

Evaluation of the Breakthrough Success Community

2019–20 Through 2022–23

H. Alix Gallagher
Benjamin W. Cottingham
Leah Faw



February 2025

Evaluation of the Breakthrough Success Community

2019–20 Through 2022–23

H. Alix Gallagher, Benjamin W. Cottingham, and Leah Faw
Policy Analysis for California Education (PACE)

Acknowledgements

The authors would like to thank the CORE Districts for welcoming us to their public events, inviting us to internal planning sessions, and granting us access to their internal documents. We are also grateful for the time and insights that CORE staff and BTSC school team members provided during interviews. We would like to thank Belen Chavez for her technical support of the quantitative analysis and Shawn Bernardo for his steady support of our data collection. 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's contents.

Suggested Citation

Gallagher, H. A., Cottingham, B. W., & Faw, L. (2025, February). *Evaluation of the Breakthrough Success Community: 2019–20 through 2022–23* [Report]. Policy Analysis for California Education. edpolicyinca.org/publications/evaluation-breakthrough-success-community

Contents

Executive Summary 3

Introduction 4

 Main Influences on CORE’s Approach to BTSC 5

 The COVID-19 Pandemic as Context for Early Years of CORE’s BTSC 9

Research Approach. 11

 Research Questions 11

 Data and Methodology. 12

 Analytic Approaches 13

 Frameworks. 14

CORE’s Breakthrough Success Community 16

 Outcome Measures and the BTSC Aim. 16

 Recruitment. 25

 Content of the Change Package 28

 How Do We Know if a Change Is an Improvement?. 45

School Engagement With BTSC 49

 CORE Struggled to Get Traction. 51

 Learning About Building Traction From Positive Outliers 57

Student Outcomes 69

 Selecting a Comparison Group and the Resulting Analytic Sample. 70

 Descriptive Statistics on BTSC and Comparison Schools 73

 Results of the Regression Analysis. 75

Summary and Conclusions 78

References 82

Appendix A. 83

Author Biographies. 89

Figures

Figure 1. Indicators of Ninth-Grade On-Trackness in CORE Districts Before BTSC Launched 5

Figure 2. Criteria for BTSC On-Track Lanes, Checklist Method 18

Figure 3. Criteria for BTSC On-Track Lanes, Compensatory Points Method 20

Figure 4. BTSC Checklist (1.0) Versus Compensatory Points (2.0). 21

Figure 5. Traction Framework for Reforms Affecting Student Outcomes at Scale. 50

Figure 6. Sample 1-Page Overview of Key Actions for April–June 2022 54

Figure 7. D/F Rate at Magnolia, 2018–19 Through 2022–23. 64

Figure 8. Average Ninth-Grade On-Track Points at Magnolia, 2018–19 Through 2022–23 64

Figure 9. D/F Rate at Aspen, 2018–19 Through 2022–23 67

Figure 10. Average Ninth-Grade On-Track Points at Aspen, 2018–19 Through 2022–23 68

Figure 11. Average BTSC On-Track Metric Points in BTSC Schools Versus Comparison Schools. 73

Figure 12. Average School D/F Rates in BTSC Schools Versus Comparison Schools 74

Tables

Table 1. Timeline and School Participation 10

Table 2. Team Structure Proposed at BTSC Launch 32

Table 3. School-by-School Highlights of Change Ideas Being Tested Reported by CORE 52

Table 4. Year-to-Year Variation in Cohort Size for Aspen and Magnolia High Schools. 68

Table 5. Values of Matching Variables for BTSC and Comparison Sample 72

Table 6. Distribution of BTSC and Comparison Schools Across the CORE Districts for the Analytic Sample . 72

Table 7. Results of the Confirmatory Model. 76

Table A1. Codebook: Definitions of Variables Used in Report Regressions and Analysis 83

Table A2. Covariates for Regression Analyses 84

Table A3. Logistic Model Regression Results 87

Table A4. All Regression Output 88

Executive Summary

The CORE Districts received a Networks for School Improvement grant from the Gates Foundation to launch a network focused on improving ninth-grade on-track rates. CORE based its approach on the Institute for Healthcare Improvement's Breakthrough Collaborative model, where a hub identifies areas where practices diverge from research and translates those research findings into actionable recommendations and measurements to gauge success. CORE invested significantly in measures to support the network, developing the Breakthrough Success Community (BTSC) on-track metric, which improved on binary metrics by estimating five lanes of "on-trackness." However, CORE paused the metric in 2019–20 due to concerns and introduced the Developmental Relationships Survey as a measure of adult–student relationships. The revised on-track metric returned in winter 2021–22, and although educators valued the measure, there were questions about its validity and transparency.

