

# Generating Traction with Continuous Improvement

Lessons from Two Learning Networks

H. Alix Gallagher  
Benjamin W. Cottingham  
Kiley O'Meara



July 2022

# Contents

<b>Executive Summary</b> . . . . .	<b>1</b>
<b>Background</b> . . . . .	<b>2</b>
<b>Outcomes for On-Track Networks 2020–21</b> . . . . .	<b>7</b>
Student Outcomes . . . . .	7
Traction of Improvement Work . . . . .	9
What Hubs Did to Create Traction . . . . .	13
<b>Conclusion and Implications</b> . . . . .	<b>30</b>
<b>References</b> . . . . .	<b>33</b>
<b>Author Biographies</b> . . . . .	<b>34</b>

# Generating Traction with Continuous Improvement

## Lessons from Two Learning Networks

H. Alix Gallagher  
Benjamin W. Cottingham  
Kiley O'Meara

### Acknowledgements

The authors would like to thank the CORE Districts and California Education Partners for welcoming us to their public events and internal planning sessions. We are also grateful for the time and insights that the staff at these organizations and all of their partner districts provided during interviews. We would like to thank Shawn Bernardo for his steady support of our data collection and the entire PACE team for their ongoing thought partnership. This report, like all PACE publications, has been thoroughly reviewed for factual accuracy and research integrity. The authors assume full responsibility for the accuracy of the report contents.

### Suggested Citation

Gallagher, H. A., Cottingham, B. W., & O'Meara, K. (2022, July). *Generating traction with continuous improvement: Lessons from two learning networks* [Report]. Policy Analysis for California Education. [edpolicyinca.org/publications/generating-traction-continuous-improvement](https://edpolicyinca.org/publications/generating-traction-continuous-improvement)

---

## Executive Summary

Continuous improvement has a prominent place in California’s approach to educational accountability. But while there are proof points that show the potential of continuous improvement, currently there is not evidence that continuous improvement efforts are consistently leading to sustainable improvement in student outcomes and system functioning. This report analyzes the experiences of two organizations serving as the hubs of improvement networks, both of which led networks seeking to increase the proportion of students on track for postsecondary success during the 2020–21 school year. We analyzed their work to learn lessons about how to improve the consistency of success of initiatives using continuous improvement.

We analyzed data to examine the improvement work of each school or district (which we refer to as “sites”) participating in a learning network. We used Coburn’s (2003) four dimensions of reform at scale—depth, spread, ownership, and sustainability—as indicators of *traction*, that is, work that is on a trajectory to affect student outcomes at scale. We then analyzed each site’s work with the hub to ascertain the level of traction the work obtained. Based on that analysis, we identified six strategies that hubs can use to support teams in gaining traction towards impact at scale:

1. Build trusting relationships to understand and support the sites that the teams are serving.
2. Hold space for teams to meet and provide accountability for moving work forward.
3. Make the work coherent with sites’ priorities and context.
4. Ask sites to identify cross-functional improvement teams.
5. Present continuous improvement as straightforward logic for making decisions using varied data, simple tools, and minimal jargon.
6. Help districts to select promising ideas to test.

These hub actions helped teams develop improvement work that was aligned with the rest of the site’s work and gained traction during the 2020–21 school year, even though most sites saw a pandemic-related rise in the number of students failing one or more classes.

We close with conclusions and implications for others (such as County Offices of Education working with districts or districts trying to support schools to improve in a specific area) serving as a hub for a group of districts or schools that are striving to improve.

## Background

In most education systems, many students—especially Black, Latinx, and socioeconomically disadvantaged students—do not finish K–12 with a full range of postsecondary options available to them. The Bill & Melinda Gates Foundation sought to address this challenge through the Networks for School Improvement (NSI) grant, which has dedicated hundreds of millions of dollars to this cause. NSI funds external support providers to serve as hubs for improvement networks focused on supporting high school students to be on track for college and career readiness. *Hubs* are organizations at the center of an improvement network that convene members and create the conditions under which the organizations in the network can learn from both the hub and one another. This report examines the work of two network hubs in California—California Education Partners (hereafter Ed Partners) and CORE Districts (hereafter CORE)—both of which have NSI-funded on-track networks, have prior experience as network hubs, and work with Policy Analysis for California Education (PACE) as their research partner. We use data from both hubs collected during the 2020–21 school year to generate lessons on how hubs of improvement networks can support the uptake of continuous improvement practices and improve students’ on-track outcomes.

*Continuous improvement* is a relatively broad term that encompasses multiple specific methodologies (such as improvement science, deliverology, the Shingo Model, and Baldrige), all of which share five main assumptions that distinguish them from traditional ways in which organizations try to improve:

- *systems*, not individuals, produce outcomes;
- improvement efforts focus on the design and operation of *processes*;
- improvement requires collective and systematic *learning*;
- *frontline workers* have unique expertise that is essential for improvement; and
- effective practices should be *spread* to become standard work (Grunow et al., 2018).

Both hubs’ approaches embraced these principles. Each hub also enrolled its district or school partners in the network by first building district-level interest in engaging in a network focused on improving on-track rates. Because Ed Partners typically works with smaller districts, all secondary schools in the district participated in most cases; in some mid-sized districts, the district selected a subset of schools for participation. The districts working with CORE selected schools for participation; none of these districts had all secondary schools participating. Each *site*—typically a district for Ed Partners and a school for CORE—created a team charged with leading the continuous improvement work. Table 1 describes notable similarities and differences in the hubs’ approaches and their networks.

**Table 1.** Similarities and Differences in Hubs' Approaches and Networks

	Ed Partners	CORE
<b>Network goal</b>	Increase the proportion of students on track for college	
<b>Network-wide measure</b>	At the start of the 2020–21 year, both hubs planned to use an innovative On-Track metric <sup>1</sup> developed by an external partner that used predictive analytics to determine how to weight student grades, course difficulty, attendance, and GPA in order to assign each student to one of five categories representing how on track a student was. During the school year, problems with the On-Track metric led both organizations to suspend use while it was revised. Ed Partners changed their focus to reducing Ds and Fs in mathematics and English language arts while CORE continued to work to improve the metric but did not create an interim replacement measure.	
<b>Network members</b>	Mostly smaller to midsize districts, typically located outside major metropolitan areas or in rural areas	High schools located in nine large urban districts
<b>Number of sites</b>	19	24
<b>Grade focus</b>	Two subnetworks: one focused on getting middle-grade students on track for high school and one for ninth grade	Ninth grade
<b>Approach to continuous improvement<sup>a</sup></b>	<ul style="list-style-type: none"> <li>• Informed by improvement science and deliverology</li> <li>• Greatest emphasis on the logic of setting a student-outcome goal, diagnosing system problems, testing a change believed to address a problem, and regularly monitoring progress towards the goal</li> </ul>	<ul style="list-style-type: none"> <li>• Informed most by improvement science and the Institute for Healthcare Improvement's Breakthrough Series model</li> <li>• Greatest emphasis on the Plan-Do-Study-Act (PDSA) cycle for learning from and aggregating small tests of changes</li> </ul>
<b>Network support structures</b>	<ul style="list-style-type: none"> <li>• <b>Professional learning events:</b> At these events, hubs present research-based information about improving on-track rates, provide time for teams to work, support teams in learning continuous improvement processes, and offer opportunities for cross-team learning.</li> <li>• <b>Connections with expert thought partners:</b> These are events where prominent researchers and practitioners present research-supported interventions—known as <i>change ideas</i>—to address the problems of practice that participants have identified.</li> <li>• <b>Data and continuous improvement tools:</b> Data dashboards help to track progress, and the tools guide teams through a continuous improvement approach to the work.</li> <li>• <b>Improvement coaches:</b> Each site is assigned a coach from the hub to support a site-based improvement team lead, facilitate team meetings, and provide differentiated scaffolds to support the team's progress; coaches in each hub typically work with five to seven sites.</li> <li>• <b>Creation of change packages:</b> These are designed to present relevant knowledge from research and practice that could inform network members' reforms.</li> </ul>	

<sup>a</sup>For a comparison of deliverology and improvement science as well as additional continuous improvement models, see LeMahieu et al., 2017; for more information on the Breakthrough Series model, see Institute for Healthcare Improvement, 2003.

<sup>1</sup> In this report, we capitalize "On-Track" to refer to the specific metric and use "on track" more generally to describe the chances that students will graduate high school with a full array of postsecondary options available.

The work that is the focus of this report took place during the 2020–21 school year, when the COVID-19 pandemic reshaped many facets of American life, including education. Disruption in education began in March 2020, when schools across California abruptly ended in-person instruction and many failed to provide consistent instruction for multiple weeks. To address the disruption, the California Department of Education required each local education agency (LEA) to develop a distance learning plan and stated that those plans should “ensure that grading policies hold students harmless for their spring grades and the transition to distance learning” (California Department of Education, 2020). Districts enacted a variety of policies that ensured students did not receive lower grades than they had when in-person schooling ended, and many districts instituted either pass/fail or credit/no-credit grading (Fensterwald, 2021). As a result, there was a sharp decrease in Ds and Fs in the second semester of 2019–20.

In 2020–21, California districts offered less in-person instruction than many states across the country. An analysis of state data showed that even at the end April 2021, 53 percent of California public school students were still in distance-only learning, 31 percent were receiving hybrid instruction, and 16 percent were receiving fully in-person instruction. Students in lower-poverty communities, in lower grades, and in the rural eastern parts of the state were more likely to have access only to hybrid or remote learning (Willis & Fensterwald, 2021). Almost all schools/districts participating in both improvement networks spent the school year in largely remote instruction, far removed from any sense of normalcy. Regardless of the mode of schooling available during 2020–21, most districts reverted to prepandemic grading policies, and the D/F rate increased statewide. Sites served by both hubs mirrored the statewide trends, with the vast majority seeing an increase in students struggling or outright failing one or more classes during the 2020–21 school year.

