



Impact by Design

Using Qualitative Comparative Analysis
to Identify Pathways to Instructional
Transformation

Impact by Design: Using Qualitative Comparative Analysis to Identify Pathways to Instructional Transformation

Ariana Audisio and Rebecca Taylor-Perryman

with Cassandra Cross, Tim Tasker, Ingo Rohlfing, and Andrew McEachin

Introduction

School systems face increasingly complex choices about how and where to spend limited resources to maximize student outcomes, and they operate from incomplete information about which investments will make a difference (Rauf, 2025). The need for evidence-based solutions has never been greater.

Student proficiency in math and reading has not recovered since the pandemic (Nation's Report Card, 2025). Leaders must contend with how to raise student achievement against today's standards while rapid changes in the world increase demands for future readiness.

- Research into what works is becoming increasingly scarce, with billions of dollars in federal funding being cut from universities (Center for American Progress, 2025).
- At the same time, teachers are overwhelmed and stressed, working an average of 10 hours over their contracted hours per week and experiencing worse well-being than other working adults (Steiner et al., 2025).

When school districts choose to invest both scarce funds and limited teacher time in new professional learning initiatives, leaders must ensure that the investment will have a meaningful impact on students. Professional learning has the potential to make a significant

difference in student learning (Hill et al., 2022) by strengthening educators' knowledge, skillfulness, and adaptability.

But how can school systems identify *which* professional learning initiatives will make a difference, and which will waste precious time, effort, and money?

Unfortunately, even initiatives with strong results in one context can fall short of these goals when implemented in new contexts under different conditions (Hill et al., 2021). Exciting results from one study often don't hold up when others try to replicate them (Gelbach and Robinson, 2021). Even more, when results do not hold up, those findings are rarely made public. Consequently, when education leaders aim to sort through the evidence to make decisions about professional learning, a significant portion of the picture remains invisible.

Non-profit organizations can play a crucial role in helping education leaders bridge the gap between research and practice. One such organization, **Leading Educators**, has designed and delivered job-embedded professional learning initiatives designed for durable capacity building that have led to student impact in school systems across the country since 2011.

- The Leading Educators' improvement approach has achieved unique results, including enduring impacts on student proficiency two years after the end of a professional learning partnership (Audisio et al., 2024) and a gold-standard randomized controlled trial that found improvement in student learning after just one year (Mihaly et al., 2022).
- At the same time, some partnerships do not show progress on summative student assessments in the first or even second year. One potential cause may be the school and system conditions needed to facilitate impact, but evidence is lacking about how to ensure the most important conditions are in place.

In reality, professional learning must occur under imperfect conditions and with limited resources to support *all students*.

That is why **Leading Educators**, through the support of the **Research Partnership for Professional Learning** (RPPL), conducted one of the most comprehensive mixed-methods studies of professional learning ever completed in K-12 education. They examined 42 partnerships across school systems nationwide from 2015 to 2024 to identify places where students have made gains and uncover the "recipes"—the specific combinations of conditions and supports—that lead to meaningful student growth.

The findings from this study confirm that professional learning, when *strategically* designed, can lead to significant improvements in student learning. Moreover, this study provides an important roadmap to support that goal by identifying the combination of features of professional learning that can achieve results in various contexts.

The research team identified three key features necessary to achieve student impact through professional learning across various contexts, as well as specific pathways to achieve impact in different contexts. The studied contexts included schools with large populations of

multilingual learners, schools demonstrating the lowest results for students (typically schools on the state improvement list), schools that did not opt to participate in the program, and very large districts.

This study involved an ambitious research effort, bringing together researchers and practitioners from Leading Educators with external researchers from the University of Passau, ETS Research Institute, and Civilytics Consulting in a research practice partnership. The team applied Qualitative Comparative Analysis, a well-established research methodology commonly used in political science and sociology, but rarely in educational research (Mello, 2021), in conjunction with quasi-experimental research methods.

- First, the research team expanded the number of rigorous evaluations of impact on summative student assessments.
- Next, the researchers carefully described the contextual factors, focus areas, and program components of each initiative and used Boolean algebra to identify “recipes” for impact. These recipes are the combinations of contextual factors, foci, and program components that are sufficient to achieve the desired outcome, meaning whenever that particular combination was in place, the outcome was consistently achieved.
- This comprehensive effort yielded promising evidence that standards and curriculum-based improvement efforts can yield real outcomes, counter to the dominant narratives, as well as concrete design considerations for future professional learning initiatives in different contexts.