From 2018–19 to 2021–22, CORE recruited three cohorts of high schools (41 schools total) to BTSC, with each cohort receiving 2–4 years of support during the period covered by this evaluation. COVID-19 interrupted Cohorts 1 and 2, and Cohort 3 launched after the pandemic. Recruitment saw limited engagement and did not include system analyses to help schools identify gaps for BTSC to address or see existing system assets that could serve as a foundation for improvement.

CORE's review of research led to five focus areas, or "drivers": adult teaming, relationships, grading, transitions, and master schedule. BTSC provided a Key Actions Checklist to guide testing ideas in these drivers, emphasizing one-on-one adult–student interactions and building on student assets over scaling system-level changes. Some schools made progress with BTSC, characterized by strong relational trust, engaged administrators, adaptation of BTSC ideas, and monitoring of student outcomes. However, most schools did not fully implement BTSC practices across the ninth grade.

To assess student outcomes, we compared BTSC schools in the CORE Districts to a matched set of schools. Analysis of trends in BTSC on-track metric points and ninth-grade D/F rates showed comparable patterns for BTSC and comparison schools. Regression analyses found no evidence of an impact of BTSC on on-track points in either 2021–22 or 2022–23 and no effects on the D/F rate in 2022–23. The results were consistent across a range of specifications of the regression models.

Since our data collection concluded, BTSC has completed another year, and CORE will enroll a fourth cohort. Recent adjustments to BTSC may have addressed some issues. We conclude with recommendations to improve the success of similar initiatives in education.

Introduction

In 2010, large urban districts in California put in a joint proposal for Race to the Top, forming a nonprofit called the CORE Districts (hereafter CORE).¹ The proposal originated out of their collective frustration about the No Child Left Behind (NCLB) Act's approach to measuring Adequate Yearly Progress and determining which schools were struggling versus which were effective. CORE's first application was unsuccessful, so they reapplied in 2013 and secured a waiver from some sanctions of NCLB in exchange for using an innovative, multiple-measures accountability system that CORE had developed to evaluate school performance. In 2016, with the conclusion of NCLB, CORE pivoted to focus on "continuous improvement," launching an improvement network focused on improving mathematics outcomes for African American and Hispanic students in Grades 4–8.

In June 2018, CORE won a Networks for School Improvement (NSI) grant from the Gates Foundation. The NSI grant program was designed around the belief that continuous improvement networks could support schools to serve students better—especially those students from traditionally underserved communities—so that they would thrive in high school and successfully continue to and through postsecondary education. NSI grants could focus on one of several areas related to secondary and postsecondary success, including "ninth-grade on-track." Focusing on ensuring that students have a successful ninth-grade year is supported by research (e.g., *The Make-or-Break Year* and multiple related publications from the University of Chicago Consortium on School Research) that shows that student performance at the end of ninth grade is strongly related to high school graduation and positive postsecondary outcomes. Additionally, CORE conducted a preliminary analysis of both student performance in the CORE districts and gaps in performance between African American and Hispanic students and other students in the CORE districts on two on-track metrics: the percentage of students who received no Ds or Fs in ninth grade and the percentage of students who finished ninth grade with a GPA of 2.5 or higher. Results showed that African American and Hispanic students in poverty were less likely than their peers to finish ninth grade without Ds or Fs and were less likely to have a GPA of 2.5 or higher (see Figure 1). These data confirmed that an on-track-focused network in the CORE districts could support the Gates Foundation's goal of improving student on-track rates for historically underserved student groups.

¹ The districts that are members of CORE have changed slightly over time. During the years covered in this report, CORE included Fresno, Garden Grove, Los Angeles, Long Beach, Oakland, Sacramento, San Francisco, and Santa Ana Unified School Districts.

Figure 1. Indicators of Ninth-Grade On-Trackness in CORE Districts Before BTSC Launched

	Ninth-grade on-track (percentage)	
	No Ds or Fs in ninth grade	
African American/Hispanic and in poverty	No Ds or Fs in ninth grade	34
NOT African American/Hispanic and in poverty		59
African American/Hispanic and in poverty	Ninth grade GPA of 2.5 or better	50
NOT African American/Hispanic and in poverty		83

Source. Adapted from a slide from “Improving 9th Grade ‘On-Track’ Rates in California Public Schools,” a training presentation at the Institute for Healthcare Improvement, CORE Districts, [February] 2018.