The pandemic also wreaked havoc with education data, which matters because data play a central role in continuous improvement. Organizations engaged in improvement use data to analyze the current shortcomings of a system, set goals, and assess progress towards meeting those goals. Yet COVID-related remote learning made it more difficult to interpret key data points about students’ engagement with and success in school, such as attendance and grades, making it more challenging for the networks to track their progress in improving on-track rates. Across California, the number of students failing or getting Ds in classes increased; this distorted the reference point for understanding whether a school or district was doing “better.” (Better than what? Their historical trends? Other districts in the state?) This also meant that across the board, fewer students remained on track than in a typical school year. Pauses in state-level testing also limited the amount of summative data available to schools and districts, making it difficult to assess year-over-year changes in student learning.

Finally, both networks had contracted with the same third-party vendor to develop a sophisticated early warning measure of how likely it was that each individual student was on track, referred to as the “On-Track” metric. The On-Track metric was based on analyzing historical data in all districts served by CORE and identifying ninth-grade outcomes (GPA, grades in A–G courses,<sup>2</sup> credits attained, and attendance) that were predictive of students graduating high school having met the A–G requirements and then successfully completing two years of college. The metric weighted the ninth-grade outcomes based on the extent to which they “predicted” students’ later success and placed students into five categories (ranging from “highly vulnerable” to “postsecondary competitive”) to indicate their likelihood of graduating high school, meeting the A–G requirements at graduation, and being successful in college. The hubs and vendor spent the 2020–21 school year negotiating revisions to the metric, so neither hub was able to offer this outcome measure to their network.

In response to the pandemic, both hubs—which initially paused most work during the spring of 2020—had to adapt their support structures to a fully virtual model. One positive implication for both networks was that school and district improvement teams generally found it easier to meet once travel time was removed from consideration and fewer nonteaching duties (such as teacher sponsorship of clubs) were competing for educators’ time. However, the hubs found themselves competing for attention with the effects of a pandemic, which typically hit schools and districts in the networks hard because most served low-income communities of color that were disproportionately affected. Individual educators on *improvement teams*—those charged with leading the network work at their sites—experienced the pandemic differently; some found themselves overwhelmed personally and professionally while others eventually hit a stride in the new environment. The pandemic’s effects on schools shaped the nature of our analysis and reporting as well. (For more information about the data and analytic approach, see the box “Data and Analytic Approach.”)

---

<sup>2</sup> In California, A–G classes are a sequence of high school courses in six disciplines that students must complete with a grade of C or better to be eligible for admission to the University of California and the California State University systems.



## Data and Analytic Approach

Data from the study for the 2020–21 school year include interviews with hub staff (some of whom were interviewed more than once) and participating educators, observations of events, and analysis of artifacts collected from events and teams’ work. The interviews included all hub leaders and coaches involved with the on-track networks as well as team members from 17 of 18 districts working with Ed Partners and 21 of 24 schools working with CORE.

- For districts working with Ed Partners, six districts considered to have mid-range or high engagement in the work were selected for virtual site visits (three to six interviews based on district size), and two additional districts were selected for in-depth case studies (11 and 13 interviews respectively) that focused specifically on how they were scaling continuous improvement processes in their districts. We interviewed team leads in the remaining nine districts during the fall and spring.
- For schools working with CORE, we attempted to conduct a virtual site visit (interviewing three to four team members per school) at all participating schools. Three schools declined to participate in interviews.

Role type	Ed Partners	CORE
<b>Hub staff interviews (coaches, leaders)</b>	20 interviews, 10 individuals	28 interviews, 15 individuals
<b>District/school staff interviews</b>	66 interviews 17 districts	77 interviews 21 schools, 9 districts
<b>Event/coaching observations</b>	51 hours	27.5 hours

Interviews were recorded and transcribed for analysis. Interview data were complemented by observation notes and documents produced by schools, districts, and the hubs themselves as part of the network activities, including reports on student outcomes being tracked by the schools, districts, and hubs.

In this context of disrupted schooling, interrupted data collection, and a sharp rise in Ds and Fs across the state, it is hard to calibrate data on student outcomes as a measure of network progress. Despite these challenges, we still present these data both as a reference point for later years and because students' academic performance still had real consequences for the students' on-track trajectories. It was nonetheless clear that improved student outcomes were an insufficient signal of where the improvement work was more or less successful. Instead, the research team turned to existing education research to identify attributes of reform associated with sustained, long-term improvement in student outcomes, which are presented next. Using a framework we developed based on this extant research, we present results of our analysis on the extent of *traction*—defined as work that is on a trajectory to affect student outcomes at scale—attained in both networks. The data we collected showed that the work gained traction in about half of the sites, but some sites in both networks almost entirely failed to engage with the hubs. The crux of the report is our analysis of which hub actions appear to be related to increased traction of the work at participating sites. We conclude with a summary and policy implications for hub agencies (such as County Offices of Education working with districts in differentiated assistance and districts trying to support schools to improve in a specific area) leading learning networks.

## Outcomes for On-Track Networks 2020–21

Examining student outcomes is one way to understand what these networks accomplished during the 2020–21 school year. Interpreting student-outcome data as a measure of hub success during 2020–21 was challenging, however. The typical pattern across California was a sharp drop in Ds and Fs in the second semester of 2019–20, followed by a rise to higher-than-typical rates of Ds and Fs in 2020–21 (Tadayon, 2021). This is because, as described previously, the number of Ds and Fs dipped sharply at the end of the 2019–20 school year even while many students made less academic progress than would have been expected of them during a typical school year (Pier et al., 2021), leaving them less prepared for their courses in 2020–21. Because student outcomes were such a poor indicator of progress during the 2020–21 school year, we also analyzed qualitative data on the work in both networks and found notable differences in the extent of traction (that is, the progress and momentum) that this work attained in schools and districts across the networks, which research suggests is a harbinger of impacts on student outcomes.

### Student Outcomes

Both hubs set goals that focused heavily on ninth graders' grades, although they tracked progress in slightly different ways internally. Using available data, we describe student outcomes for 2020–21 while acknowledging the changes to schooling and education policy associated with COVID-19.

**Student outcomes tracked by Ed Partners.** When it became clear that the predictive On-Track metric (described in Table 1) was not going to be available for the 2020–21 school year, Ed Partners permanently switched their outcome focus within the network to the D/F rate (that is, the percentage of student grades that were Ds or Fs). All ten districts that submitted data for both the 2019–20 and 2020–21 school years had an increase in the D/F rate: Four had an increase of 1–10 percent, and six had an increase of 11–20 percent across those years.

Ed Partners also asked districts to track their progress towards an aim related to reducing the D/F rate. Aims typically focused on reducing the D/F rate in a student subgroup, for example: “By June 2021, we will decrease the D/F rates of Latinx eighth-grade students in [regular] Math 8 by 5 percentage points to support the increase of on-track achievement.” Of the ten districts that shared data about progress towards their aim, six reported a decrease in their D/F rate, two reported an increase, and two reported mixed results (an increase in one subject area and a decrease in another). Combined, the consistent pattern of an increase in the overall D/F rate for all districts and the mix of results that teams self-reported on the more focused aims suggest that many districts saw success at improving outcomes, but it was at a small scale.

**Student outcomes tracked by CORE.** In the absence of the On-Track metric, CORE did not have schools set an overarching student performance goal for the 2020–21 school year. Instead, teams set shorter term aims, typically improvement in students’ responses on a survey administered at three points during the year that measured students’ relationships with adults at school. This measure, the Search Institute’s Developmental Relationship Survey, was selected based on work on the importance of developmental relationships for student success.

CORE used the On-Track metric to assess progress for the 2020–21 school year retroactively. CORE’s network defined *off track* as being in the lower two lanes of the metric, which meant that students had ninth-grade outcomes similar to students who historically had not graduated high school or, if they graduated, did not meet A–G requirements. Of the 21 schools for which data were reported, 19 had an increase in off-track rates in 2020–21 compared to the previous year, and two had a minor decrease (1 and 2 percentage points, respectively) in their off-track rates.

In sum, neither network saw the positive impact on student outcomes that it would have wanted during 2020–21, but the outcomes most sites achieved mirrored statewide trends for this period. To get a better sense of whether the work proceeded as desired, we shifted to focusing on evidence that participating teams were beginning to gain positive traction in their work.

## Traction of Improvement Work

Analysis of qualitative data showed that some teams in each network struggled to participate while others made notable progress towards making changes in their districts or school systems that could lead to improvements in student outcomes. We anchored our analysis in the existing research base to identify features of early implementation that are likely to lead to reforms gaining traction towards having a substantial and sustained impact on student outcomes.

Existing research includes ample evidence about the factors that support or impede the take-up, success, and institutionalization of reform. In this analysis, we use Coburn's (2003) four dimensions of scaling reform to identify whether schools/districts are gaining traction on changes to their system that are likely to have an impact on student outcomes at scale:

- **Spread:** Principles central to the reform are taken up by increasing numbers of people across a system and are embedded in formal policies and organizational culture.
- **Depth:** Consequential reforms shift education system beliefs, norms, and pedagogy.
- **Ownership:** Reforms are often required and/or supported by external mandates or organizations, but districts, schools, and teachers must own the reform for it to stick.
- **Sustainability:** Adopted reforms are only significant if they persist over time.

