. Consider using student misconception data in guided adaptation processes. Qualitative data provided examples of different ways teachers adapted their practice in response to student misconceptions, with many identified as learning-enhancing. This study provides evidence of one PL context-student misconceptions data set that can provide the structure for guided adaptation support.

Exploring the Impact of Thinking Routines and Professional Learning Delivery Models on Teacher and Student Outcomes in Mathematics Education

Throughline Learning

Malika Ali | Chief Innovation Officer, Throughline Learning

Tier 4: Demonstrates a Rationale

Study Overview

This study examined whether providing teachers with CBPL on math thinking routines, delivered through monthly PLCs and 1:1 coaching, improved teacher practice and students' confidence, mindset, and success within the [Eureka Math Curriculum](#).

Prior research suggests that PL programs focused on creating concrete changes in teachers' instructional strategy and practice may be more effective than those that focus on building teachers' content knowledge alone.⁶ One such instructional strategy is the use of thinking routines for students. Thinking

routines are short, easy-to-learn mini-strategies that make visible, then extend and deepen students' thinking.⁷ A growing body of research supports the use of thinking routines in math classrooms, showing that students who are provided with thinking routines are more engaged, have more positive perceptions of learning, are more ready to explore new content, and are better positioned for higher-order problem-solving.⁸ ⁹ ¹⁰ Evidence also suggests that students who use thinking routines show improved grades, exam scores, and stronger reasoning skills.¹¹ ¹²

Study Methods and Research Questions

For this study, schools were randomly assigned to either the treatment or control conditions. The pilot year included one treatment school (n=2 teachers) and three control schools (n=10 teachers). The second year included eight treatment schools (n=32 teachers) and three control schools (n=10 teachers). Researchers integrated quantitative analysis of student performance on a math critical thinking problem-solving rubric, student benchmark math assessment performance, and student experience survey data with teacher survey and teacher competency rubric data to answer the following research questions:

1. What is the effect of participating in PLC sessions focused on the 7-step thinking routines protocol, along with 1:1 coaching sessions, on teacher outcomes?
2. What is the effect of having a teacher who participated in PLC sessions focused on the 7-step thinking routines protocol, along with 1:1 coaching sessions on student confidence, mindset, and success within the Eureka Math curriculum?

[More information on study methods is available here.](#)

Inside the PL

Treatment group teachers participated in eight PLC sessions focused on the 7-step thinking routines protocol and received 1:1 coaching sessions to further support their implementation.

The PLC sessions focused on presenting teachers with the structured 7-step process for scaffolding a thinking routine for problem-solving aligned with Eureka Math's word problems, ensuring that the instructional practices were embedded within HQIM rather than layered on top. The 7-step protocol for scaffolding and developing cognitive skills was created to separate the development of students' problem-solving skills from the development of their math content skills before integrating the two, providing a framework for solving math word problems that makes their thinking visible. Teachers were taught to use the protocol to support students in showing what they know, identifying where they may have misconceptions, and learning strategies to persevere through challenges. Teachers

were also taught how to provide specific feedback, resources, and structured support in response to students' use of the protocol. The PLC sessions then supported teachers' adaptations, choices, and/or differentiation opportunities for using the thinking routine in their practice. The PLC sessions also included time for data analysis, during which teachers collected and analyzed data on students' problem-solving performance using a validated rubric. Teachers examined student work and rubric data to determine what students needed and then planned differentiated next steps.

Then, the 1:1 embedded coaching model provided personalized support for each teacher to meet their implementation needs, including co-planning, co-teaching, modeling, observation, scoring, and data analysis, with time for debriefing, discussion, and reflection.



Findings

01 Teachers who received the PL on thinking routines reported more productive reflection on their practice, stronger collaborative learning, and more growth in implementing instructional strategies than teachers in the control group.

Across both years, teachers in the treatment group reported that engaging in professional learning led to instructional improvement. In year one, all teachers in the treatment group reported feeling comfortable sharing classroom artifacts, discussing ideas about teaching even when uncertain, and raising questions about their practice. They also reported receiving feedback more frequently and being prompted to reflect on their instruction more often than teachers in the control group.

By year two, teachers reported translating this reflection into changes in classroom practice: all teachers implemented an instructional strategy introduced through the PL, and 89% reported growth in their teaching practices. Most teachers described improvements in supporting students to persevere through challenging problems, use of structured approaches to communicate their thinking, and improved ability to solve complex mathematical tasks. Results also suggest that teachers who participated in more PL sessions reported significantly greater confidence in implementing Eureka Math, as well as higher ratings of student problem-solving skills and perseverance, than teachers who participated in fewer sessions.