Across two grants starting in June 2018 and ending in June 2025, the Gates Foundation awarded CORE almost \$22.8 million to support the creation of an improvement network focused on ninth-grade on-track. In addition to funding the basic network, the grant included funding to develop a data and measurement system using predictive analytics to identify annually (in Grades 4–12) the extent to which students were on track for high school graduation and likely postsecondary academic success; the grant also supported ongoing research and developmental evaluation support from Policy Analysis for California Education (PACE).

Main Influences on CORE’s Approach to BTSC

CORE designed its NSI work to be “[d]isciplined by improvement science methodology with the network structure of the Breakthrough Series Collaborative model developed by the Institute for Healthcare Improvement” (CORE Districts, 2018). One reason to examine both the implementation and the outcomes of BTSC is that it was an important attempt to instantiate two related approaches to improving organizational outcomes—improvement science and a Breakthrough Series—in education.

Improvement science is a particular methodology for continuous improvement that seeks to help organizations use a disciplined approach to understanding their systems, identifying high-leverage changes they could make to improve outcomes, testing possible changes in iterative cycles, and spreading those changes that lead to improved outcomes. The Breakthrough Series Collaborative model is a vehicle for using improvement science in a network that has particular features, as the Institute for Healthcare Improvement (IHI) explains:

The driving vision behind the Breakthrough Series is this: sound science exists on the basis of which the costs and outcomes of current health care practices can be greatly improved, but much of this science lies fallow and unused in daily work.

There is a gap between what we know and what we do. The Breakthrough Series is designed to help organizations close that gap by creating a structure in which interested organizations can easily learn from each other and from recognized experts in topic areas where they want to make improvements. (IHI, 2003, p. 1)

As this quotation shows, the Breakthrough approach originated in health care and is designed for spaces where: (a) “current prevailing practice deviates from the best scientific knowledge,” (b) improvements would have a large impact on outcomes and costs, and (c) some organizations have already demonstrated the feasibility of adopting the evidence-based practices (IHI, 2003, p. 3). In short, Breakthrough Series are for research-based interventions that have been shown to be feasible to implement and, when implemented, have been shown to cause improved outcomes. The Breakthrough Series approach has made a quantifiable impact on patient outcomes and health care costs for multiple medical issues around the world (e.g., by reducing the rates of cesarean sections).

Key Features of IHI’s Breakthrough Collaborative Model

The focus of an IHI Breakthrough Collaborative is figuring out how to take a research-based consensus and put it into practice. Given that the entire approach is built on a premise that a strong research base exists, the collaborative needs a mechanism for centering that knowledge within the network. That mechanism is the *faculty*, a mix of international subject matter experts and expert clinicians who know how to apply the research base in practice. The “chair” of the faculty typically spends 1–2 days per week working in the collaborative, developing a vision for new practice based on the research base and collaborating with IHI to support teams and the network. Combined, the faculty and IHI staff are the network hub, charged with:

- identifying an appropriate aim for the collaborative,
- distilling research-based knowledge into a change package that explains shifts in processes that should lead to substantial improvement in outcomes, and
- developing measurement strategies to support teams using the Model for Improvement to conduct small-scale tests of change and support successful changes being adopted as standard practices in the organization.

The hub also invites organizations into a network, known as a *collaborative*, to learn collectively how to implement these changes. The hub offers “learning sessions” three times during the roughly 15 months of the collaborative, where a small team from each member organization comes together with teams from other member organizations to

learn from the hub, faculty, and one another about how to make the changes. The teams that participate in the learning sessions lead a larger team back in their organizations and test process changes that have strong research backing to figure out how to implement them in ways that dramatically improve outcomes.

The cornerstone of the improvement work is the Model for Improvement, where teams posit and answer three questions:

1. **What are we trying to accomplish?** This is the team's aim. In a collaborative, each team's aim would be related to an overall network aim.
2. **How will we know that a change is an improvement?** This requires the hub to identify leading indicators that provide evidence that changes are an improvement; teams are required to maintain "run charts" (i.e., a visual representation similar to a line graph that shows progress over time on key process and/or outcome measures) to track their progress.
3. **What changes can we make that will result in an improvement?** In the Breakthrough Model, the hub is responsible for providing evidence-based changes for participating organizations to test.