Coburn's four dimensions of scale are highly relevant to continuous improvement. For reforms to make a difference in student outcomes, changes must penetrate beyond the superficial and instead affect instructional practices and beliefs. Educators within a system must feel a sense of ownership over reforms as they will be responsible for adapting changes to make reforms most effective in a specific context. In alignment with continuous improvement's focus on systems and processes, teams—not individuals—must also determine how to spread reforms within an organization. Without all three elements of depth, spread, and ownership, sustainability is impossible; even if changes are taken up by some within a school or district, the work will remain isolated and will often be abandoned with shifts in resources or staffing. For this reason, we see a progression through Coburn's dimensions of scale over time, where sustainability cannot be engaged without the work having depth, local ownership, and some degree of spread already in place. To see where the work was "gaining traction," which we defined as movement on Coburn's dimensions, we used the four dimensions of scale to examine the work being done by sites in both networks.

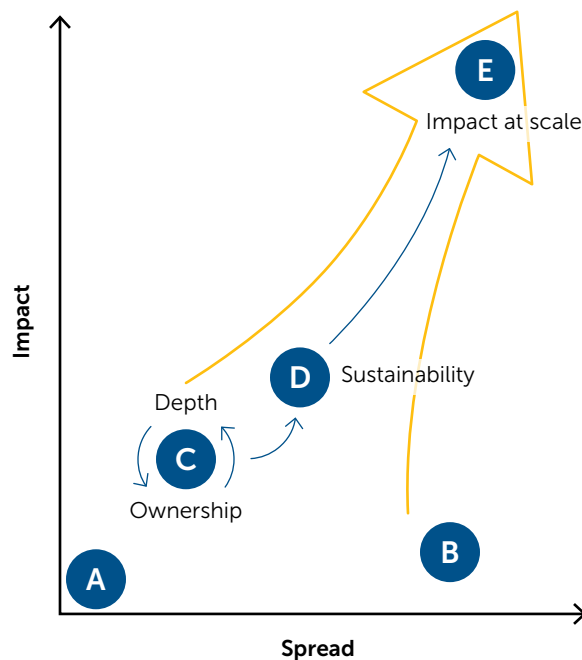
There is no expectation that any site would have fully realized all dimensions of scale midway through the grant period, but teams' levels of traction could be examined along a continuum using Coburn's dimensions. Analyzing patterns in the data, we extended Coburn's theory to place the trajectory of sites' work along two axes (Figure 1). The x-axis can be thought of as the spread of the work, starting with minimal engagement with the hub or in improvement work on the left and moving to organization-wide spread of the work to affect all of the desired

students on the right. All sites started the work with a team, but sometimes the team had only one or two engaged members. Spread would mean increasing the number of adults directly engaged in work with the hub, having teachers or other school staff who are not directly engaged with the hub either trying or implementing ideas, and ultimately having the work occur across organizational structures (for example, departments and grade levels) as necessary to reach all students. The y-axis represents a progression of impact of the work, showing a range of effects on student outcomes from insignificant to a high level of impact.

We theorize that there is a natural progression of Coburn’s dimensions on their way to achieving impact on a large scale that runs through the arrow from the bottom left to the top right quadrant. At a foundational level are ownership and depth in a mutually reinforcing feedback loop. If there is no champion who owns the work, deep work will never get off the ground; unless the work is of sufficient depth to make a difference, over time no one will champion it and the status quo will return. In-depth work can only spread if it becomes sufficiently owned, in which case members of the site’s community will assume leadership roles in driving increasing depth and spread of the improvement work. To achieve sustainability, the work must be spread into the organization’s structures and processes, making it less vulnerable to changes in leadership and having it become “the way we do things here.”

Figure 1 also represents, running from left to right just above the x-axis, the fact that changes can spread that are shallow, lack ownership, and/or are unsustainable; no matter how broadly a shallow change is spread, it will not scale or have an impact on student outcomes. This is the fate of many attempts at reform in education.

**Figure 1.** Zones of Traction of the Work in On-Track Improvement Networks



In the remainder of this section, we describe the results of our analysis of the traction of work in sites across both networks. The following are descriptions of what work in each zone looked like in the schools and districts participating in improvement networks with both hubs:

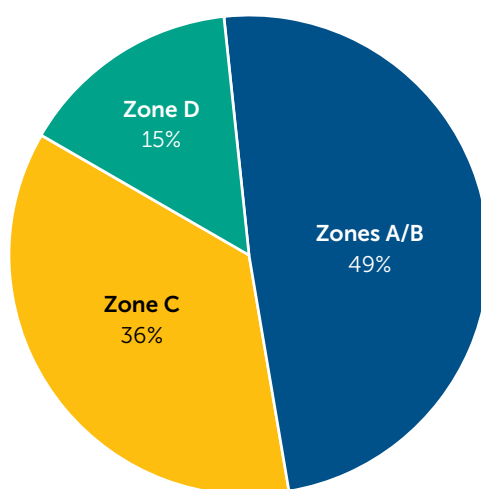
- **Zone A—no traction, little to no evidence of improvement work:** Indicates sites in which improvement work has had little to no traction. Some sites were clear in that they were “taking a break” from the work or had yet to establish fundamental structures, such as assembling a team dedicated to the improvement project. Others in this category had a person or two devoted to improvement, but they were the only ones attempting to engage, and there were obvious structural and cultural barriers to scaling the work.
- **Zone B—low traction, superficial change, and teams spinning their wheels:** Indicates sites in which some team members had been able to launch some improvement-focused initiatives, but the change ideas were either dubiously related to outcomes or not grounded in evidence from research or established practice. In these instances, people may have been doing reasonable things, such as sending out email requests to parents to sign up for an app, extending email invitations to students to join Google classrooms, or hosting an ice cream social. All of these examples can be part of the constellation of strategies that support students, are relatively easy wins, and have easily quantifiable outputs (e.g., counting parent sign-ups and/or event attendance). However, the strategies are at best distantly related to the goal of substantially increasing the number of students graduating from high school ready for college. Thus, even when improvement teams had actively tried strategies such as these with increasing spread (for example, across the entire ninth grade), their overall initiative still landed in Zone B—low traction/low impact—because the work lacked depth and ownership and so had no path to sustainability and scale.
- **Zone C—some traction, small tests of change, and with potential to scale:** Describes small individual-level change and includes sites with a limited number of engaged team members testing research-based strategies with the potential to improve student outcomes substantively. These sites did not have all the structures, policies, or administrative support required to scale and sustain change, but they did have that initial spark of improvement work and a fully invested—if small—team (that is, depth and ownership but low spread). In one site, the improvement team was testing one-on-one interventions with at-risk students, tailoring supports and check-ins with faculty and staff. Another was testing intervention models with math instructors at the high school. In these instances, changes were being tested, but learning generated was often not being spread to the broader site community.

- **Zone D—traction, spreading changes that have been proven effective in small tests in this context:** Exemplifies reform efforts that are of sufficient depth to affect student outcomes and are positively affecting system functions. Sites in Zone D operated with a team engaged in continuous improvement and supported by administration (site ownership) as well as structures and policies to help spread learning and innovations. Two examples of this type of work came from small districts with deeply engaged superintendents who were able to help engineer and support systemwide changes to bolster and spread the change launched by the team working directly with the hub.<sup>3</sup> Another example was from a large comprehensive high school with a strong relationship with a County Office of Education coach and a long-tenured principal who had invested significantly, over time, and in multiple ways in developing staff capacity around improvement. The improvement team met weekly with a high degree of transparency and accountability, used supporting data to define challenges and growth areas, and scaled changes via collaborative learning structures.
- **Zone E—impact at scale:** In this most advanced state, sites would have a demonstrated track record of scaling changes that are having a positive impact on student outcomes and change ideas would be successfully spread across the system. In the process, there would be evidence of all four dimensions of scale named by Coburn. The data set did not yet include any examples of teams that had attained this, as is clear from the student-outcome data previously presented.

We analyzed interview data to identify the level of traction in each site according to the five zones (Figure 2). The analysis did not attempt to distinguish sites with minimal work occurring (Zone A) from those where the work lacked depth (Zone B) because functionally, sites in both these zones were not gaining the traction necessary to achieve impact at scale. But we did analyze data to identify whether the work had sufficient depth (Zone C) or had both depth and initial spread of testing beyond just the improvement team (Zone D). The results show a modest proportion (15 percent) of sites have substantial traction (Zone D); the work is deep and is increasingly owned by educators at the site, they are positioning themselves for sustainability, and they are intentionally spreading the work to colleagues who are not on the improvement team. In an additional set of sites (36 percent), there is some degree of traction (Zone C); the work is deep, and a small group of people are engaged and feel ownership. About half of the sites (49 percent) did not gain traction during 2021–22 (Zones A and B).

---

<sup>3</sup> For more information, see Cottingham et al. (2022).

**Figure 2.** Distribution of Traction Zones for Sites in On-Track Networks

*Note.* These data represent the districts and schools where we had qualitative information about how they were engaging in the work of the on-track network. These include all 19 districts working with Ed Partners and 21 of the 29 schools that started the 2020–21 school year with CORE. Reports from CORE and Ed Partners staff suggest that the sites that did not respond to requests for interviews were more likely to have low levels of traction (Zone A/B) in their work.

The next step of our analysis focused on identifying strategies that enabled the hubs to support teams' movement from Zones A and B (where minimal scalable change is occurring) to Zones C and D (where tested strategies are showing evidence of impact and are beginning to be scaled throughout the system).

### What Hubs Did to Create Traction

Schools and districts did not experience their work in these networks the same way because they entered the work in different places, hubs purposely differentiated the work based on coaches' understanding of sites' needs, and there was variation in how individual coaches within each hub enacted their roles. As the preceding section described, sites have made different amounts of progress. As we analyzed data from both networks to see where the work was gaining more or less traction, we identified strategies the hubs were using that appear related to sites building traction with this work. We identified six major hub moves that increased the chances that the work would gain traction: (a) building trusting relationships, (b) providing space and accountability for the work, (c) aligning work with site priorities, (d) supporting selection of cross-functional teams, (e) presenting continuous improvement as straightforward logic, and (f) funneling sites to promising change ideas.