Ultimately, these findings and approach offer a model for others to follow in accelerating the pace of learning about what works, for whom, and under what conditions.

Leading Educators’ Improvement Approach

Leading Educators partners directly with school systems to create sustainable professional learning systems that improve teacher and student outcomes. Grounded in research on effective professional learning, this work is tailored to each partner’s unique context and shared vision of excellent teaching and learning, yielding promising results (Mihaly et al., 2022). These partnerships vary across geographies and evolve over time in response to each system’s context, resources, and needs. Partnerships are often, but not always, targeted towards the selection and implementation of high-quality instructional materials. Some partnerships target principals and assistant principals, while others focus on developing teacher leaders. Most partnerships target leaders at multiple levels to support instructional coherence.

Professional learning typically takes the form of advising at the system level to develop conditions, one-on-one coaching, professional development workshops that bring together schools across a system, and job-embedded professional learning communities that occur during the school day.

Phase 1: Using Rigorous Methods to Evaluate the Impact of Each Partnership

Leading Educators leverages sophisticated research methods and carefully reviews best practices to ensure partners experience reliable impact. The goal is to understand not just how much students grow in supported schools but also how likely it is that the support model contributed to student growth.

The first phase of this study involved assembling a large pool of year- and subject-specific interventions and applying rigorous methods to evaluate their impact. Building on three prior rigorous quasi-experimental or experimental studies, the researchers identified nine multi-year partnerships to study.

Longitudinal Student Achievement for LE and Comparison Schools: Data and Sample

To evaluate the effect of LE's support on math and ELA student achievement, the researchers constructed a longitudinal school-grade-year dataset for both supported schools and schools not supported by Leading Educators, typically for a minimum of 3 years before the program starts and all available programming years.

The datasets contained grade-level data from the state or district, including mean scale scores. The researchers converted raw test scores into standard deviations to compare outcomes on a uniform scale. This works because a standard deviation measures the amount of variation in a dataset.

The datasets also contained the school-grade level percentages of Black or African American students, Hispanic or Latino students, multiracial students, students eligible for free or reduced-price lunch (FRL), multilingual learners, and students receiving special education services. For each case, the researchers identified a set of treatment schools that received the same support from LE, while comparison schools did not receive any support from LE. This allowed them to identify a set of schools that received a particular recipe.

Identifying the Best Comparison Samples and Estimating Effects: Empirical Approach

After constructing the dataset, the researchers compared the trajectories of LE-supported schools with those of schools of similar socioeconomic status in the same district or across the entire state. Next, the researchers identified a comparison group that matches the treatment population as closely as possible. They aimed to approximate what would have happened to treatment schools if they had not participated in the intervention.

The researchers used a statistical technique known as propensity score weighting (PSW) in combination with machine learning models¹ to create a strong comparison group, then they used difference-in-differences (D-in-D) strategies to estimate the effect of LE's partnership. They employed a weighted D-in-D approach, which compares changes in math standardized scores in treatment schools over time to changes in scores for students in comparison schools.

These models controlled for demographics, test scores in the years *prior* to the start of LE support, and grade—and school-level differences. Together, these methods helped ensure comparability across groups, thereby increasing confidence that the effects are due to the program, rather than other factors ongoing in each context.

Finally, to interpret the results, the researchers employed a framework developed by Deke, Finucane, and Thal (2022) to determine if the findings are educationally meaningful. This approach involves using probability models that incorporate existing knowledge from prior research on a similar subject and grade level. This helped the team get a realistic understanding of an intervention's likely impact. Ultimately, for each case, the researchers reported the percent growth for supported schools and comparison schools and then provided the probability that LE supported student growth.

Understanding Support at Scale: District-Level Analyses

Leading Educators' mission focuses on building partnerships across school systems to transform instruction and student outcomes at scale. In programs scaled across an entire district, we took a different approach to identifying a strong comparison group since all schools in that system receive LE support.

We used district-level data from the National Assessment of Educational Progress (NAEP) for all large city districts with data available in tested years from 2011 to 2024 that had not received district-level support from LE. Sixteen cities were available for comparison. The data also included district-level percentages of school-age students in poverty, Black students, Hispanic students, multilingual learners, and students receiving special education services. Data was obtained from the Urban Institute's Education Data Explorer.