In a Breakthrough Model, while the hub and faculty provide guidance about shifts that teams need to make, teams are responsible for conducting testing cycles, where they select changes and iteratively test ways of implementing them to make them responsive to organizational context and culture. Based on the measured outcomes from their tests, they then adapt those changes until they see improvement—at which point, they work to spread adoption of the changes in the organization—or abandon them if they do not lead to improved outcomes. At the conclusion of the Collaborative, the learning and results are documented by the hub and participating organizations so that others can learn from their work (IHI, 2003).

CORE titled their NSI grant the "Breakthrough Success Community" (BTSC), although CORE planned some adaptations of the Breakthrough model from the outset. For example, while an IHI Breakthrough Series typically lasts from 6 to 15 months, the NSI grant funded CORE for a planning year and 4 subsequent years of work with schools. Within those years, CORE planned to retain the Breakthrough Collaborative structure of three learning sessions per year, where all schools in the network would send a team, which was then charged with engaging with testing and spreading ideas during the "action periods" between learning sessions. The methodology that CORE taught schools to use in testing ideas was improvement science, specifically the idea of a Plan-Do-Study-Act (PDSA) cycle (Bryk et al., 2015), which is the methodology typically used in a Breakthrough Series.

CORE did not engage a typical IHI faculty chair to spend 1–2 days per week for the duration of the collaborative to ensure that BTSC was infused with key findings from research and lead the development of a change package and associated measurement system (as the Breakthrough Collaborative model would suggest). It also soon became apparent that the research base in education has not produced a small handful of tightly specified changes that, when implemented, have been shown to predictably yield dramatic improvements in ninth-grade on-track rates. Ultimately, CORE decided to focus their network on five research-based ideas: (a) the transition into high school, (b) adult collaborative structures for supporting students, (c) adult–student relationships, (d) strategic master scheduling, and (e) grading reform. We describe the research base for each of these and how the work on these topic areas evolved over time in subsequent sections.

Instead of hiring a Breakthrough faculty, CORE invested in supporting schools in doing BTSC work by hiring improvement coaches to work with a small set of schools—typically four to six schools each. The coaches provided a higher level of improvement support than is typical in a Breakthrough model, ideally checking in with each of their teams at least monthly and sometimes regularly attending team meetings in schools (in person or virtually). In some cases, the coaches reached out to experts in the field for support in developing components of the change package, but CORE never got expert support to create a measurement system related to all elements of the change package.

CORE Improvement Coaches and the BTSC On-Track Metric

CORE supported schools to apply the principles of improvement science in BTSC by hiring several improvement coaches. All coaches were former teachers or principals, most of whom had formal training in improvement science (e.g., attending one or more workshops sponsored by the Carnegie Foundation). In fact, two coaches previously worked for the Carnegie Foundation, and another had played a key role in leading improvement in a large urban district, serving as the lead author of a handbook about how to do improvement science in education. Additionally, several senior members of CORE’s team, who served in leadership roles in BTSC, had deep training in improvement science from the Carnegie Foundation as well as training in IHI’s Breakthrough Collaborative model. These CORE leaders also had prior experience leading one or more improvement networks. Over time, through strategic hiring decisions as BTSC expanded, CORE’s set of improvement coaches had an increasing depth of knowledge about improvement science in education.

Some BTSC coaching was provided by people outside CORE as part of CORE’s exploration of a staffing model that might support long-term sustainability. In Cohort 2, BTSC worked with five schools that were not in any of the CORE Districts. These schools were coached by improvement coaches from their local County Offices of Education (COEs). In Cohort 3, a new CORE district added a set of four schools and gave one of their existing teachers on special assignment (TOSAs) release time to work with CORE to learn how to coach schools in BTSC as a strategy for sustaining the work beyond the grant.

In addition to identifying drivers, CORE developed a BTSC “on-track metric.” This measure was of crucial importance for BTSC because schools’ aims and the overall network aim were going to be based on students’ on-trackness at the end of ninth grade. By creating a more precise way of monitoring the progress for each individual student towards attaining various degrees of on-trackness, CORE believed schools could most effectively target the appropriate students with additional resources. Developing a metric like this is technically complicated and requires trade-offs that we will explain in greater detail in subsequent sections.