**Coaches started by building trusting relationships to understand and support the sites they were serving.** Because coaches were external and coming to work within an unfamiliar system, coaches' relationships with the teams they supported were named by each organization as foundational for the work. Continuous improvement sometimes requires examining unflattering data and challenging long-held assumptions. Coaches first needed to establish trust and familiarity with the improvement teams they supported to engage productively in the conversations necessary for driving change.

Coaches faced pushback when they attempted to implement changes and new strategies without having established trusting relationships and deep contextual understandings of the sites they were serving. In these instances, site leaders expressed that they felt disconnected and that the coaches undervalued site expertise. One site leader reflected:

*There are times when I feel like [the hub] has deadlines to meet with whoever funds them so they need all of us to stay on the deadline so that they can turn in whatever they need to get their funding. ... I don't often feel that our coaches are invested in what our projects are or even know.<sup>4</sup>*

The central strategy of successful coaches was to take the stance of a thought partner, which one hub leader described as a mix of listening, supporting, and at times nudging for change but not leading it. The hub leader explained that sites are not expecting hubs to solve their problems but instead need assistance to build the muscles necessary to sustain continuous improvement—a process that sites must own if they are to meet future challenges after the hub has left. This positions coaches not as the leaders of reform but as supporters of the leaders, a role coaches can accomplish only if they have knowledge of the local context.

One strategy coaches used to establish trust with improvement teams was taking time at the beginning of a new partnership to understand the full context of ongoing work in the site's system. By conducting a system analysis, hub coaches were able to work collaboratively with teams to understand system priorities, identify relevant system structures and processes, highlight levers for potential change, and analyze existing outcomes. This process gave coaches an empathetic entry point, establishing themselves as learners first and foremost before engaging in critical conversations about what needs to change. System analyses provided opportunities for sites to name potential shifts and needs prior to the hubs' introduction of new strategies and concepts. Additionally, the teams led these conversations while hub staff listened, giving team members a sense of ownership of the work and an active role in framing and diagnosing their own problems as opposed to waiting to be told what to do and what to change.

---

<sup>4</sup> Throughout this report, potentially identifying information—such as names of people and organizations—has been redacted from quotations.

One coach explained the connections between doing a system analysis, developing relationships, and establishing the trust necessary to support sites' reform work:

*Having authentic relationships and developing context, building your understanding of where the district is; we say that a lot, but I'm realizing that was a huge lever for me. And having those authentic relationships gave me the ability to build their trust. Them being more open to be coached, them being more open to trying new things. ... [D]eveloping a better understanding about their context helped me build that trust also, because they were like, "Oh, you really do listen. You're really understanding. You really do know us." And it helped me further focus them.*

A site leader's perspective corroborates this understanding of why coaches' learning about sites' context was critical:

*Unlike all the other program people we've had [through prior initiatives], [our coach] is the best at actually listening to the needs of [the site], and helping us to adapt things properly. We've never ever had that before. ... I feel like [they're] listening to what we say.*

Taking the time to understand ongoing system work and surface the deep contextual knowledge that improvement team members brought to conversations gave coaches a window into the system that allowed them to target more specifically the high-leverage actions that would further improve the work.

**Hubs played a critical role by holding space for teams to meet and providing accountability for moving work forward.** Both hubs sought to create and protect time for sites to collect, analyze, and reflect on information about their systems as well as to design and test changes. When hubs were not able to get sites to provide these spaces, the work never launched. Given all the demands on educators' time, this hub function was consistently named by sites as critical for traction. Participants across both hubs noted that coaches played a critical role in holding their team accountable to meet and make progress in their improvement work. In conversations with teams, coaches acted as strong thought partners, listening and asking relevant probing questions to guide teams as they developed, executed, and analyzed tests of change. One site leader described how they felt the hub supported their improvement team in continuing and improving existing work:

*We were already doing some [improvement] work, but we didn't have a system in place in order to perpetuate it, or a system in place to go back and double check to make sure that we're being accountable for what we say we're going to do. So, I would say the [hub] is kind of keeping us on track with a lot of endeavors that we already initiated. ... [I]t's really easy to lose things when you don't have a system in place of how these pieces fit together.*

Dedicated time and space with an experienced coach helped teams build traction by staying accountable, developing change management capacities, and ultimately building depth and ownership in their work.

**When the work with the hub was coherent with sites' priorities and context, it was more likely to gain traction.** It is widely recognized that if a new initiative does not become coherent with the broader goals, norms, and culture of an organization, it will not be sustained. Honig and Hatch (2004) note that coherence is not a permanent state but is instead crafted by the people within organizations. Park et al. (2022) have named *weaving*—integrating the work into the existing fabric of the school—and *stacking*—adding new work and specific processes or tools on top of the existing system—as two main ways that actors in systems engaging in continuous improvement have attempted to make the work coherent in their systems. Interviews with teams participating in these networks suggest that when hubs facilitate weaving, sites are more likely to gain traction than when teams see the continuous improvement work as either additional work or in opposition to existing priorities and approaches.

When work was not woven into the system and was seen as a stacked “add-on,” it greatly decreased the likelihood that practitioners and leaders would engage with the work at all. In some instances when stacking occurred, site leaders did not see how the work with the hub aligned with the priorities they were trying to set for a site. One coach described that experience:

*I would have team meetings with the lead and then I would have meetings where the administrator would want to meet with me and then undo everything that the leads were doing. And you could tell that they had very different goals. ... We've planned PDSAs, but [the site leader] would then come back and say, “No, you can't do that.” So that was a really rocky start.*

Even if the new approach was not seen as oppositional to the existing site, if it was seen as a separate initiative, it risked not having sufficient support to gain traction. One stressed site leader explained his unwillingness to push the “hub agenda” on educators at the site given everything else competing for their time and attention, especially during the COVID-19 pandemic:

*The problem is we are trying to implement things [as a site] as well, and that's already difficult. ... With all the things that we're trying to also do at the same time, we haven't pushed a [hub] agenda on the team and I just don't think that the other VPs have time for it.*

The work at this particular site ended up stalled with a set of individuals who were personally willing to engage but without any viable options for scaling the work (that is, without any traction).

On the other hand, weaving the work with the hub into existing site-based goals allowed site leaders and practitioners to take ownership and control of the work, and ultimately, it was more likely to gain traction. Coaches typically played the role of bridging the on-track initiative with the site, based on a deep appreciation for local context and leveraging system leaders' knowledge of history, politics, relationships, and capacities required to make new initiatives successful in the local context. A coach described this process:

*Meeting [sites] where they are, they're seeing some needs. How can we then incorporate that into the work? It's valuing what they have and making sure that we're not extra. We're doing the things that you're trying to work on already. It brings them to the table more often.*

As the work developed, coaches also recognized that scale required that they gradually turn over ownership to educators in the site to build local ownership of the work:

*I think it comes down to their true conviction of, "Do you believe this work is your [site] work," right? If they see it as something different, it's going to be harder for somebody to move on things in addition to their thousand other things that they have to do that day. However, if they see the benefit of why they're doing all the hard work and their brain's hurting now for the benefit of their students or for the benefit of the support of their teachers, they're more likely going to do it and they're more likely going to have that buy-in and understanding of the time needed to do that.*

When hubs prioritized coherently weaving new ideas and approaches into the existing fabric of sites' systems, instead of stacking on too many externally developed requirements and tools, leaders were more likely to make the work a coherent part of their systems.

**Hubs were better positioned to affect systems when sites developed cross-functional teams.** Starting an improvement team can be seen initially as "additional" work, and leaders often faced a quandary when selecting members for improvement teams. Many understandably took the path of least resistance, selecting those who were most willing. Experienced teachers were often reluctant to sign on for the work, which led to new teachers being overrepresented on many teams. In other cases, the team composition represented one group of adults in a site—in the most extreme case, one team had no teachers and was composed solely of counselors and staff who ran a supplemental program—making it difficult for the work to influence the site as a whole. When the improvement team was composed of novice teachers or staff who identified with the same subgroup of adults in a school, the work did not have many opportunities to gain traction as a schoolwide initiative.

Membership on the most effective teams typically included both administrators and teachers as well as other staff, such as counselors or data specialists, if the focus of the work overlapped with specific roles or expertise. Teams that cut across typical hierarchies and silos positioned the work for success in a variety of ways, such as increasing opportunities for the work to spread across the system. Both hubs emphasized the need to include staff who crossed vertical hierarchies as well (for example, teachers, school administrators, district staff, and district administrators).

Administrator participation at the school and, ideally, the district level signaled the importance of the on-track work, which engendered confidence and accountability and showed that participants' time was valued. As one team leader explained, having an administrator on the team reinforced the legitimacy of the work, and that "creates a better or a higher sense of urgency to support the work that is being done." Having a team member with decision-making authority was also helpful for allowing teams to test and, when appropriate, spread their ideas quickly. Interviewees expressed their sense of disempowerment that occurred when administrators were disengaged from or unaware of the work, as one frustrated participant shared: "[Our administrator] has not shown to our last two meetings, no emails, no cancellations. ... He says he kind of knows what [the improvement work] is, but I'm not really sure that he does." The most common concern raised was that meeting time was inconsistent or taken over by other priorities if administrators did not protect the time. Across the board, involved administrators signaled the importance of the improvement and helped teams garner the necessary resources to support growing, promising work.