02 **Students whose teachers received PL on thinking routines saw growth in their mathematical reasoning rubric scores, but there was not a significant impact on study iReady or MAP scores.**

During year two, on average, students in all eight intervention schools showed significant growth of nearly 50% (49% growth, $p < 0.0001$) from baseline to mid-year in their mathematical reasoning rubric scores, with increases across all assessed domains. The final domain of “Evaluates Solutions” showed the most growth, with a 65% increase in scores, followed by the domain “Defines what is known,” with an increase of 45%. It’s important to note that these rubrics are teacher-completed and therefore may be influenced by teacher perceptions. In addition, students of teachers in the treatment group showed insignificant increase in overall scale scores on math assessments compared to students of control teachers.

Making Sense of the Results

What This Means for Teachers and Students



The Thinking Routines Study offers preliminary data on how to implement practices that result in positive trends in teacher confidence and student growth in mathematical reasoning through existing PL structures.

The following takeaways can inform PL design:

1. Consider pairing PLCs with sustained 1:1 coaching to create structured opportunities for teachers to discuss instructional challenges, reflect on classroom practice, and receive individualized feedback over time. Emerging insights from this study raise the possibility that combining collaborative learning spaces with embedded coaching may help normalize deprivatized practice and collaborative problem-solving, which are important conditions for continuous instructional improvement.
2. Consider incorporating PL on thinking routines to support students' mathematical reasoning. Mid-year data from this study offer examples of students developing stronger mathematical reasoning skills in classrooms where teachers engaged in this PL, compared with classrooms where teachers did not have the opportunity to engage in this PL.
3. Consider implementing and evaluating thinking routines over multiple years to capture changes in both problem-solving and broader mathematics outcomes. This study suggests that students' problem-solving rubric scores, which provide a more proximal measure of learning, may improve before changes are reflected on broader interim mathematics assessments. PL designers may benefit from supporting teachers to implement thinking routines consistently over two years to better understand how gains in problem-solving skills translate into grade-level mathematics achievement.

Unlocking the Potential of EL Education's Math Approach to Create Equitable, Capable, and Courageous Math Communities for All

EL Education, Math Culture Lab (MCL), and National Implementation Research Network (NIRN)

Megan Bookhout | Research Scientist, National Implementation Research Network

Ximena Franco-Jenkins | Senior Research Scientist & Co-Director, National Implementation Research Network

Linda Grein | Founder and CEO, Math Culture Lab by EduQuate

Deepa Patel | Founder and Principal Consultant, DPA Strategies LLC

Jordan Templeton | Chief Operating & Program Officer, Math Culture Lab by EduQuate

ESSA Tier: Not Applicable.

This mixed-methods case study was designed to provide rich insight into implementation and school-level instructional change efforts rather than causal impact estimation and, therefore, does not meet ESSA evidence tier criteria.

Study Overview

This study examined how PL focused on building a strong, school-wide math culture and teacher-student relationships in math classrooms translated into change in teacher and student outcomes. EL Education, Math Culture Lab (MCL) and the National Implementation Research Network (NIRN), examined how a PL model focused on strong teacher-student relationships to support HQIM implementation impacted middle school students' mathematical identities, sense of belonging, and academic achievement.

Prior research suggests student-teacher relationships are an important element of students' classroom experiences and that the quality of these relationships may influence student outcomes.¹³ Many students, particularly those from historically marginalized backgrounds, may not be in math classrooms with strong student-teacher relationships.¹⁴ Research demonstrates that teacher practices related to classroom culture and relationships can be intentionally developed through PL and coaching.^{15 16}

Study Methods and Research Questions

Researchers employed a mixed-methods case study approach in a single school, integrating quantitative analyses of student assessments and teacher and student survey data with qualitative analyses of coach and teacher interviews, student listening sessions, and classroom observations. These data informed the following research questions:

1. Across the academic year, was there growth in:
 - a. Teachers' confidence in establishing a positive classroom environment with strong teacher-student relationships?
 - b. Use of equitable math instructional practices, as evidenced by integrity to the Illustrative Mathematics (IM)/Desmos Math 6-8 curriculum? and
 - c. Students' math identity, sense of belonging in math, and academic achievement, particularly for students who are Black, Latine, and those experiencing poverty?

2. Does implementation of the PL (e.g., acceptability, feasibility, appropriateness of the curriculum) predict teacher and/or student outcomes?
3. Are associations between PL implementation and student outcomes moderated by teacher–student relationships?

[More information on study methods is available here.](#)

Inside the PL

The MCL intervention is a PL model that builds the capacity of a School-based Math Culture Lead (SMCL) to establish a clear vision for high-quality, equitable math instruction. The program begins with a foundational three-hour ‘Vision Institute’ for the school’s leadership and math teaching staff to establish shared goals and a common understanding of strong mathematics teaching and learning. Afterward, the Math Culture Lab provides monthly 1.5-day onsite support and twice-monthly 1:1 coaching for the SMCL, focused on clarifying leadership roles, building coaching capacity, collecting and analyzing learning walk data, and planning next steps for coaching and PL.