To create a strong comparison group, we used Synthetic Controls, a leading econometric method used to systematically select comparison groups for interventions that are implemented at an aggregate level, like a district or state, and there is a small number of comparison units to use as a comparison. This method combined data from a set of comparison districts that did not receive LE support to create a "synthetic" comparison, or a fictional district that closely mirrors the treated district. A data-driven approach weighted the comparison districts so that the synthetic control group reproduces the treated districts' pre-program trends and characteristics as closely as possible. Then, the differences between treatment and the synthetic control after the treatment can be attributed to the intervention.

¹ <https://www.rand.org/statistics/twang.html>

Phase 1 Findings: What Was Learned about Impact

Drawing on nine studies conducted as part of this project, along with three prior Leading Educators evaluations, we assembled a set of 42 cases. We examined results for individual “cases,” defining a “case” as one set of schools or one school district that received support from Leading Educators with the goal of improving student achievement in one subject (English language arts or math) in one school year.

For each case, we evaluated whether the evidence was sufficient to conclude that Leading Educators contributed to student growth of an educationally meaningful magnitude. For partnerships that supported a subgroup of schools within a district, we assessed whether there was at least 70% certainty that Leading Educators contributed to at least 0.05 SD (equivalent to approximately 2.4 months of learning)² of student growth². Of the 38 school-level cases, nearly half met this rigorous standard of positive impact.

- LE schools **grew 8–34% in math and 12–18% in ELA**, outpacing comparison schools. On average, this growth represents about **9 months of additional learning** for students.
- In these cases, LE-supported schools began with clear areas of need, with scores below those of 70–93% of schools in the state (ranging from -0.54 to -1.47 SD).

For district-wide partnerships, we assessed whether student growth increased by at least one point more than the most comparable district or group of districts on the NAEP exam.

- **All four cases** across the three districts examined met this standard, **demonstrating a positive impact** from Leading Educators' support.
- Before the program started, NAEP scores in the three supported districts were 2 to 29 points lower than those of all other large city districts that participated in the NAEP during the study period. These **districts grew 1 to 3 points more** than the comparison districts in the years immediately after the partnership.

In evaluating each case, we also considered whether the statistical methods were able to identify a comparison group that, on average, closely matched the treatment group on important characteristics (called “baseline equivalence”), whether the units in the treatment group had a unit in the comparison group with similar characteristics (called “propensity score overlap”), and whether the treatment and comparison groups had similar trajectories before LE support (called “parallel trends”). We also examined descriptive achievement trajectories to determine whether LE-supported schools experienced greater growth than similar schools during the period of LE support.

In some cases, it was difficult to find perfect baseline equivalence and overlap, given how different the supported schools were from other schools in the state. In such cases, we also

² We used Hanushek (2012) benchmarks to convert to months of learning across this report.

examined alternative statistical methods, including unweighted differences in differences. For all cases in the analysis, we compared results from all methods to decide if there was a significant chance that the program was successful at improving outcomes in one year, and then we moved to phase 2 of the study.

Phase 2: Identifying Recipes for Impact

Qualitative Comparative Analysis (QCA) is an established technique in political science and sociology. It is a mixed-methods approach that helps identify the specific combinations of conditions under which desired outcomes occur. Instead of focusing on average effects, it shows which factors or combinations of conditions are necessary and sufficient for success across different cases (Mello, 2021). This approach directly responds to calls in the education research literature, such as Borko's (2004) emphasis on the need to generate comparative evidence about the implementation, effects, and resource requirements of well-defined professional development programs, which is essential for informing policy decisions about resource allocation.

QCA is especially useful in situations where the outcome depends on the complex interplay of conditions. In schools, many complex factors intersect to influence student outcomes. Sometimes, a strategy that works in one school backfires in another. QCA is designed specifically to anticipate and address this type of complexity.

In this study, we examined the conditions that shape the success of teacher professional learning programs, including factors from both the school and system contexts, as well as features of the programs themselves.

To identify which conditions to study, we conducted a set of interviews with program staff and completed a literature review of 14 review papers and meta-analyses. In the interviews, we asked staff to reflect on the service activities and characteristics of the teachers and leaders' context that they believed contributed the most to the presence or absence of positive outcomes in student achievement. We created transcripts from the interviews and prompted the Large Language Model (LLM) llama3.1-405b to conduct a thematic analysis, identifying the frequency of each theme and producing a sentiment score for each. This produced a list of 17 themes.

In the literature review, we conducted a Google Scholar search to identify recent review papers or meta-analyses, and noted the activities and characteristics identified as contributing to the outcome of professional learning. We grouped similar themes together. From those 14 papers, we recorded 24 themes. Next, we compared the two sets of activities and characteristics, identifying commonalities and opportunities to combine. From those two sources, we compiled a list of 17 potential elements, including context, focus areas, and program features.