The COVID-19 Pandemic as Context for Early Years of CORE’s BTSC

During winter 2018–19, CORE recruited its first cohort of schools for BTSC. The first cohort, composed of 14 schools from five of the CORE districts, launched in April 2019. In early April 2020, Cohort 2 launched with 16 schools (12 from the same five CORE districts and four schools from three other districts that were not part of the CORE districts). CORE launched a third cohort of 11 schools (from three of the CORE districts, including one district new to BTSC) in fall and winter 2021. The timeline in Table 1 shows major milestones in BTSC, the number of schools participating over time, and the districts they came from. It is worth noting that the seven schools that dropped were not equally distributed across cohorts or districts: six were from Cohort 1, and six were from the same district.

Table 1. Timeline and School Participation

	2018–19	2019–20	2020–21	2021–22	2022–23
Cohort 1	April launch (14 schools)	Year 1 (14 schools)	Year 2 (10 schools)	Year 3 (9 schools)	Year 4 (8 schools)
Cohort 2		April launch (16 schools)	Year 1 (15 schools)*	Year 2 (15 schools)	Year 3 (15 schools)
Cohort 3				Fall/winter launch; Year 1 (11 schools)	Year 2 (11 schools)
COVID-19		In-person school closes March 13, 2020	School online/ hybrid	In-person school resumes	
Focal drivers	Transitions Adult teaming	Transitions Adult teaming Relationships	All five	All five	All five

* One of these 15 schools did not participate in 2020–21 but rejoined in 2021–22, so we do not count them as attriting.

The timeline makes it clear that BTSC’s early years coincided with an interruption of enormous magnitude: the COVID-19 pandemic. When the pandemic shut schools in March 2020, the California Department of Education required that schools hold students “harmless” in terms of their final grades, so that variation in home environment, access to technology, and/or illness and its repercussions did not unfairly disadvantage some students. Districts interpreted this requirement differently, but the effect in participating districts (and the state of California as a whole) was a radical reduction in the number of students who received Ds and Fs at the end of the spring semester of 2020. The state also paused the administration of Smarter Balanced Assessment Consortium (SBAC) assessments. These policy decisions altered the two greatest pieces of information about how students were progressing through secondary school—namely, their grades and their standardized test scores, which had a substantial impact on the BTSC on-track metric among other things.

In August 2020–21, students arrived in virtual school generally having had little schooling for the 5 preceding months. Educators in BTSC schools designed virtual high school orientation for ninth graders and teachers came up with new strategies to build relationships and teach materials to students who were typically not required (in these districts) to turn cameras on during class over Zoom. The number of students receiving Ds and Fs increased dramatically in some districts (Johnson, 2021).

CORE also needed to reinvent its approaches to supporting the network of schools and engaging educators whose personal and professional lives were turned upside down. Before COVID-19, CORE held three in-person BTSC learning sessions each year, which prioritized community building, fun, and celebration along with learning. They were held off site in hotel ballrooms that could host multiple team members from all participating schools at the same time, and they offered a mix of information for teams and celebration of the work that teams were doing. But when COVID-19 hit, like the educators CORE was supporting, CORE needed to adapt its supports for educator learning to Zoom and figure out what was reasonable to ask of educators given their daily personal and professional realities. The result was that during the 2020–21 school year, CORE held three sets of 2-hour Zoom meetings—in the fall and spring these were for the entire network, but in the winter they were focused on specific drivers—instead of their typical 2-day events. While CORE, like everyone else, had a steep learning curve around effectively leading virtual meetings, CORE was ultimately able to leverage some of the advantages of virtual work, and while educators missed the community-building elements of in-person learning sessions, they remained engaged in the network. This report does not focus on CORE’s adaptations for supporting schools virtually because (perhaps optimistically) we hope that those lessons have little relevance for the near future. However, it is worth noting that CORE actively and ultimately successfully redesigned some of its delivery approach midstream to respond to COVID-19. CORE’s successful adaptations notwithstanding, the pandemic—and the toll it took on students, educators, and society as well as policy responses in education—is in the background of our description of CORE’s work on BTSC and educators’ attempts to implement ideas in their schools from March 2020 onward.

Research Approach

The data and research methods used in this study were heavily influenced by our role as CORE’s developmental evaluation partner. We worked to balance our ability to provide CORE with rapid turnaround feedback with our ability to learn independently from CORE’s experiences designing and leading BTSC and to share that learning with the field at large.

Research Questions

Our goal in writing this report is to share what we, as CORE’s research partner, have learned about the following research questions:

1. How did CORE design and implement BTSC? How did BTSC evolve as a result of CORE’s learning?
2. How did participating schools implement BTSC? What enabling conditions supported strong implementation?