Then, the SMCL provides direct PL to teachers, including both PLC sessions and 1:1 coaching. The SMCL facilitated monthly, data-informed PLC sessions for teachers, responsive to learning walk data and focused on pedagogy, content, and equitable math practices. These sessions also focused on relationships and identity, with the intent of building teacher capacity to cultivate positive math identities and strong teacher–student relationships. As part of these PLCs, teachers had structured opportunities to learn and problem-solve together. The SMCL also provided 1:1 coaching support focused on strengthening instructional practice and HQIM implementation.

Practitioner Insight

Make Desired Practice Visible Through Demonstration Classrooms

“ *We learned that systems-level leadership development alone was not enough to accelerate instructional change. Sustainable improvement required making expectations visible and pairing leadership development with direct support for teachers.”*

To support school-wide math culture change, Math Culture Lab introduced Lab Teachers as a strategy to clarify what high-quality implementation looked like in practice. Through intensive coaching, Lab Teachers developed classrooms that served as concrete examples of an inclusive, rigorous, and engaging math culture within their school context. These classrooms became spaces where leaders

and teachers could observe instruction together, reflect on evidence of practice, and build a shared understanding of what the vision required in action. For example, focused coaching strengthened teachers’ implementation of problem-solving lesson structures, such as “Launch, Explore, Discuss, Synthesize,” and key instructional practices, including anticipatory planning, boardwork, purposeful questioning, and discourse and engagement strategies.

This approach helped surface and address differences in leader and teacher perceptions of classroom practice by grounding conversations in observable instruction and student experience rather than aspiration alone.

— ADAPTED FROM REFLECTIONS SHARED BY
Jordan Templeton, Math Culture Lab



Findings

01 Teachers' self-reported high-levels of confidence in their ability to establish a positive math classroom environment with strong teacher-student relationships were not aligned with leadership assessments of teacher practice.

Despite teachers reporting high levels of confidence in establishing a positive classroom environment and teacher-student relationships across the fall and spring survey administrations, learning walk observation data captured by the MCL coach and SMCL revealed that equitable and ambitious mathematics instruction was not yet consistently enacted across classrooms. Teacher confidence scores also remained largely unchanged from fall (8.29/10) to spring (7.76/10). There were no significant shifts in teachers' perceptions of curriculum acceptability, appropriateness, or feasibility over the same period.

02 The challenges school leaders identified in developing coaching and departmental capacity in math instruction included teacher mindsets, skill gaps, unfamiliarity with IM curricular materials, inconsistent expectations, limited resources, and onboarding new teaching staff.

School leaders identified implementation challenges during their walkthroughs. Implementation challenges included teacher mindsets, skill gaps, unfamiliarity with IM curricular materials, inconsistent expectations, limited resources, and a largely new teaching staff. Despite these barriers, leaders reported that MCL's support is building coaching and departmental capacity and contributing to a gradual shift toward more inquiry-based instruction that will take time to fully take hold.

03 There was no significant growth in student outcomes over the study year.

Across measures of student math identity, sense of belonging in math, and math academic performance, there was no significant growth in student outcomes over the intervention year.

Making Sense of the Results

What This Means for Teachers and Students



By exploring how implementation of a novel professional learning approach intersects with on-the-ground challenges, PL providers can learn more about supporting lasting math culture change.

The following takeaways can inform PL design:

1. Consider clarifying the leadership inputs required to address implementation barriers for PL focused on shifting school-wide culture. This study offers an example of why PL orgs might design and clearly communicate criteria for successful school-wide math culture change, including the leadership actions, organizational structures, and ongoing supports required to translate PL into sustained shifts in classroom practice. This may address common implementation barriers. This clear articulation can then be used alongside the results of a needs assessment of the school's current systems and practices to outline a pathway forward including a realistic timeline for change.
2. Consider creating opportunities for teachers and leaders to examine differences in perceptions of instructional practice using shared classroom evidence. This study raises the possibility that teachers may benefit from seeing concrete data on classroom culture to help create alignment—not just on vision, but also on the current state of their practice and a clear pathway towards improvement.

Formative Feedback Study: Enhancing Observation-Based Feedback for Coaching Conversations

University of Virginia, Teach For America, and Annenberg Institute at Brown University

Claire Allen-Platt | Research Scientist, Annenberg Institute at Brown University

Emily Bosch | PhD Candidate in the Dept. of Curriculum and Instruction,
University of Virginia School of Education and Human Development

Julie Cohen | Professor, University of Virginia School of Education and
Human Development

Robin Greatrex | Senior Managing Director, Learning & Insights, Teach For America

Brendon Krall | Project Director, Annenberg Institute at Brown University

John Papay | Director of Annenberg Institute at Brown University

Hallie Parten | Data Scientist, University of Virginia School of Education
and Human Development

Grant Van Eaton | Senior Researcher, Teach For America

Tier 2: Moderate Evidence

Study Overview

This study examined the impact of different observation-based feedback formats on novice teacher development and coach perceptions of feedback conversations.