We then recruited a pool of 13 school and district leaders to discuss the draft list of conditions, prioritizing the most important factors, identifying potential gaps, and discussing how to evaluate each factor. Participants engaged in an hour-long interview, during which they

ranked the most important conditions, identified additional barriers to impact, and answered questions about how they would be able to identify the presence or absence of these factors. Again, we created transcripts from the interviews and prompted llama3.1-405b to conduct a thematic analysis, identifying the frequency of each theme and producing a sentiment score. We carefully reviewed the output and the transcripts. Overall, the input from practitioners largely affirmed the set of conditions and provided useful insights on which conditions to prioritize.

There were four conditions identified as important from the practitioner input and literature review that we were unable to incorporate: these were a program focus that balanced fidelity and adaptation of high-quality instructional materials, a coherent system-wide instructional vision, having an "executive champion" for the work, and teacher and leader retention.

- We lacked high-quality data on these conditions to assign strong ratings to the cases.
- Our inability to rate these potentially important conditions is a notable limitation of the study; however, other conditions intersect with these conditions in ways that may mitigate the impact of this limitation.
- For example, although we were unable to measure the presence or absence of an executive champion, we were able to assess whether multiple levels of leadership were engaged in the program.

With the final list of conditions, we then specified quantitative rules to rate the presence or absence of each condition or characteristic. As summarized in Table 1, the data sources were a mix of detailed artifact analysis and publicly available sources. We used both established conventions and data-driven thresholds to identify the calibration rules, i.e., what constituted a 0 (the condition is fully absent) and what constituted a 1 (the condition is fully present).

After completing the data table, we first conducted a necessity analysis, which identifies the conditions that can be considered prerequisites for the outcome. We set a threshold of 0.85, meaning 85% of the time, the outcome was present only when the conditions were present. Next, we conducted an analysis of which conditions are sufficient to produce impact, which can be considered the different recipes.

The analysis sorts the different combinations of the conditions into three sets: combinations that *are sufficient* for the outcome, combinations that *are not sufficient* to achieve the outcome, and combinations for which *we cannot infer* whether they are sufficient for either outcome (see Figure 1).

Ultimately, the three sets are compared to isolate the combinations with a positive outcome that do not also have negative outcomes. Finally, we interpreted the truth table, reviewing the recipes that are only associated with a positive outcome. To facilitate interpretation, we grouped recipes with shared contextual elements and then identified the program focus areas and program components that consistently led to success within each contextual factor.

TABLE 1: CONDITIONS

Name	Description	Assessment Method
<i>High-quality instructional materials (HQIM)</i>	Curriculum and assessment materials that are standards-aligned and effectively support teaching and learning	We reviewed program design documents to identify if the program was focused on a specific curriculum. If the partner did not have a single curriculum, we reviewed survey data from schools on their primary curricula. Then, we consulted Ed Reports ratings. If >60% of schools have a curriculum in place with an Ed Reports Meets Expectations rating, this was rated 1, otherwise, 0.
<i>Very large districts</i>	A district that serves a very large number of students	We collected enrollment data from the National Center on Education Statistics. If the district served over 100,000, we record it as a very large district.
<i>Very large achievement gap</i>	A very large baseline achievement gap between the supported schools/district and the state	We calculated the mean scale score for supported schools at baseline, standardized within the state. Schools with an achievement gap of more than 1 standard deviation were rated 1, and schools with an achievement gap smaller than 1 standard deviation were rated as 0.
<i>Large proportion of Multilingual Learners</i>	A large proportion of students classified as Multilingual or English Learners	We collected data from each state and calculated the mean percentage for supported schools at baseline. Schools with a percentage of multilingual learners above 12% were classified as high percentage of multilingual learners.
<i>Focus on subject-specific pedagogical practices</i>	The program has a goal of building teachers' knowledge in how to effectively convey complex, content-specific concepts in ways that are accessible to students	We asked project teams to list the specific topics covered in their professional learning sessions, and also reviewed project goals and step-back materials. If these sources listed one of these three foci (subject-specific pedagogical practices, teacher-student relationships, or teachers' mindsets) as a focus area, we marked this as 1. If the projects focused on content-agnostic instructional leadership skills, we marked this as a 0.
<i>Focus on teacher-student relationships</i>	The program has a goal of enhancing teachers' skills in building positive and supportive connections with their students	