3. How did BTSC affect student outcomes?
4. What are the key lessons from BTSC that could inform ongoing development of BTSC and other efforts to improve student outcomes?

Data and Methodology

Our data collection approach evolved with the network as it grew and CORE's and our questions about the network shifted:

- During 2018–19, most of the year was spent in planning. CORE launched the first cohort of schools in April. That year, we interviewed CORE staff and leaders from other partner organizations, and we observed an expert convening and other planning meetings as well as the launch of Cohort 1 with BTSC schools.
- During 2019–20 (prepandemic), we observed all learning sessions in person, and we attempted to conduct in-person site visits to all participating Cohort 1 schools (though some of the spring visits were canceled due to COVID-19).
- During the 2020–21 school year, our data collection moved online with everything else, but we still observed all events and attempted interviews in all participating schools. PACE staff also attended most internal BTSC coaching team meetings (weekly 2-hour meetings where CORE coaches and leaders planned major events, discussed progress in the network, and fine-tuned their strategies based on what they saw schools doing).
- Starting in 2021–22, there were many more schools in the network, and there were three sets of learning sessions: a joint learning session strand for Cohorts 1 and 2, a learning session strand for Cohort 3 schools in one district, a third strand of learning sessions for the Cohort 3 schools in two other districts, and virtual sessions for team leads and principals. We began to focus our data collection more strategically. We worked with CORE to identify the 12 schools they believed would be most informative (i.e., typically those schools that were making the most effort to engage in the work) while striving to include schools supported by all CORE coaches across the year. In those 12 schools, we interviewed the CORE coach and up to four team members. In all other schools, we attempted to interview the team lead and the CORE coach.
- In the final year of data collection (2022–23), we focused our data collection on “positive outliers” identified by each CORE coach (i.e., those schools most engaged, where we could study the most successful BTSC work). In these schools, we attempted to interview four team members at the most engaged school each coach was working with—ideally, the team leads, driver champions, and a knowledgeable administrator. We also observed 4 days of learning sessions, including all schools aside from Cohort 3 schools in two districts. To ensure that our findings were not biased by our smaller, more purposeful sample, we corroborated our high-level findings with CORE's internal tracking of schools' implementation of change ideas.

Across the years, we conducted 329 interviews:

- 82 interviews with CORE staff (i.e., CORE’s executive team and CORE coaches that supported BTSC schools);
- 12 interviews with BTSC coaches who supported BTSC schools as part of their roles in COEs or one of the CORE school districts that provided its own coach for Cohort 3;
- 7 interviews with individuals working at partner organizations (e.g., Education Analytics, IHI); and
- 228 interviews with educators (e.g., teachers, principals, counselors) at schools participating in BTSC.

We observed approximately 200 hours of BTSC events across the years. Our sample included the full range of events: learning sessions for all cohorts of BTSC schools, meetings specifically for team leads and principals, and a range of in-person and virtual events each year. Additionally, we were participant observers at a BTSC expert convening during the planning year; BTSC’s improvement reviews, where BTSC leaders presented information about the program and questions they wanted feedback on to external thought partners; CORE’s board meeting three or four times a year, where BTSC leadership presented information about the program and its outcomes to their board; and a set of internal BTSC team meetings.

Finally, CORE gave us unrestricted access to their internal files, which included background and planning materials, annotated facilitators’ agendas for events, and copies of internal network tracking documents. We do not make any direct claims based on CORE’s internal tracking because they are designed for CORE’s internal use, not as a research tool (and, as a result, we cannot verify interrater reliability, for example). However, for 2022–23, when we had a smaller and purposive sample of schools where we collected data, we did check to ensure that our findings were not contradicted by data that CORE collected on schools where we collected limited data.

Analytic Approaches

Because of our role as a developmental evaluation and research partner, we presented formative feedback to CORE several times per year. This formative reporting included emerging insights for the overall research questions listed previously in this report as well as evidence on how some specific changes that CORE was making to the program were working. We presented our thinking to CORE informally (through conversations in regular check-ins with key leaders) as well as formally (in written memos presented up to four times each year).

Our analytic approaches varied over time, mostly in response to work we did collaboratively with CORE and how we were sharing our learnings. Initially, we coded interview transcripts based on emergent themes. In 2021–22, we used a rubric we jointly developed with CORE, aligned

to their and our learning/research questions, and chunked the data as evidence under each domain. In 2022–23, we used a process of developing structured analytic memos, which we then shared with CORE.