The University of Virginia (UVA) and the Annenberg Institute at Brown University (AIB) partnered with Teach For America (TFA) to examine TFA's enterprise-wide adoption of an observation tool designed to assess and improve teacher-student interactions, the Classroom Assessment Scoring System (CLASS). Alongside qualitative interviews exploring coach and teacher experiences, the team conducted a

randomized experiment to test the effectiveness of different formats for providing teachers with feedback in coaching conversations.

While observation rubrics are a longstanding, standard feature of teacher evaluation systems, there is an ongoing debate in the field regarding the efficacy of providing rubric-based scores directly to teachers.^{17 18} While scores can motivate improvements to effort and skills, they can also run the risk of impeding development by damaging the relational trust between teacher and coach.^{19 20 21 22}

Study Methods and Research Questions

The research team randomly assigned TFA coaches to provide early-career teachers with one of three approaches to sharing classroom observation-based feedback in coaching debriefs: 1) numeric score reports; 2) numeric score reports with comparative data; and 3) business-as-usual descriptive feedback. The enterprise-wide experiment randomized 129 coaches (n=4,166 teachers) to these three conditions.

The study examined impacts on coach and teacher perceptions, coaching conversations, and teacher performance growth. Researchers integrated qualitative and quantitative data, including teacher and coach surveys, CLASS observation scores, coaching transcripts, and interviews with central staff, coaches, and teachers, to answer the following research questions:

1. How do TFA teachers, coaches, and national staff view the CLASS observation system and related feedback?
 2. Do teachers and coaches report that numeric score reports, numeric scores with comparisons, or descriptive feedback promote better (or worse) coaching conversations?
 3. Do teachers improve more (or less) after receiving numeric feedback, numeric feedback with comparisons, or descriptive feedback in coaching conversations?
 4. What are the characteristics of coaching conversations with and without numeric scores (e.g., use of CLASS language, sentiment, emotional tenor, etc.)?
- Additional questions will explore whether the relative impacts differ by teacher, coach, or the match between teacher and coach.
- [More information on study methods is available here.](#)

Inside the PL

The intervention in this study centered on coaching grounded in feedback from CLASS observation data. The CLASS tool provided information about teacher practice in three domains: 1) emotional support; 2) classroom organization; and 3) instructional support.²³ Beginning in 2024-25, TFA rolled out enterprise-wide CLASS observations. Nearly all corps members were observed twice a year with follow-up coaching conversations to provide instructional feedback. The central goal was to provide coaches and teachers across the country with a shared language for discussing teaching. The data also provides a common outcome to assess instruction across the organization and evaluate the impact of professional learning efforts.

Given the substantial investment in CLASS data to inform coaching debriefs, TFA was interested in learning about how best to use this information in

feedback conversations. In particular, the organization debated whether to have coaches provide teachers with the numeric ratings from the observation or to provide more descriptive feedback. As a result, researchers randomly assigned coaches to deliver post-observation feedback in one of the three formats:

- **Numeric Only (Treatment 1):** A numeric score report displaying the teacher's CLASS scores contextualized within the range of possible scores (1-low to 7-high)
- **Numeric + Comparative (Treatment 2):** A numeric score report as described in (1) with an additional comparison to the average CLASS scores for all other teachers
- **Business-as-Usual (Control):** Descriptive feedback provided purely at the coach's discretion, without the standardized numeric reports



Findings

01 Across feedback formats, teachers and coaches were overall satisfied with feedback conversations grounded in CLASS data. Teachers predominantly felt "reflective" (80%), "encouraged" (79%), and "valued" (53%) after coaching conversations.

Interest holders across TFA saw CLASS as a tool that establishes a common language and focuses coaching conversations. Both teachers and coaches found the

coaching conversations during the enterprise-wide RCT helpful. Both coaches and teachers reported that the conversations were clear in purpose, highlighted teacher strengths, and were useful for identifying specific areas for growth. Teachers reported looking forward to their next observation. Qualitative interviews with 37 stakeholders (national staff, coaches, teachers) suggest similar patterns.

02 Coaches preferred the numeric score reports, particularly those with comparison data to the business-as-usual descriptive feedback format. Teachers appear to be more satisfied with the coaching debriefs with numeric score reports and comparison data.

Coaches preferred the two numeric treatment formats over business-as-usual feedback, with effect sizes of 0.32 SD and 0.49 SD in the numeric-only and numeric-plus-comparative conditions, respectively. Many more coaches reported that they understood the purpose of the conversations, that the conversations were useful for teacher development, and that the conversations provided clear areas for strength and growth when they were accompanied by numeric reports. While these estimates are not statistically significant at traditional levels, the magnitude of the impacts is quite large. Interviews suggest that the reports improve conversations even when scores are low.