In schools or districts where the improvement work gained traction, administrators viewed the teachers and counselors—whose daily work shaped students' experience—as an essential source of knowledge, vital to exploring how promising approaches from elsewhere could be adapted and integrated into their local context. As described previously, continuous improvement relies heavily on the knowledge of people on the ground who implement most of the changes being tested. Communication in schools is often constrained by hierarchies (such as the district office, school administrators, department chairs, and teachers) and siloes (such as the department and/or grade and academies within high schools). Information about new ideas and practices traditionally flows top-down and within personal and professional networks. Those improvement teams that achieved more traction elevated teachers' voices and pushed their expertise into professional learning communities, departments, and leadership meetings. Teachers in these schools felt listened to, as a teacher explained:

*Even though I am a lowly teacher, I know that I can say something to our superintendent, and he's not going to ignore or put me down in any way. That's something that we learned [in work with our hub]. Every single voice is definitely important, and we make time to hear from each one.*

Teachers' knowledge and perspectives drove the improvement work, and administrators had enough understanding of teachers' work to respond to challenges supportively, address needs, and elevate early wins. This shift in knowledge flow is a central tenet of continuous improvement and was consistently evident in sites that gained the most traction.

Highly functional teams considered the informal influence as well as the positional authority of members and strategically structured teams so that work could be spread across the system. For example, in one meeting we observed a team that was planning whom to recruit the following year to test a new and potentially contentious change idea. The team ultimately tasked a team member (who was also a department chair at one school) to recruit a colleague who was the department chair at a different school, knowing that if the department chair at the second school came on board, the team would have made inroads for the work in two schools. Hubs' attention to strategic team formation from the beginning positioned teams well for picking up traction as the work got underway.

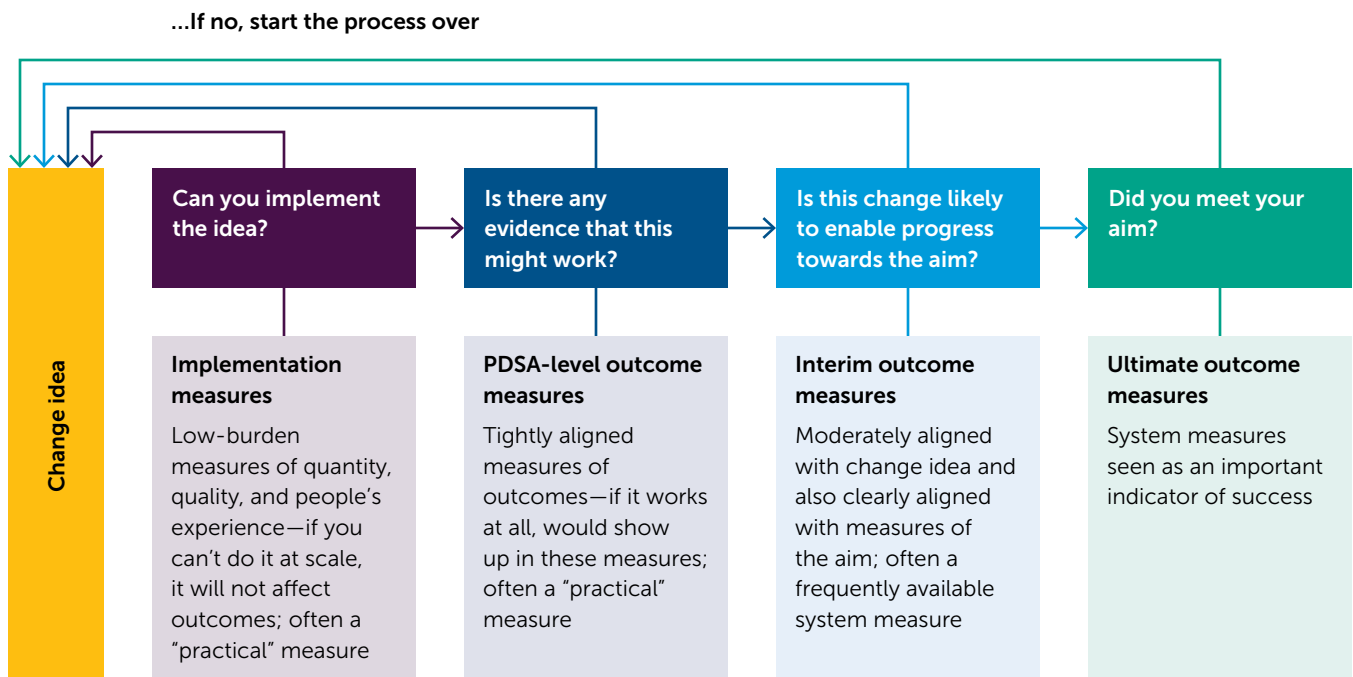
**Continuous improvement gained traction when hubs presented it as straightforward logic for making decisions using varied data, with simple tools and minimal jargon.** Continuous improvement requires frequent data use and, often, types of data that are not commonly used in educational practice (Bryk et al., 2015; Hough et al., 2017). For example, annual test-score data provide information that is too time-lagged to support iterative improvements in instruction or other student supports (students typically take tests in the spring, and results become available in the fall). Using data for improvement requires a different approach, one tied closely to the logic of the change being tested (Grunow et al., 2018; Hough et al., 2017). This section describes what we learned about how hubs can support improved data use to help sites gain traction in their improvement work in terms of (a) the logic and associated data of continuous improvement, (b) the challenges hubs and sites faced around data use, and (c) high-leverage data moves that the hubs made.

***The logic and associated data of continuous improvement.*** Continuous improvement makes a sharp departure from typical reform after the problem is identified and people select a change to test in their context. The Model for Improvement (Langley et al., 2009) starts with two questions that are typical in any strategic attempt at reform: "What problem are we trying to solve?" and "What changes might we introduce and why?" After this point, continuous improvement diverges from more traditional approaches to implementation by engaging people in cycles of learning in order to answer the question, "How will we know that a change is an improvement?" The iterative work to answer that question follows this straightforward logic:

1. Can a change be implemented with fidelity and feasibly in a given context (tracking quantity and/or quality indicators for implementation)?
2. Might the idea matter for student outcomes when implemented reasonably well (on a small-scale indicator of progress that provides relatively rapid feedback)?
3. When the idea is spread to others, does it affect interim measures of student outcomes and ultimately lead to progress towards the aim (without causing undesired negative consequences)?

If at any point in time, the answer to the question is “no” (for example, “no, we cannot implement it as designed” or “no, it does not seem to make a difference at a small scale or on a fine-grained measure of outcomes”), then the process returns to the start of the cycle for adaptation or to find another change idea to try. If the answer is “yes” to the first two questions (for example, “yes, we’ve figured out how to get students to participate and/or adults to do their work differently” *and* “yes, scores on the posttest showed more growth than with typical practice”), then the next step is to spread that practice and ultimately conduct a broader test of the change to see if the change leads to improvement towards the aim. Throughout, the team should monitor interim outcomes (such as nine-week progress reports), look for impact on the less frequent measures of student progress that are most closely related to the project aim when those become available (such as semester grades or state test scores), and assess whether the change is causing unintended negative consequences (e.g., students losing access to electives to participate in an intervention). Figure 3 represents those questions along with measures typically associated with them.

**Figure 3.** The Logic of Continuous Improvement and Associated Measures



---

Traditional accountability approaches have inspired practices where someone implements a change and then assesses at the end of the year whether they met the aim; the data are frequently insufficient to understand why they did or did not meet the aim, and those data are too lagged for midcourse corrections. Data-driven decision-making approaches improve upon traditional accountability by also using interim outcome measures to see if there is any progress towards the aim. Continuous improvement asks two additional questions that are neglected in most other approaches: (a) Can you implement the idea, and if not, can you modify it so it is feasible in this context? and (b) Before you ask everyone to implement it, is there any evidence that it might work in this context? These additional questions are central to the power of continuous improvement.

Attempts at reform often fail because some aspect of the context is a barrier to good implementation (for example, there may be conflicting demands for time or teachers may need additional support to implement the new practice well). Refining implementation at a small scale is one way that continuous improvement can prevent the churn of new programs that do not lead to improvement in school systems. Concurrent with initial tests of implementation are small-scale, typically quick tests for evidence that the idea works. These cycles are ideally rapid and conducted long before the change might be expected to have any noteworthy outcome. One central task in both on-track networks was to reduce the number of students failing classes. The example in the box illustrates how a hypothetical school might apply the logic of Figure 3 to testing approaches to decreasing student failure rates.



## Testing Approaches for Reducing Student Failure

As described previously, students failing classes is a key barrier to their being on track for college and career success. If a system analysis suggested that one reason students were failing classes was because of missing work, a teacher might hold an afterschool make-up and support session. A sign-in sheet would provide information about whether the teacher was able to hold the session with the desired students attending (Can the idea be implemented?). The teacher could track student completion of assignments as a PDSA-level measure (Is there any evidence that the idea might work?). If some other demand on the teacher's time meant the session was canceled, students did not attend, or they did not complete work, there is no need to wait for a more interim measure (such as nine-week progress reports) to go back to the drawing board to adapt or abandon the idea.

Although answering these questions is appropriate for initial small-scale tests of an idea, it is crucial that an interim outcome measure also be examined to learn if the change is likely to lead to progress towards the actual aim. Sometimes the initial theory that led to selecting a specific change idea is incorrect or incomplete. Continuing the previous example, it would be important to examine students' progress reports after a few weeks of afterschool make-up sessions because they provide an interim measure of students' achievement in classes and so are likely more predictive of students' final grades than turning in work alone. Examining that measure would let teachers know if the change they tried put them on target to meet their aim or if they might also need to provide some other intervention (such as tutoring) to help an additional set of students raise their grades.