<i>Focus on teacher mindsets</i>	The program has a goal of helping teachers and leaders develop a growth mindset, recognize the potential of every student, and foster high expectations for all	
<i>Leader engagement</i>	The program provides school- and system-based instructional leaders with role-appropriate ways to participate in and reinforce teacher learning	If multiple role levels (for example, both principals and teacher leaders) participated in the same topics for professional learning, we rated this as a 1. Otherwise, we rated this condition as a 0.
<i>Opt in</i>	Whether participants chose to participate	If the principal or teacher elected to participate in the program, this was rated as a 1. If the district assigned schools to participate, this condition was rated as 0.
<i>Strong dosage of professional Learning Cycles (PLCs)</i>	The program provides ongoing, shared opportunities for teachers to engage in hands-on, participatory activities in shared learning to improve instruction, build relationships, practice, plan, and collaborate	Each year, program leaders complete a survey detailing program components. They revisit and update their responses two to three times per year. Finally, we reviewed the information a final time with each team to ensure accuracy. In this survey, lwsWEA listed the hours of Professional Learning Cycles or other similar school-based learning that participants were expected to engage in as a program requirement. Sometimes LE staff facilitated or participated in these cycles, but in many cases, LE provided materials and training, and school-based staff facilitated these cycles. If the number of hours listed in the survey was 18 hours or more (equivalent to 1 hour of meetings every two weeks), we rated this as 1; otherwise, we rated it as 0.
<i>Learning Walks</i>	The program provides a structured process where educators, often in groups, visit classrooms to observe teaching and learning practices	On the program component survey, staff also recorded the number of hours they spent hosting learning walks or co-observations alongside school and/or district staff. If the program supported learning walks or co-observations, we rated this as 1; otherwise, we rated it as 0.

Phase 2 Results: Recipes for Impact

Before beginning the analysis, we identified the following hypotheses about impactful program components to guide program design, based on a preliminary comparative analysis and evidence from the literature. This approach is detailed in Table 2.

TABLE 2: LEADING EDUCATORS' APPROACH TO SCOPING

<i>Transformational outcomes for teaching and learning will be achieved if...</i>				
We provide Strategic Advising to strengthen conditions ,	And we focus our support design on a combination of: - Subject-specific pedagogical practices - Specific curricular materials - Teacher-student relationships	And we lead at least 32 hours of group learning with peer-to-peer collaboration and follow-up for a school-level role,	And support one-on-one application with at least 8 hours of coaching ,	And engage at least two groups of roles in some shared learning to reinforce coherence across role levels.

The Truth Table analysis (Table 3) identified 15 recipes that consistently led to positive impact; when a recipe produced both positive and non-positive outcomes, it was inferred to be insufficient for impact.

- These 15 recipes all had perfect scores on the measure of fit identified through an analysis of sufficiency, known as consistency.
- These indicate that there were no inconsistencies, or cases where these configurations led to a different outcome.

In reviewing these 15 pathways and the results of the necessity analysis, we found continued broad support for our current approach to scoping professional learning, especially for our emphasis on coherence and alignment across levels. We found that three elements of the PL format are necessary to achieve positive impact. These conditions were present in all of the positive pathways. These are:

- engaging leaders at multiple levels in learning on the same topics,
- providing eight or more hours of coaching for at least one role (school-based or district-based instructional leader),
- focusing on subject-specific pedagogical practices.

This third condition aligns well with existing research, as summarized by RPPL's Building Better Professional Learning brief (2022), which identified subject-specific pedagogical practices as one of six characteristics of PL that strongly impact student outcomes.

The 15 pathways revealed patterns that may support program designers when scoping initiatives in particular settings. When grouping the recipes by contextual factor, we identified a few additional PL elements that were consistently present alongside the three necessary conditions when a positive outcome was observed.

- In both the lowest-performing schools and schools that did not opt in to participate, successful recipes all included learning walks and strategic advising.
- In schools with large proportions of multilingual learners, successful recipes all included learning walks, high-quality Instructional materials, at least eight hours of coaching for a district-level role (often alongside school-level coaching), and strategic advising.
- Finally, in very large districts with over 100,000 students, successful recipes all included learning walks and school-level coaching.

These combinations make intuitive sense. Strategic advising and learning walks can ensure that leadership support and conditions are aligned for schools with additional needs, such as turnaround schools or those with a large proportion of multilingual learners.