Teachers reported more satisfaction with coaching debriefs when they received numeric score reports with comparison data (effect size = 0.10) when

compared with business-as-usual, although the differences were not statistically significant. However, these findings suggest that teachers may have preferred quantitative feedback with comparison data.

03 Numeric score reports may better support coaches with translating CLASS data into actionable coaching.

Coaches reported that CLASS was useful for focusing conversations. However, they report widely varied approaches to delivering feedback; integration of CLASS data into feedback routines is uneven but increasing over time. Teacher interviews similarly revealed varying levels of familiarity with CLASS language and CLASS-aligned growth goals. Coaches reported needing more support to translate observational data into actionable steps and connect that data to concrete resources for the teacher. Recent interviews suggest that numeric score reports can serve as one such support, accelerating TFA corps members' internalization of CLASS language and increasing coaches' comfort with and use of the CLASS tool.

Practitioner Insight

Use Shared Frameworks to Make Coaching More Consistent

“

Participating in the research process did more than generate findings—it created opportunities to reflect on our coaching model and identify ways to strengthen coherence across professional learning experiences.”

Insights from the study are contributing to the shift in integrating classroom observation data and comparison reports more directly into coaching conversations to make practice visible and actionable. TFA is introducing a roadmap for teacher learning and development that explicitly connects developmental milestones, instructional quality indicators, coaching actions, and Cultivate. These tools are designed to

work in multiple directions—allowing coaches and teachers to move from evidence of classroom practice to developmental goals, and from developmental goals back to concrete instructional moves.

As additional findings become available—including analysis of coaching conversations—TFA plans to continue refining coaching guidance to ensure coaching interactions are increasingly focused, evidence-based, and effective at translating professional learning into classroom change.

— ADAPTED FROM REFLECTIONS SHARED BY Robin Greatrex, Teach For America

Making Sense of the Results

What This Means for Teachers and Students



This study provides evidence that quantitative feedback can be quite valuable for coaches and teachers, particularly when combined with comparative information. However, a central question is whether and how feedback translates into shifts in instructional practice, which researchers will examine when the full-scale RCT concludes at the end of the 2025-26 school year.

The following takeaways can inform PL design:

1. Consider how standardized feedback routines and formats can provide important benefits. Teachers and coaches report that having a standardized element to their feedback routines can help support effective conversations. This shared framework can also support effective alignment across the organization. PL providers could consider establishing shared feedback routines alongside their observation tools. This may reduce the cognitive load on coaches and help teachers anticipate a common format for debriefs.
2. Consider providing comparative information to contextualize feedback scores. Preliminary data suggest that providing teachers with peer-comparison averages might enhance the perceived usefulness of feedback. PL providers should consider experimenting with contextualizing scores against peer averages, provided they train coaches to frame this data developmentally rather than evaluatively to protect relational trust. Providing explicit scores—and particularly comparative insights—could actually be supportive for coaches.
3. Consider building robust evidence that turns problems of practice into actionable evidence. The study highlights several key lessons about using evidence to support organizational decision-making. TFA's substantial investment in classroom observation infrastructure facilitated learning across the organization. Its willingness to test different feedback practices in an A/B testing framework suggests the value of examining different approaches to PL. It can serve as a model for other organizations.

Cross-Cutting

Emerging Insights and Design Implications

S.03

Cross-Cutting Emerging Insights and Design

1. Peer collaboration around concrete instructional decisions



In this section, we identify initial cross-study insights for designing effective PL that build on what we know from [RPPL's Building Better PL brief](#). Across varying PL models, the most promising approaches created structures that help educators make disciplined and collaborative instructional decisions using data from students in ways that remained aligned to the goals of math HQIM. While some findings require further research investigation, the patterns surfacing across these recent studies offer important early insight into how certain PL design features can improve implementation and instructional practice.

Below are the three main categories of insights that emerged:

Research has repeatedly demonstrated that teacher collaboration can improve teacher practice and student outcomes. However, our evidence base is weaker about how to design and facilitate these collaborative opportunities.⁶ RPPL's Building Better PL brief suggested that collaboration may be most effective when it “embodies joint work around shared and specific goals.” The studies included here offer examples of how collaborative PL can be structured to achieve this goal.

The Guided HQIM Adaptation Study brought teachers together to analyze student misunderstandings, learn about related pedagogical moves, and plan specific curricular adaptations to address them. Teachers engaged together in structured routines—unpacking standards, solving problems as learners, and anticipating errors—that resulted in concrete, specific decisions about what to do in their classrooms. In this way, the treatment PL fostered collaborative decision-making.

Similarly, the Thinking Routines Study convened teachers in PLC sessions to learn about a specific instructional routine for fostering student thinking with Eureka Math problems and to make implementation choices grounded in structured analysis of student work using the SCALE rubric.