Because this type of approach is unusual in most education systems, understanding the existing data sources, how data are accessed and by whom, and how the data are used is critical knowledge for a hub because it helps them see existing assets as well as areas of need. Data analysis also serves an important function for hubs in motivating teams to undertake substantial work that might otherwise be ignored: identifying system needs that might have been overlooked and helping teams assess if what they are testing is meeting desired outcomes. The experiences in these networks around data use are in line with prior research—namely that continuous improvement required a broader range of purposes for using data shown in Figure 3, which in turn required different types of data (Table 2).

**Table 2.** Purposes and Types of Data Used by On-Track Networks

Purpose	Type of data
Understand initial performance, set an improvement aim, and assess whether the aim is met.	These are system data that measure an outcome that matters; they may not need to be frequently available (for example, student A–G completion rate or semester grades).
Get rapid feedback on whether new approaches (a) are feasible and (b) are promising.	Sometimes called a “PDSA-level” measure, these data are rapidly available and tightly aligned to a particular idea being tested (such as student exit slips, end-of-week quizzes, and informal teacher notes on student engagement). These are unlikely to be system-level data, so care needs to go into designing practical measures that are good enough to inform learning about improvement but do not take too much time to collect.
Monitor progress towards the aim.	These are interim, regularly available outcome measures that are predictive of aim (such as six-week grade reports, daily attendance, and classroom observations).
Learn how people experience the status quo as well as attempts to change the system.	These data are typically qualitative and used in different parts of the continuous improvement process (for example, to gain insights into the initial performance gap, understand experiences of people trying to implement changes, or understand how students experience changes being tested). These often are also practical measures.

To support the range of data use required, the hubs developed data infrastructures (for example, they created dashboards to support data visualization or brought new measures into the networks); taught routines for data use (for example, they developed protocols for examining data or made templates to embed data analysis into the continuous improvement process); and had coaches provide customized support to each team (for example, they filtered data to make it easier for teams to see progress for focal students or modeled leading improvement-oriented data conversations). Given the time and resources invested by hubs, it’s important to understand the strategies that were more or less effective in helping sites gain traction in their improvement work. In some instances, sites needed the full support of their hubs to build data infrastructure and their data capacity while other sites needed a lighter touch and highly specialized coaching. The hubs’ work suggests that elevating straightforward logic and using existing or easy-to-gather data are more likely to be successful, although there are times when it could be worthwhile for hubs to introduce technically advanced data sources or tools.

**Challenges the hubs and sites experienced around data use.** Both hubs made innovations around data quality, which had some benefits for their networks, but both also encountered the same challenges at sites where complex data visualizations did not match users’ data capacity or align with site-specific data needs. One example of the intersecting challenges of innovations around data quality came from a coach who described sharing a new dashboard that displayed data from an innovative measure:

*I gave the [site] team leads access to the dashboard for the first time. They love data, and they looked at it and they were like, “I don’t know what to do with this. It’s cool but I have no idea where I could even begin with this. I can’t even look at it.”*

A second example from a site that relied on its own internal data even though the hub had built a more sophisticated dashboard illustrates additional challenges to successful data innovation by the hub:

*I have used [the hub dashboard]. [But w]e have a data dashboard for [our site]. I do three-week grade pulls for our site already because of monitoring. How do I say this? It's helpful, but I feel like we've got a lot of tools in that category.*

A third illustration comes from an innovative survey that one hub asked sites to administer to all students in a focal grade level. Many sites liked the survey, and most thought it provided insights into their students. But several improvement teams focused much of their work on getting a high enough response rate to make the surveys useful, in the process diverting the improvement teams' time away from testing change ideas designed to improve students' outcomes directly.

These quotations and examples illustrate a series of interrelated challenges to the widespread uptake of the innovative data sources and tools that the hubs offered:

- **Time:** Collecting and analyzing data required substantial educator time. Coaches spent a lot of team time explaining new data and how new metrics related to existing work. Hubs devoted substantial time and other resources to creating data sets and tools.
- **Frequency:** Continuous improvement relies on getting rapid feedback on whether change ideas appear promising in a context or need to be adopted or abandoned. Existing student information systems typically provided more up-to-date information than the hubs could on important outcomes like grades and attendance.
- **Redundancy and convenience:** In multiple sites across both hubs, administrators reported having an existing measure in their system that was similar enough to the one offered by the hub that they used the familiar measure rather than the new one. While hubs valued a shared outcome measure for the network, local sites often continued to focus on their local measures instead of on the shared measure.
- **Infrastructure:** When hubs created new tools and metrics, sites' infrastructures did not necessarily have the capacity to enable relevant educators' access to the data.
- **Expertise:** In many sites, educators did not have established, productive ways of using data (and may have had ingrained suspicion of data based on prior punitive use for accountability). In smaller districts, there also may have been limited staff resources for basic data functions beyond state-required data uploads. Finally, when data resources were complicated (such as the On-Track metric), even members of the hub staff encountered challenges communicating about the data clearly and convincingly.

Of these challenges, the easiest to overcome were expertise and infrastructure, both of which could be at least partially addressed in the short term by an influx of hub resources.

The others proved more intractable, which does not imply that hubs should never have a new data source or infrastructure as the center of their improvement networks. It could be worth the investment to overcome challenges inherent in building new data infrastructure when the systems in a network consistently lack valid and reliable measures of a substantively important outcome. These hubs' experiences, however, also suggest that the default course of data action should be focused on helping sites take up the straightforward logic of continuous improvement into their work and select data purposefully.

**High-leverage hub moves around data use.** Most often, simple approaches created the greatest return on investment because improving infrastructure was costly. Sites had varying levels of data capacity, and this affected how they engaged with data and how hubs supported them. Given the variations in data infrastructure across the sites these hubs served, it is not surprising that the hubs needed to provide a range of data supports to facilitate continuous improvement. But, as a recent report on scaling innovation in education noted:

Schools need to be pragmatic about how often they ask their teachers to adopt something new. Thus, products have a better chance of scaling if they are designed to be easy to use and fit with existing practices and workflows of the schools and districts. (Wu et al., 2021)

Across hub efforts, we saw that they accomplished more relative to the effort expended when they focused on building on existing data and processes and optimizing the logic surrounding data use rather than on improving the quality of the data infrastructure. This finding does not imply that improvements to data infrastructure are not sometimes important—just that their cost is so high that technical improvements should not be prioritized lightly. Because data needed to be fit to the logic of the specific improvement project and the assets and constraints of local context, coaches were the best-positioned hub staff to provide critical supports for high-quality data use.

Coaches could play a big role in helping teams to focus on their priorities and the associated measures through the questions they asked when they met with sites' improvement teams. One coach described the process of helping sites to select changes for testing:

*Say you want to improve on this, so let's define what is the change idea and what are the measurements that we're going to actually look at. I frame it as, "What is it that we need to know to know that this is an improvement? What is it that we need to know, and how are we going to collect that?"... So we frame them as question[s] ... and then after they implement [the change ideas] and we come to the study meeting, I bring those questions into the study meetings. Now we're just going to answer them. So it's just very much like, "We're going to talk about these ... questions that we've posed for ourselves." That's helped to focus those, and then record those down. ... [And] there I'm saying, "Look at the data and let's think about it."*

The basic framing questions this coach provided created opportunities for the team members to course correct and prioritize their actions if someone suggested testing an idea that might be interesting but was insufficiently deep to improve student outcomes measurably or was not central to the team’s goals. It also created a chance for the team to discuss their data collection plan and find data that were aligned enough to their change and easy enough to collect while still being sufficient for the logic of their project.

For the work to gain traction, coaches needed to make continuous improvement tools and relevant data accessible to school personnel to help teams focus on the straightforward logic of continuous improvement while finding appropriate evidence. Coaches often adapted and selected the most useful and relevant pieces of existing improvement frameworks to meet teams’ needs and avoid inundating them with overly technical information. Similarly, coaches in sites with more traction prioritized collecting and sharing only useful and relevant data to the teams they supported, such as providing customized snapshots of outcomes for a target student group. In some instances where coaches did not buffer or scaffold data analysis, participants shared that they felt “overwhelmed” or did not have time to digest all the available information. At sites where continuous improvement gained traction, team members described coaches partnering with them to make meaning—for example, disaggregating data to answer site-specific questions or supporting them in discussing everything from broad-scale change strategy to testing the introduction of a new assessment tool in a classroom.

At all participating sites, teams had to identify *practical measures*—data that staff at a site could easily collect and analyze—to guide their improvement efforts (Yeager et al. 2013). Because educators have generally been trained to think of only the subset of data needed for accountability, coaches played a critical role by reinforcing the value of practical measures, helping teams better use existing data assets (such as school information systems), and supporting team members to select logical measures for their tests. (Our data also provided examples of times when coaches did not support selection of relevant, practical measures and teams ran tests that did not support the inferences they needed to move their work forward.)

One type of measure that many sites found eye-opening for assessing the impact of their work was the *empathy interview*, which provides qualitative data that can be used at various points in an improvement process to understand the experiences of people—including both students and adults—involved in the system the project is trying to change (see the box).

## Practical Measures in Continuous Improvement: Empathy Interviews

Empathy interviews are frequently used to understand students' experiences of the system believed to be causing the gap between existing and desired performance, but they could also play an important role in providing insights into implementation, as the following example shows.