Similarly, in very large districts, ensuring there is sufficient support at the school level is logical, as district-level support would likely have a more diffuse impact on each individual school. Notably, learning walks emerged as a key component of 12 out of 15 recipes. This finding aligns well with past research on the value of robust calibration on observing instructional practice with administrators ([Ho & Kane, 2012](#)). Finally, high-quality Instructional materials may be more important for teachers that are supporting high proportions of multilingual learners to create a strong foundation for customizing to different students' needs.

The vast majority of the positive recipes included at least eight hours of coaching for a school-based role. However, four of the recipes included only coaching for a district role.

- These recipes may have achieved this growth by also including at least eight hours of coaching for district-based roles, strategic advising, and multiple years of support.
- Furthermore, these four recipes also included districts with high-quality instructional materials and that were not very large districts (with fewer than 100,000 students).
- The relatively smaller district size may have made district-level coaching more impactful. School-level coaching may be necessary to achieve impact in very large districts.

We also found a surprising finding that went against our initial expectations: Eight positive cases had fewer than 32 hours of group learning. These cases all received two to three times

the recommended coaching dosage, with an average of 14 hours at the school level and 25 hours at the district level. They also had an average of 24 PLC hours and incorporated strategic advising. More coaching and PLCs could have compensated for the lower levels of workshops or more alignment brought about by strategic advising.

Implications and Future Directions

These findings provide a blueprint for program design as professional learning organizations and school systems develop new initiatives to respond to emerging needs.

The two necessary conditions —engaging leaders at multiple levels and providing at least 8 hours of coaching per year —align well with past research (e.g., Kraft, 2018) and Leading Educators' School and System Conditions (Meili and Tasker, 2023). Additionally, the findings as a whole align with our emphasis on the importance of attending to dosage; whenever a recipe had dosage below the hypothesized thresholds in one component, e.g., workshops, the dosage in the other program components (e.g., coaching or PLCs) far exceeded the minimums. This suggests that overall, developing a clear theory for how the learning will be sustained, emphasized, and followed up on throughout the year can facilitate impact.

Importantly, the decisions about which components can compensate for lower levels of others must be tailored to each specific context.

1. Forgoing intensity at the school level to provide more support at the district level may only work in smaller systems, where system leaders are more likely to be closer to the schools and understand and influence their unique characteristics.
2. In larger systems, the complexity added by additional layers of leadership may make an impact unlikely without direct training of school staff.

Likewise, schools serving populations with higher needs may require more attention to address systemic issues.

1. Serving these populations may require additional resources that system leaders can provide.
2. Additionally, teachers may have experienced a fragmented series of programs that have not achieved their goals (Kutash et al., 2010). Ensuring leaders at multiple levels are invested in an approach is necessary for teachers to focus on improving instruction.

This approach is not without limitations. QCA results are sensitive to calibration choices and the selection of cases and conditions. We addressed this by providing thorough justifications for all analytical decisions and conducting robustness tests to ensure the validity of our findings. We included all available cases, except for two where the data could not clearly distinguish between participating and non-participating students.

Some cases also had less-than-ideal comparison groups; however, since our goal was to identify which program and context combinations improved student outcomes after one

year, rather than to measure the exact size of the effect, we employed alternative methods to confirm the direction of the results.

Additionally, QCA does not directly account for the timing or sequence of causal pathways. To partially account for this limitation, we introduce a condition representing the duration of exposure to the program, later conceptualized as “sustained learning” when schools participated for more than one year. Finally, we aimed to measure conditions that lacked sufficient high-quality data.

In future analyses, we plan to explore questions such as:

- How do recipes differ in contexts with frequent teacher or principal **turnover**?
- What more can we learn about **dosage** and how that intersects with the number of roles served and dosage of other program components (e.g., more coaching and fewer workshop hours)?
- What can we learn about the **qualitative features of coaching and workshops** and their connection to recipes?

Overall, these findings provide confirmation of the value of Qualitative Comparative Analysis as a tool for education researchers, practitioners, and program designers. The value of ensuring that an initiative is successful and that the same approach is *not* associated with a null outcome is immense for system leaders making decisions with scarce resources.

As we explore new approaches in various contexts, we’ll continue to assess our impact using rigorous methods and utilize Qualitative Comparative Analysis as a tool to refine our recipes.