While the specific routines differ across studies, the underlying mechanism is similar: **collaboration may develop teachers' skills more when it provides structured routines that foster collective decision-making about concrete next steps for instruction.**

What does this mean for PL design?

The Guided HQIM Adaptation Study and Thinking Routines Study suggest **PL providers should consider collaborative learning on specific, in-the-moment instructional decisions essential for lesson implementation.** Focusing on a small number of targeted choices—such as anticipating misconceptions or using questioning strategies to elicit student thinking—can be more effective than addressing broad sets of

instructional strategies in general (e.g., inquiry-based learning at large or engagement activities).

Examples of this include:

- Lesson-study routines
- Shared protocols for analyzing student work
- Curriculum-based planning cycles
- Structured problem-solving discussions that focus teachers on concrete instructional choices tied to upcoming instruction

To ensure these activities and decisions are relevant, PL providers may benefit from thoughtfully grouping teachers using the same curriculum, objectives, and grade bands so they can engage in joint instructional decision-making.

2. Coaching anchored in common frameworks and artifacts



Research shows a large variation in impacts within and across instructional coaching programs.^{24 25} Many coaching programs try to maintain quality by training coaches in a common model and providing structured conversation protocols, particularly given the challenge of managing differing levels of coach knowledge, skill, and experience. For instance, some programs use structured debrief protocols that guide coaches through a consistent sequence of reflection, evidence review, and goal-setting in each conversation. The Formative Feedback Study and Thinking Routines Study highlight the potential of another approach for guiding the work coaches do: **anchoring coaching conversations in common instructional frameworks, data (e.g., classroom observation scores and student work rubrics), and classroom artifacts (e.g., curricular materials and instructional routines).**

The Thinking Routines Study anchored coaching conversations in Throughline Learning's 7-step student thinking routine, an aligned annotated graphic organizer, the SCALE problem-solving rubric, and the Eureka Math curriculum. These tools provided a common lens for both coaches and teachers to analyze instruction and set goals for improvement, potentially increasing teachers' willingness to deprivatize their practice, use feedback, and implement new strategies. Such tools may help coaches foster psychological safety and support teachers individually in applying PL content in daily practice.

The Formative Feedback Study's coaching was anchored in the CLASS observation tool. Survey and interview data suggest that coaches and teachers found the shared language of CLASS helped to make coaching conversations more effective. While positive impacts of the Thinking Routines Study PL may relate to other PL features, the Formative Feedback Study provides specific evidence of the impact of numeric CLASS score reports on teacher and coach perceptions of coaching.



At the same time, the Formative Feedback Study highlights that shared data and artifacts are likely not enough on their own to ensure productive coaching conversations. The study surfaced a “coach translation gap.” Some coaches reported having access to observational (CLASS) data but lacked sufficient support to translate it into concrete, actionable next steps for teachers. Similarly, the Math Culture & Relationships Study hints at the challenge the School Math Culture Lead faced in providing effective coaching for teachers while also developing their own internal understanding of equitable math instruction.

These studies both offer strong examples of concrete ways to help coaches negotiate instructional improvement goals with teachers in ways that can ultimately lead to shifts in teacher practice. Effective coaches identify specific steps for improvement and also facilitate teacher reflection, offering feedback in ways that result in shared goals for improvement that both the coach and teacher find worthwhile.²⁶ **Instructional frameworks, data, and classroom artifacts can help facilitate this alignment by providing a shared vision of effective instruction and clear tools to ground conversations around improvement.**

What does this mean for PL design?

The initial implications of these studies suggest that PL providers can benefit from using specific instructional frameworks, data, and classroom artifacts to ground coaching conversations with teachers. Materials that help communicate a concrete vision of effective instruction and/or concrete data on how to improve instruction aligned with the broader improvement goals of the coaching program proved particularly effective in these instances.

This may require **PL providers to more intentionally design coaching routines around shared tools and clear improvement protocols that help coaches translate evidence into actionable next steps.**

Concrete tactics include:

- Standardized observation tools
- Lesson analysis templates
- Structured debrief routines

For districts, where coaches often work within the same structures they are trying to improve, this also means strengthening conditions like protected time, clear roles, and leadership alignment are needed.

3. HQIM adaptation driven by data on student misconceptions



Designers of CBPL face the challenge of designing PL that supports implementation integrity, while enabling effective, context-specific curricular adaptation.^{1 27} This balancing act requires clear, evidence-based guidance and guardrails, as thoughtful teacher adaptations can improve student learning, but some changes may undermine the impact of high-quality materials.^{1 27 28} Despite its importance, little evidence exists on how best to guide teachers in making quality adaptations.

Across the Guided HQIM Adaptation Study and the Thinking Routines Study, a consistent pattern suggests that CBPL better supports teachers in adapting curricular materials when it uses guided analysis of student learning data. In both studies, data-driven implementation planning allowed teachers to adapt practices while maintaining intellectual rigor and positive student learning impacts.