Sites' experiences showed that collecting basic implementation data was sometimes crucial for learning, although sometimes educators and even coaches in one of the hubs thought it would be overly burdensome. For example, one site designed a mentorship program for 66 students who received an F in mathematics at the end of their first semester of ninth grade to try to help them get a C or better by the end of the second semester. Their idea was to assign these students a junior or senior as a peer mentor. A team member explained how they made multiple small tweaks to the idea over time as they tried to increase student participation:

*We started it by having them meet during office hours in the beginning, which was outside of class period. And then we realized we weren't getting students ... attending. And so we went to doing it ... during first period. We also realized the way we were communicating with the teachers [sending all teachers a notice that the mentors/mentees program was meeting that day] ... wasn't effective enough ... [because the teachers] could have had one of the 66, or they may not have had [one]. ... [W]e changed it to just targeting those teachers [who had one of the 66 students during the target period] and sending it ... the day of. ... [But teachers still reported it was disruptive. S]o we created a Remind app thing to send to all the mentees and mentors. ... So we changed a lot of things ... [and] ... we learned a lot about communication and planning for it.*

The question of when to schedule additional support for students—after school, on weekends, during lunch, or during a class period—was a topic where several teams across both networks found that successful implementation was a central challenge. Sites were often able to make productive adaptations to how they implemented a change idea based on basic count data paired with empathy interviews—how many students attended, what students said about why they could not or would not attend some of the time slots tested, and how teachers experienced each version of the intervention—that were nonetheless sufficient to support the logic of the improvement project. The ease and simplicity of data collection are the defining features of the practical measures that are central for continuous improvement and are largely absent in traditional conversations about data.

Looking across the interviews, a clear pattern emerged about the nature of interactions that hubs created for teams around data and the continuous improvement work. As illustrated by all the quotations in this section, when sites had traction, they and their coaches described their work in jargon-free language, explaining how they steadily worked through persistent challenges to make progress. Engaged participants in sites with traction did not always know specific continuous improvement terminology or precise processes, but they had internalized the overarching concept of using data to diagnose the system, making a plan, getting rapid feedback to adjust the plan, and tracking progress towards a clear goal. Hubs provided support to improvement teams by systematically walking through some of the data available to the team and providing supports that made complex problems manageable through a combination of a straightforward logical process and aligned data.

**Hubs can improve the depth of work by supporting districts to select promising ideas to test.** Continuous improvement posits that when ideas are put into new contexts, they often need to be adapted to work effectively, yet there is still a strong role for evidence from research and practice for guiding decisions. This evidence is an important indicator of the potential depth of the change because only deep changes produce a scalable impact on student outcomes. As Penuel et al. (2017) found, school and district leaders are most likely to take findings from research and move them into practice when they learn about the research in their professional networks. Both hubs stepped into this role by creating what they called a “change package” with research-based suggestions about changes likely to improve on-track rates. The data in this study suggest that change packages can be a strategic way to increase the chances that ideas are sufficiently deep to matter for student outcomes. When hubs identified deep change ideas for sites to try, they also decreased the chances that teams spent a lot of time trying to reinvent the wheel.

Hubs are well positioned to inform sites about evidence-based change ideas to consider, but the ways they designed and shared them with sites influenced the traction those change packages received. Each hub’s change package first identified a few important areas to focus on, but the nature of the change packages differed across topics and hubs, and not all change packages worked equally well. When a change package provided an overview of major findings in a research domain or a large number of potential changes to try with minimal guidance about how to select an idea to test, the process for selecting a change idea was haphazard. As one coach explained, when teams had little guidance to support selecting a research-based change idea, “the change ideas come from, ‘I kind of have heard about this,’ or almost like the first person who talks frames what the change ideas are.” Hubs were more successful at getting sites to integrate deeper ideas from research when they paired an overview of big ideas with a relatively small number of clearly explained practices for sites to try. When hubs boiled ideas down for busy site leaders, it facilitated action. One leader described it this way:



*The [hub] kind of presented to us a menu of options. One of them included using these sort-of contracts or these behavioral or academic plans to support students. Another one involved, I think like an intentional professional development pathway for teachers. ... [T]he team thought that these personalized contracts would be the way to go.*

When hubs distilled research in a way that sites could understand the principles identified through research and picture what it could look like to operationalize those principles, hubs were able to funnel change idea selection towards ideas with stronger evidence behind them.

A well-designed change package presented by experts—either researchers or practitioners—elevated the importance of the work. Experts could provide tailored advice on implementation directly to sites, which increased the frequency with which sites in both networks used the change package to select research-based ideas to test. Experts helped build motivation around the need for change and helped sites see how to operationalize changes, giving sites momentum behind the idea and a picture of what it looked like to implement it—both of which supported deeper work. One coach shared the impact of an expert thought partner on a site’s change idea:

*The change idea ... came after a conversation at [a hub convening with an expert] who reiterated the importance of quantity of reading. ... [The team] used that conversation in conjunction with [a digital learning] platform and started with a pilot, the teacher group, that said, ‘Get your kids reading this much time.’ The pilot teachers did that, and those kids in that classroom, their lexile levels grew.*

Having experts interact directly with participants had a bonus of increasing the credibility of new ideas with the improvement team and their broader school and district communities. One participating teacher explained that she was able to use an expert’s knowledge to gain more authority for the ideas than she would have otherwise: “It’s not just me saying it. I have somebody behind me saying, ‘You know what? She’s right. She’s right. That’s what we’re looking for, and I think that’s what’s missing.’” A strong hub strategy paired information about evidence-based ideas that balanced motivation, principles of the reform, and a description of how to operationalize it with access to experts and integrated coaching support.



---

## Conclusion and Implications

This report examines the work of two on-track improvement networks through the lens of Coburn’s four elements of scale. Work observed during the 2020–21 year touched mostly on the first three—spread, depth, and ownership—to varying degrees, with a smaller number of sites taking steps that increased their chances of sustaining the work. As the initiatives progress, drawing closer to the official end of collaboration for both hubs in spring 2023, questions regarding ownership and sustainability will come more to the forefront.

The early stages of the networks required collaboration between the hubs and sites to start the work and find leverage points in participating education systems to drive change. Both hubs asked sites to form an improvement team and attend convenings, requiring dedication of time and resources. About half of the sites built little traction for the work during 2020–21. Others that launched did not pause and determine the specific changes that were most effective for their particular needs (so the work was shallow). Some teams failed to move from launch to a steady rhythm for the work. The improvement work was seen as “work for the hub” or an isolated initiative (so there was little ownership). Under these circumstances, there was nothing impactful to sustain.

In other sites, however, improvement work was becoming more integrated and coherent with the rest of the organization’s work. In these sites, team members selected promising change ideas, and their testing gave them the information they needed to adopt or adapt the change so that it could be implemented in their system in ways that positively affected students and shifted system practices (depth), or the tests helped them abandon an idea that did not work for them. Through the process of testing and making a decision based on the data they gathered, teams came to own the change (ownership), and positive results helped them build broader support for and use of the idea, integrating it into existing processes and norms in the school (sustainability). The results obtained in each site were due to an interaction between the conditions at the site and the strategies that hubs enacted via the events they offered, the tools they provided, and the coaches who worked with sites’ improvement teams. This report highlights hub strategies that were more or less likely to build traction for participating teams in their networks and support them in moving reform practices towards scale.

From the experiences in these networks, we have identified recommendations to inform the work of hub organizations like philanthropic funders, external support providers (such as County Offices of Education or nonprofits leading networks), and technical assistance providers. Additionally, we believe these lessons apply to districts working as a hub to support a set of their schools to make progress in a given area:

- **Prompt participant ownership of the work early and often.** Ultimately, the participants, not the hub, will need to lead improvement work on their own if it is to be sustainable. Hubs must prioritize discussions about a gradual release of ownership of the work early and often, including conversations on how to scale the work throughout the system from the beginning. When hubs are intentional about designing work so that it revolves around local priorities, sites can weave the new ideas into their existing systems, increasing the likelihood that the work will gain traction.
- **Approach continuous improvement principles as practical and common sense.** While complicated technical tools can facilitate refined approaches to continuous improvement, when they are introduced before the work has traction, they can overwhelm educators. For districts, schools, and practitioners, continuous improvement is about positively changing their systems to improve practices and student outcomes. Hubs can facilitate achieving the desired outcomes by helping sites ask the right questions and by portraying continuous improvement in ways that are conceptually rigorous yet technically simple.
- **Flexibly integrate tools, structures, frameworks, and data so that teams can more easily digest and ultimately sustain them.** Related to the previous point, there are many continuous improvement resources and tools that teams can use, yet many teams do not understand how and when to use which tools to address their needs. Improvement coaches who invest time and energy in building an understanding of ongoing work and context in each LEA are positioned to pair teams with the resources that are important for a given time and task. When hubs prioritize helping a team to improve a system, as opposed to prioritizing teams' adherence to a continuous improvement methodology, they provide an authentic opportunity for sites both to improve and to learn improvement processes. Relatedly, using locally available data sources and teaching approaches (such as empathy interviews) that provide practical measures for improvement decreases sites' reliance on complicated data infrastructure that they may not have the resources to support after the network sunsets.
- **Value the existing expertise, assets, and constraints of participating LEAs and take on a stance as a partner in the work.** Fundamental to continuous improvement is the understanding that local context plays a critical role in figuring out how an approach can be implemented effectively. Many external support providers come into schools, present great ideas, give evidence that they have worked somewhere else, and explain to educators what they need to do differently. Such attempts generally fail to yield sustainable improvement because they do not consider the local assets that can be leveraged to make a new practice feasible to implement well and what challenges might need to be addressed to attain high quality consistently.

Hubs can provide a unique vantage point compared to all participants in the improvement team in that they do not operate within the system that the team is trying to change. Hubs that take time to build relationships with teams and deeply understand system context are better positioned to help teams test changes and gain traction in their work. Hubs and coaches that take a partnering approach to building shared understanding, rather than a top-down approach, more easily build trust with teams and help teams feel a sense of ownership over the work. Once trust is established, coaches can ask questions to support strategic thinking and nudge team members to make sure tasks get completed on a reasonable timeline.