This summative analysis is based first on an analysis of the memos written for CORE at regular intervals throughout the grant. These memos had embedded quotations from transcripts, sometimes supplemented by notes from observations, to support analytic claims. Given the relatively steady production of these memos over time—roughly two each semester—they provide a good record of the questions that CORE and PACE found most pressing as well as the evidence that seemed most salient at the time for answering those questions. Analyzed retrospectively, these memos helped us track the evolution of CORE’s approach as a network hub.

Our findings are also informed by the many meetings and debriefings we engaged in with CORE across the duration of the grant. As participants in these meetings, we did not always take notes (since we attended as participating partners, not researchers), so we triangulated the understandings developed in this manner by returning to the original interview transcripts or observation notes. Finally, in select cases, we returned to interview transcripts to verify evidence and gain context around claims.

Frameworks

As we addressed these research questions and analyzed our data, three lenses were prominent in our analysis. One of the underlying frames for our analysis is the Breakthrough Collaborative model itself. While CORE’s work on BTSC is important in its own right, it is also important as a relatively early attempt to bring an improvement methodology that has a demonstrated track record of success in medicine into education. However, education and medicine differ in many ways, some of which necessarily require adaptations of the model. We analyze CORE’s approach in reference to the Breakthrough model to understand the extent to which the fundamental features of a Breakthrough Collaborative were instantiated in BTSC so that we can assess the extent to which BTSC is a test of the model or deviated far enough from it that it provides little information about the possible efficacy of the Breakthrough approach for education.

The second frame we brought was an understanding of continuous improvement more broadly. As a term, *continuous improvement* includes a range of methodologies (e.g., improvement science, Baldrige, Shingo Model, etc.) that share a set of common beliefs and assumptions:

1. *Systems* (not individuals) produce outcomes.
2. As a result, change efforts should focus on key *processes*.
3. Improving processes requires continual *learning*.
4. *Frontline* workers are uniquely situated to learn how to improve outcomes.
5. Effective practices must be *spread* in organizations to change the system and improve outcomes (Grunow & Park, 2019).

As with the Breakthrough Model, we analyze CORE's enactment of BTSC to see the extent to which BTSC tests the efficacy of improvement science networks in education.

Finally, we turned to the expanded definition of *scale*, as defined by the research by Coburn (2003) on literacy reforms that improved student outcomes over the long term. Coburn pushed back on the simplistic understanding of scale—that is, how many people have been touched by an initiative. Instead, she named four dimensions of scale necessary for reform to have a long-lasting impact.

- **The depth:** how meaningful the reform is, including ultimately changing culture and beliefs.
- **Local ownership:** whether people in the school assume responsibility for the reform's success.
- **The spread:** making sure enough people implement the new approach that it can reach its target population of students.
- **Sustainability:** factors that enable the reform to continue after engagement with external support or a grant ends.

There is a natural fit between central principles of continuous improvement and Coburn's understanding of scale in that both are focused on how new ideas can be introduced and taken up in organizations in ways that lead to long-term improvements.

Compatibility of Coburn's Definition of Scale and Continuous Improvement

While Coburn did not study continuous improvement, continuous improvement is fully compatible with Coburn's definition, perhaps because they both examine purposeful attempts at change that lead to improved outcomes. Continuous improvement focuses on *systems*: the structures and processes that determine outcomes. When working to implement a reform, systems offer a path to *spread* and *sustainability* by creating opportunities to hardwire reform into an organization and institutionalize it over time. Additionally, continuous improvement's focus on the expertise of ground-level workers provides opportunities to build *ownership* of the reform. Continuous improvement centers measurement as a way of repeatedly asking and answering the question of whether a change is an improvement. By doing so, continuous improvement regularly assesses the *depth* of a reform. If a reform is being implemented superficially, regular measurement will show that implementation as ineffective, providing the chance to adapt or deepen implementation or to abandon the reform entirely. On the other hand, if a reform is

institutionalized, it becomes “the way we do things here,” reshaping organizational norms and culture and achieving deep implementation through an organization’s systems. In prior research at PACE (Gallagher et al., 2022), we have extended Coburn’s theory in the context of improvement networks, noting that when continuous improvement is used with integrity to its central principles, the result should be traction towards scaling the reform in a system.