TABLE 3: TRUTH TABLE

HQIM	Very Large District	Very Large Baseline Gap	High % of MLL	Subject-Specific Practices	Student Teacher Relationships	Teacher Mindsets	Leaders Engaged	Opt In	Learning Walks	PLCs >= 18 hours	Coaching >= 8 hours		Workshop >= 32 hours		>= 2 Years	Strategic Advising	# Cases
											School	District	School	District			
Recipes that are associated only with a positive outcome																	
0	1	0	0	1	0	1	1	0	1	1	1	0	0	0	1	1	1
0	1	0	0	1	0	1	1	0	1	1	1	0	1	0	0	0	1
1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	1	1
1	0	0	0	1	0	0	1	0	1	0	0	1	0	0	1	1	1
1	0	0	1	1	0	0	1	0	1	1	1	1	0	0	0	1	1
1	0	0	1	1	0	1	1	0	1	1	1	1	0	0	0	1	1
1	0	0	1	1	1	0	1	0	1	0	0	1	0	0	1	1	1
1	0	0	1	1	1	0	1	0	1	1	1	1	0	0	1	1	1
1	0	0	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1
1	0	0	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1
1	0	1	0	1	0	1	1	0	1	1	0	1	1	1	1	1	2
1	0	1	0	1	0	1	1	1	0	1	1	0	1	0	0	0	1
1	0	1	0	1	0	1	1	1	0	1	1	0	1	0	1	0	1
1	1	0	0	1	0	0	1	0	1	0	1	0	0	0	0	1	1
1	1	0	0	1	0	1	1	1	1	0	1	0	1	1	0	1	1
Recipes that are associated with a non-positive outcome or with both a positive and a non-positive outcome																	
0	0	1	0	0	0	1	1	1	0	1	1	0	1	0	0	0	2
0	0	1	0	1	0	1	1	1	0	1	1	0	1	0	0	0	2
0	0	1	0	1	0	1	1	1	0	1	1	0	1	0	1	0	2
0	0	0	0	0	0	1	1	1	0	1	1	0	1	0	0	0	2
0	0	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	3
0	0	0	0	1	0	1	1	1	0	1	1	0	1	0	1	0	1
1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	2

TABLE 3: TRUTH TABLE

HQIM	Very Large District	Very Large Baseline Gap	High % of MLL	Subject-Specific Practices	Student Teacher Relationships	Teacher Mindsets	Leaders Engaged	Opt In	Learning Walks	PLCs >= 18 hours	Coaching >= 8 hours		Workshop >= 32 hours		>= 2 Years	Strategic Advising	# Cases
											School	District	School	District			
1	0	1	1	0	0	0	1	0	1	0	1	0	0	0	0	1	2
1	0	1	1	1	0	1	1	0	0	1	0	0	0	0	0	1	1
1	0	1	1	1	0	1	1	0	1	1	1	1	0	0	1	1	2
1	1	1	0	1	0	1	1	0	1	0	1	0	1	1	0	1	1
1	1	1	0	1	0	1	1	1	1	0	0	0	0	1	0	1	1
1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	1	1	1
1	1	1	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1
1	1	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	2
1	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0	1	1

References

- Anderson, M., Baker, M., Cato Czupryk, B., Coddington, C., Ehlman, K., Maus, A., Meili, L., Paek, vP., Romansky, C., Taylor-Perryman, R., Vu, H., Worthman, S., & Williamson-Zerwic, B. (2021). A learning agenda for improving teacher professional learning at scale. Research Partnership for Professional Learning.
- Audisio, A. P., Taylor-Perryman, R., Tasker, T. B., & Steinberg, M. P. (2024). Does teacher professional development improve student learning? Evidence from Leading Educators' fellowship model. *Journal of Research on Educational Effectiveness*, 1–40.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational researcher*, 33(8), 3–15.
- Center for American Progress (2025). Mapping Federal Funding Cuts to U.S. Colleges and Universities.
<https://www.americanprogress.org/article/mapping-federal-funding-cuts-to-us-colleges-and-universities/>
- Copland, M. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis*, 25(4), 375–396.
- Deke, J., Finucane, M., & Thal, D. (2022). The BASIE (BAYesian Interpretation of Estimates) framework for interpreting findings from impact evaluations: A practical guide for education researchers. Toolkit. NCEE 2022–005. National Center for Education Evaluation and Regional Assistance.
- Elmore, R., & Burney, B. (1997). Investing in teacher learning: Staff development and instructional improvement in Community District #2, New York City (No. ED416203). Philadelphia, PA: Consortium for Policy Research in Education.
- Gallucci, C. (2008). District-wide instructional reform: Using sociocultural theory to link professional learning to organizational support. *American Journal of Education*, 114(August), 541–581.
- Gelbach, J. B., & Robinson, C. (2021). Why is so little published in J-PAL's replication studies program? *Journal of Economic Literature*, 59(1), 177–194.
- Gregory, A., Ruzek, E., Hafen, C. A., Mikami, A. Y., Allen, J. P., & Pianta, R. C. (2017). My Teaching Partner-Secondary: A video-based coaching model. *Theory Into Practice*, 56(1), 38–45.
- Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). How principals affect students and schools: A systematic synthesis of two decades of research. The Wallace Foundation.
<http://www.wallacefoundation.org/principalsynthesis>
- Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, 90(7), 495–500. <https://doi.org/10.1177/003172170909000709>