- **Center improvement work on teacher knowledge.** A key component of continuous improvement is the importance of elevating practitioner knowledge. Nonadministrators (for example, teachers, counselors, or teachers on special assignment) made up the majority of all improvement teams across the networks we examined. But the power dynamics within teams varied, with administrators making most of the decisions at some sites, not engaging at some, and intentionally distributing leadership to teachers and other “ground-floor” practitioners in others. Teams that empowered staff whose daily roles were to serve students and put *their* experiential knowledge at the center of the improvement work saw more success. People in these roles had the richest opportunities for understanding student experiences as well as the deepest knowledge of the supports for and barriers to implementing new approaches, and they were most likely to be the ones testing any changes. When administrators practiced distributed leadership, they found that teacher leaders (as well as counselor leaders and so forth) were also in the best positions to convince reluctant colleagues to try practices that had proven successful in their local context.
- **Focus on work that meets the 80/20 rule (also known as the Pareto principle).** The 80/20 rule posits that 80 percent of the outcomes come from 20 percent of the causes; the implication is that hub organizations should help participants prioritize the few highest leverage actions. In this work, we saw traction develop when educators saw that their work had an impact. The greatest impact comes from finding approaches that an organization can spread to everyone who needs that approach. This typically means avoiding the tendency to chase after fancy approaches that are costly or can be accomplished only by a few superstars. Over time, those approaches are rarely sustained. Instead, persistent efforts to get consistent quality in a few important practices will yield the greatest educator engagement and the most sustained improvement.

- Transform student outcomes by testing changes that improve instruction.** System improvement does not inherently require changing instructional practices. For example, Fresno Unified used continuous improvement to diversify the California State University and University of California campuses that students applied to by changing how the district and its counselors informed eligible students and their families about college options (Aguilar et al., 2017). However, the core of schools is instruction, so there is a limit to the extent to which outcomes can be improved without tackling instructional change. Transforming instruction takes time, so it is not reasonable to expect any site to overhaul instruction during the lifetime of a typical improvement network. However, hubs do have the opportunity to help sites build support for instructional reform, start testing approaches, and gather the evidence needed to demonstrate that the new practices are promising in a specific context. By experiencing the first steps of improvement with the support of a network, sites can also build the muscle for doing continuous improvement on their own.

We examined the work of two improvement networks during 2020–21. We found that even at a time when schools and districts were heavily affected by the COVID-19 pandemic, the hubs were able to support their partner sites in gaining traction with work that could ultimately improve student outcomes at scale. Work gained traction when hubs built site leaders’ (including teacher leaders’) ownership in work that was substantively deep. This required the hubs to take a partnering stance—building relationships, valuing site assets, and centering local context—and facilitate strategic thinking about how to spread and ultimately sustain work that had initial positive effects on student outcomes. Continuous improvement holds great potential as an approach for hubs working with schools and districts as long as hubs meet sites where they are and avoid overly technical solutions unless they prove to be necessary for given circumstances.

## References

- Aguilar, J., Nayfack, M., & Bush-Mecenas, S. (2017, June). *Exploring improvement science in education: Promoting college access in Fresno Unified School District* [Practice brief]. Policy Analysis for California Education. [edpolicyinca.org/publications/exploring-improvement-science-education](https://edpolicyinca.org/publications/exploring-improvement-science-education)
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America’s schools can get better at getting better*. Harvard Education Press.
- California Department of Education. (2020). *FAQs on grading and graduation requirements*. [cde.ca.gov/ls/he/hn/gradeeducationfaq.asp](https://cde.ca.gov/ls/he/hn/gradeeducationfaq.asp)
- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational researcher*, 32(6), 3–12.
- Cottingham, B. W., O’Meara, K., & Gallagher, H. A. (2022). *Azusa/Dinuba case study* [Case study in preparation]. Policy Analysis for California Education.
- Fensterwald, J. (2021, June 24). *Grading changes, other Covid accommodations await Gov. Newsom’s signature: Remediation, pass/no pass options would help hold students harmless*. Ed Source. [edsources.org/2021/grading-changes-other-covid-accommodations-await-gov-newsoms-signature/656782](https://edsources.org/2021/grading-changes-other-covid-accommodations-await-gov-newsoms-signature/656782)
- Grunow, A., Hough, H., Park, S., Willis, J., & Krausen, K. (2018, September). *Towards a common vision for continuous improvement for California* [Report]. Policy Analysis for California Education. [edpolicyinca.org/publications/towards-common-vision-continuous-improvement-california](https://edpolicyinca.org/publications/towards-common-vision-continuous-improvement-california)

- Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher*, 33(8), 16–30.
- Hough, H., Willis, J., Grunow, A., Krausen, K., Kwon, S., Mulfinger, L. S., & Park, S. (2017, November). *Continuous improvement in practice* [Report]. Policy Analysis for California Education. [edpolicyinca.org/publications/continuous-improvement-practice](http://edpolicyinca.org/publications/continuous-improvement-practice)
- Institute for Healthcare Improvement. (2003). *The Breakthrough Series: IHI's collaborative model for achieving breakthrough improvement* [White paper]. [ihi.org/resources/Pages/IHIWhitePapers/TheBreakthroughSeriesIHICollaborativeModelforAchievingBreakthroughImprovement.aspx](http://ihi.org/resources/Pages/IHIWhitePapers/TheBreakthroughSeriesIHICollaborativeModelforAchievingBreakthroughImprovement.aspx)
- Langley, G. L., Moen, R., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance* (2nd ed.). Jossey-Bass Publishers.
- LeMahieu, P. G., Bryk, A. S., Grunow, A., & Gomez, L. M. (2017). Working to improve: Seven approaches to improvement science in education. *Quality Assurance in Education*, 25(1), 2–4. doi.org/10.1108/QAE-12-2016-0086
- Park, V., Kennedy, K., Gallagher, H. A., Cottingham, B., & Gong, A. (2022). *Weaving and stacking: How school districts craft coherence towards continuous improvement* [Report in preparation].
- Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., ... & Allen, A. R. (2017). How school and district leaders access, perceive, and use research. *AERA Open*, 3(2), 2332858417705370.
- Pier, L., Hough, H. J., Christian, M., Bookman, N., Wilkenfeld, B., & Miller, R. (2021, January 25). *COVID-19 and the educational equity crisis: Evidence on learning loss from the CORE Data Collaborative*. Policy Analysis for California Education. [edpolicyinca.org/newsroom/covid-19-and-educational-equity-crisis](http://edpolicyinca.org/newsroom/covid-19-and-educational-equity-crisis)
- Tadayon, A. (2020, November 3). *California school districts revert to A–F grades this fall—with more flexibility for some students: The decision appears to be prompted by California university systems' admissions policies, but some say letter grades are more motivating*. Ed Source. [edsources.org/2020/california-school-districts-revert-back-to-a-f-grades-this-year-with-flexibility/641962](http://edsources.org/2020/california-school-districts-revert-back-to-a-f-grades-this-year-with-flexibility/641962)
- Willis, D. J., & Fensterwald, J. (2021, May 5). *Over half of California public school students remain in distance learning: Students in wealthier schools 3 times likelier to be back in school full time*. Ed Source. [edsources.org/2021/new-data-55-of-california-public-school-students-remain-in-distance-learning/653848](http://edsources.org/2021/new-data-55-of-california-public-school-students-remain-in-distance-learning/653848)
- Wu, J., Wingard, A., Golan, S., & Kothari, M. (2021). *From research to market: Development of a transition process to integrate sustainable scaling methodologies into education innovation research design and development*. SRI International.
- Yeager, D., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Practical measurement*. Carnegie Foundation for the Advancement of Teaching.

## Author Biographies

**H. Alix Gallagher** is the Director of Strategic Partnerships at Policy Analysis for California Education (PACE) at Stanford University, where she leads the research-practice partnerships with California Education Partners and the CORE Districts. She holds a Ph.D. in educational administration from the University of Wisconsin-Madison. Alix began her career in education as an elementary school teacher.

**Benjamin W. Cottingham** is the Associate Director of Strategic Partnerships at Policy Analysis for California Education (PACE) at Stanford University. He focuses on the development of organizational conditions that support the continual improvement of student outcomes across California's education system. He earned his master's degree in education from Stanford University. He previously worked in high-needs schools as a nonprofit director and high school teacher.

**Kiley O'Meara** is a senior researcher with PACE, partnering with schools, districts, nonprofits, policymakers, and funders dedicated to improving K–12 education. She has worked in and with foundations and nonprofits, and has also taught high school English. Kiley received her master's degree in public policy from Harvard's Kennedy School.

# Policy Analysis for California Education (PACE)

*Improving education policy and practice and advancing equity through evidence*

PACE is an independent, non-partisan research center led by faculty directors at Stanford University, the University of Southern California, the University of California Davis, the University of California Los Angeles, and the University of California Berkeley. Founded in 1983, PACE bridges the gap between research, policy, and practice, working with scholars from California's leading universities and with state and local decision makers to achieve improvement in performance and more equitable outcomes at all levels of California's education system, from early childhood to postsecondary education and training. We do this through:

- 1 bringing evidence to bear on the most critical issues facing our state;
- 2 making research evidence accessible; and
- 3 leveraging partnership and collaboration to drive system improvement.



Stanford Graduate School of Education  
520 Galvez Mall, Suite 444  
Stanford, CA 94305  
Phone: (650) 724-2832 • Fax: (650) 723-9931

[edpolicyinca.org](http://edpolicyinca.org)