In the Guided HQIM Adaptation Study, teachers in both the treatment and control groups analyzed assessment data to identify common student misconceptions. These misconceptions then guided the instructional practices and responses that were later modeled for teachers during their PL. Only the treatment group received guided practice opportunities to translate key insights into targeted adaptations of HQIM. This PL feature, paired with the analysis of student data, led to stronger outcomes for the treatment group teachers and students, highlighting the importance of deliberate PL to connect student data to specific adaptation decisions.

In the Thinking Routines Study, teachers were introduced to a specific instructional routine and participated in guided PL activities to determine how to implement the routine in response to student data within the Eureka curriculum. Teachers learned 7-steps to guide students' problem-solving, worked with a coach to select appropriate moments and specific math problems to apply the thinking protocol, and analyzed student work produced after using the thinking routine. The study design does not allow us to isolate the impacts of this specific PL design feature, but hints at the benefits of this approach, potentially as a driver of teachers' increased comfort in deprivatizing their practice and opportunities to reflect on their instruction.

These studies suggest that specific, student-centered data embedded within instructional routines may be a key ingredient for shifting how teachers adopt and adapt curriculum in practice.

What does this mean for PL design?

To create meaningful learning sequences that support teachers in adapting HQIM, these preliminary findings suggest designing PL around the use of concrete student learning data—such as student errors, work samples, formative assessments, and patterns in student



thinking—to inform specific decisions about how to implement and adapt instructional materials. Student data can also serve as an important reflection tool after implementation, helping teachers evaluate the effects of specific adaptations and refine their instructional decisions over time. For districts evaluating CBPL providers, this suggests value in asking whether a provider's model includes structured, data-informed guidance for curriculum adaptation.

In practice, building coherent and repeatable routines help teachers use concrete evidence to identify when and how lesson materials should be adapted while preserving the core elements necessary to maintain curriculum integrity. This enables teachers to make disciplined, responsive adaptations that strengthen instruction without compromising rigor.

Examples include:

- Student work analysis protocols
 - Lesson-specific adaptation planning templates
 - Structured pre/post-instruction data reviews
 - Coaching conversations grounded in student evidence.
- Together, these routines help teachers translate student learning data into concrete instructional decisions.

What We're Still Learning

Beyond the promising cross-study insights, these studies reveal the limitations of what existing research can explain. The challenge ahead is understanding not only whether PL features show promise, but whether improvements endure over time, translate across contexts, and support improved outcomes for every student. The three questions to consider moving forward are:

1. How can we balance the short time horizons for decision-making with the long time horizons for change?

Across studies, early shifts in teacher practice and student outcomes suggest that even relatively brief PL experiences influence instruction in meaningful ways. At the same time, differences between teacher perceptions, implementation measures, and observed classroom practice suggest that short-term indicators may not fully capture sustained change. These findings raise questions about which measures best detect early progress, how long effects persist, and the specific PL features needed to support durable improvements over time.

2. How can smaller, exploratory studies help inform decisions at scale?

Several studies indicate what PL settings and implementation conditions are promising, but their designs were not intended to determine whether the results apply broadly across districts or instructional materials. Building evidence at scale may require coordinated portfolios of inquiry—including replication, follow-up, and studies across diverse contexts—to better understand where and under what conditions PL approaches are universally transferrable.

3. How can we better understand which PL approaches support improved outcomes for every student?

Some studies documented early signs of how implementation and outcomes may vary for different student groups and contexts. More research is needed to definitively conclude how PL approaches produce different results for various students, including Black, Latine, Indigenous, AAPI, multilingual learners, students with disabilities, and those experiencing poverty. By understanding how PL supports improved outcomes amongst these groups, we can better implement consistent, rigorous, and high-quality instruction for every student, no matter who they are, where they come from, or what unique challenges they face.

No single study can fully explain how PL produces lasting, scalable, and improved outcomes for all students. Cumulative learning across studies can highlight which PL approaches show promise and under what conditions, while identifying questions for continued investigation. This portfolio is one step forward in that broader effort.