Hanushek, E. A., Woessmann, L., & Peterson, P. E. (2012). Is the US catching up? *Education Next*, 12(4).

Hill, H., Papay, J. P., & Schwartz, N. (2022). Dispelling the myths: What the research says about teacher professional learning. *Research Partnership for Professional Learning*, 1-10.

Hill, H. C., & Papay, J. P. (2022). Building better PL: How to strengthen teacher learning. *Research Partnership for Professional Learning*, 1-19.

Hill, H., Papay, J., Schwartz, N., Johnson, S., Freitag, E., Donohue, K., Berry, R., III, Loeb, S., Anderson, M., Baker, M., Cato Czupryk, B., Coddington, C., Ehlman, K., Maus, A., Meili, L., Paek, P., Romansky, C., Taylor-Perryman, R., Vu, H., Worthman, S., & Williamson-Zerwic, B. (2021). A learning agenda for improving teacher professional learning at scale. *Research Partnership for Professional Learning*.

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1(4), 85-108.

Johnson, S. M., Reinhorn, S. K., & Simon, N. S. (2018). Ending isolation: The payoff of teacher teams in successful high-poverty urban schools. *Teachers College Record*, 120(1), 1-46.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 36(4), 476-500.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547-588.

Kutash, J., Nico, E., Gorin, E., Tallant, K., & Rahmatullah, S. (2010). *School turnaround: A brief overview of the landscape and key issues*. San Francisco: FSG Social Impact Advisors. Retrieved September, 27, 2011.

Leithwood, K., Seashore, K., Anderson, S., & Wahlstrom, K. (2004). Review of research: How leadership influences student learning. The Wallace Foundation.
<https://www.wallacefoundation.org/knowledge-center/documents/how-leadership-influences-student-learning.pdf>

Liebowitz, D. D., & Porter, L. (2019). The effect of principal behaviors on student, teacher, and school outcomes: A systematic review and meta-analysis of the empirical literature. *Review of Educational Research*, 89(5), 785-827. <https://doi.org/10.3102/0034654319866133>

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*.

Mello, P. A. (2021). Qualitative comparative analysis: An introduction to research design and application. Georgetown University Press.

Meili, L. & Tasker, T. Supporting Systemic Coherence and Impact. Measuring Teacher Professional Learning: Why It's Hard and What We Can Do About It: Case Studies 12-16. <https://annenberg.brown.edu/sites/default/files/Measuring%20Teacher%20Professional%20Learning%20Case%20Studies.pdf>

Mihaly, Kata, Isaac M. Oppen, and Lucas Greer, The Impact and Implementation of the Chicago Collaborative Teacher Professional Development Program. Santa Monica, CA: RAND Corporation, 2022. https://www.rand.org/pubs/research_reports/RRA2047-1.html.

National Center for Education Statistics. (2025). The Nation's Report Card. U.S. Department of Education. Retrieved September 11, 2025, from <https://www.nationsreportcard.gov/>

Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. E. (2020). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. *American Economic Journal: Economic Policy*, 12(1), 359–388.

Rauf, D. S. (2025, August 22). 'Start Preparing': Why Even Seemingly Stable Sources of K-12 Funding Are Uncertain. EdWeek Market Brief. <https://marketbrief.edweek.org/education-market/start-preparing-why-even-seemingly-stable-sources-of-k-12-funding-are-uncertain/2025/08>

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, 52(3), 475–514.

Steiner, E. D., Levine, P. R., Doan, S., & Woo, A. (2025). Teacher Well-Being, Pay, and Intentions to Leave in 2025: Findings from the State of the American Teacher Survey (RR-A1108-16). RAND Corporation. <https://doi.org/10.7249/RR-A1108-16>

Weddle, H. (2022). Approaches to studying teacher collaboration for instructional improvement: A review of literature. *Educational Research Review*, 35, 100415.