Endnotes

1. Alicea, S., Davis, C., Foster, E., Hornak, R., Hyler, M., Morrison, A., Papay, J. P., Richardson, J., Schwartz, N., & Wechsler, M. (2025). Defining curriculum-based professional learning: Building a common language. Research Partnership for Professional Learning, Annenberg Institute at Brown University. https://rpplpartnership.org/wp-content/uploads/2025/04/RPPL_PositionPaper_041025_FINAL.pdf
2. TNTP. (2018). The opportunity myth: What students can show us about how school is letting them down—and how to fix it. https://tntp.org/tntp_the-opportunity-myth_web/
3. Kim, J. S., Burkhauser, M. A., Quinn, D. M., Guryan, J., Kingston, H. C., & Aleman, K. (2017). Effectiveness of structured teacher adaptations to an evidence-based summer literacy program. *Reading Research Quarterly*, 52(4), 443–467.
4. McMaster, K. L., Jung, P., Brandes, D., Pinto, V., Fuchs, D., Kearns, D., ... Yen, L. (2014). Customizing a research-based reading practice: Balancing the importance of implementation fidelity with professional judgment. *The Reading Teacher*, 68(3), 173–183.
5. Relyea, J. E., Gilbert, J. B., Burkhauser, M., Scherer, E., Mosher, D. M., Wei, Z., ... Kim, J. S. (2025). Asset-based implementation of structured adaptations in an online third-grade content literacy intervention. *Reading Research Quarterly*, 60(4).
6. Hill, H. C., & Papay, J. P. (2022). Building better PL: How to strengthen teacher learning. Research Partnership for Professional Learning, Annenberg Institute at Brown University. <https://annenberg.brown.edu/sites/default/files/rppl-building-better-pl.pdf>
7. Project Zero. (2022). Project Zero’s thinking routine toolbox. Harvard Graduate School of Education. <https://pz.harvard.edu/thinking-routines>
8. Gardner, H. (1991). *The unschooled mind: How children think and how schools should teach*. Basic Books.
9. Ritchhart, R. (2015). *Creating cultures of thinking: The 8 forces we must master to truly transform our schools*. Jossey-Bass.
10. Salmon, A. (2008). Promoting a culture of thinking in the young child. *Early Childhood Education Journal*, 35(5), 457–461. <https://doi.org/10.1007/s10643-007-0227-y>
11. Delany, C., & Golding, C. (2014). Teaching clinical reasoning by making thinking visible: An action research project with allied health clinical educators. *BMC Medical Education*, 14, 20. <https://doi.org/10.1186/1472-6920-14-20>
12. Larraguibel, S.Y., & Venegas-Muggli, J. I. (2019). Effects of using thinking routines on the academic results of business students at a Chilean tertiary education institution. *Decision Sciences Journal of Innovative Education*, 17(4), 405–417. <https://doi.org/10.1111/dsji.12191>

13. McGrath, K. F., & Van Bergen, P. (2015). Who, when, why and to what end? Students at risk of negative student–teacher relationships and their outcomes. *Educational Research Review*, 14, 1-17.
14. Battey, D. (2013). “Good” mathematics teaching for students of color and those in poverty: The importance of relational interactions within instruction. *Educational Studies in Mathematics*, 82(1), 125-144.
15. Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science (New York, N.Y.)*, 333(6045), 1034–1037. <https://doi.org/10.1126/science.1207998>
16. Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of Web-Mediated Professional Development Resources on Teacher-Child Interactions in Pre-Kindergarten Classrooms. *Early childhood research quarterly*, 23(4), 431–451. <https://doi.org/10.1016/j.ecresq.2008.02.001>
17. Bell, C. A., Dobbelaer, M. J., Klette, K., & Visscher, A. (2019). Qualities of classroom observation systems. *School effectiveness and school improvement*, 30(1), 3-29.
18. Grissom, J. A., & Loeb, S. (2017). Assessing principals’ assessments: Subjective evaluations of teacher effectiveness in low- and high-stakes environments. *Education Finance and Policy*, 12(3), 369–395.
19. Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals’ human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44(2), 96–104.
20. Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7), 3628–3651.
21. Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 46(5), 234–249. <https://doi.org/10.3102/0013189X17718797>
22. Pulfrey, C., Darnon, C., & Butera, F. (2013). Autonomy and task performance: Explaining the impact of grades on intrinsic motivation. *Journal of Educational Psychology*, 105(1), 39–57.
23. Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119. <https://doi.org/10.3102/0013189X09332374>
24. Blazar, D., McNamara, D., & Blue, G. (2024). Instructional coaching personnel and program scalability. *Education Finance and Policy*, 19(3), 492–523. https://doi.org/10.1162/edfp_a_00407
25. Kraft, M.A., Blazar, D., Hogan, D. (2018). The effect of teaching coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547-588.

26. Kochmanski, N., & Cobb, P. (2023). Identifying and negotiating productive instructional improvement goals in one-on-one mathematics coaching. *Journal of Teacher Education*, 74(5), 437–450. <https://doi.org/10.1177/00224871231168198>
27. Hill, H. C., Papay, J. P., & Schwartz, N. (2022). Dispelling the myths: What the research says about teacher professional learning. Research Partnership for Professional Learning, Annenberg Institute at Brown University. <https://rpplpartnership.org/wp-content/uploads/2024/10/rppl-dispelling-myths.pdf>
28. Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting program impacts: A brief review. *Educational Researcher*, 48(9), 590–598. <https://doi.org/10.3102/0013189X19891436>