Throughout this report, the basic principles of continuous improvement, Coburn’s multifaceted definition of scale, and our understanding of how these lenses intersect form the framework that shaped our thinking and research.

The structure of the rest of the report traces the intended causal chain from BTSC to improved student outcomes.

- **CORE’s design and implementation of BTSC:** This section focuses on CORE’s work relative to the central principles of continuous improvement and the particular Breakthrough Collaborative model they were trying to use in education.
- **Schools’ engagement with BTSC, which is the intermediary between CORE’s design and student outcomes:** This section examines what schools implemented and the extent to which schools got traction towards scale with BTSC work, with the assumption that long-term positive student outcomes from BTSC require schools to have traction towards scale. This section includes more detailed descriptions of three positive outlier schools, where traction exceeded that in most other schools, to help build an understanding of conditions that fostered traction in BTSC.
- **Student outcomes:** In this section, we analyze trends in student outcomes and then report on regression analyses to see how BTSC schools compared to other similar schools in terms of ninth-grade outcomes.

CORE’s Breakthrough Success Community

In this section of the report, we go through four key aspects of CORE’s approach to BTSC and how schools experienced it: outcome measures and the network aim, recruitment of schools, the content of the change package, and the approach to continuous improvement.

Outcome Measures and the BTSC Aim

In both a Breakthrough Series and improvement science, improvement work is focused by setting an aim. Aims serve multiple functions, including defining the goal of the Breakthrough

Series, connecting the changes tested directly to their impact on the desired outcome, providing corrective feedback if the changes are not resulting in progress towards the aim, and motivating a team as they make incremental progress towards the goal. Additionally, in a network, the hub can use teams' progress towards the aim to identify which organizations are making the most progress and can then use those as exemplars that the network can learn from. Aims are typically "SMART" (specific, measurable, attainable, realistic, time-bound). They draw a line in the sand, focusing efforts and providing feedback on whether progress is being made at an expected rate.

The other types of measurement that are fundamental to continuous improvement are the measures that provide evidence that the small pilot tests that teams try are yielding improvement. These types of measures are ideally frequent enough to support rapid cycles of piloting and adapting (or abandoning) approaches until a change is found to be both feasible and worth spreading. We touch on this type of measure only briefly in this section because it is the focus of a later section on how we know if changes are an improvement.

CORE and their measurement partner developed an innovative on-track metric but paused to revise it midway, and schools never conceptualized their work as being driven by the aim.

In Chicago, where some of the ninth-grade on-track work that CORE was building on was developed, a student's ninth-grade on-track status was measured in a straightforward way as whether students accumulated 10 full credits (one for each course each semester) and had no more than one semester F in a core subject (English, math, science, or social science) by the end of their ninth-grade year (University of Chicago Network for College Success, n.d.). CORE decided it was important to create a new measure because, "[w]hile traditional on-track measures based on the 'ABCs' (attendance, behavior, and credits) do a good job of identifying students who may be at higher risk of not graduating, they under identify students who may be on track to graduate but are underprepared for the rigors of post-secondary education" (CORE Districts, personal communication, October 1, 2021). This decision was an important adaptation of on-track work to the California context where most students graduate from high school (86 percent in 2023), but only about half of graduating students (44 percent of all students) meet the eligibility requirements for the University of California (UC) and California State University (CSU) systems; success in meeting A–G requirements is far lower for historically underserved groups, with only 36 percent of Hispanic students, 32 percent of African American students, and 25 percent of Native American students meeting those requirements in 2023 (Gallegos & Willis, 2024). CORE contracted with their long-term data partner (Education Analytics) to develop a metric that would assess the extent to which each individual student was on track for postsecondary success.²

² Because CORE and Education Analytics collaborated on these metrics and the precise role of each organization changed over time, we do not attempt in this section to parse who did which part of the work, referring to it all as being done by CORE.

After convening experts and conducting focus groups of educators to understand their responses to different ways of calculating and messaging the measure, BTSC released its first BTSC on-track metric based on an understanding of the ways that progress through A–G courses, grades, and attendance were believed to effect on-trackness. The metric rolled up data on A–G courses taken, GPA, attendance, and credits earned to determine the extent to which each student was on track. Importantly, a student needed to meet all the criteria to qualify for a lane. The resulting CORE metric had five lanes of on-trackness, shown in Figure 2, which provided more nuanced information than typical binary on-track measures. Based on PACE’s analysis of the difference between the Chicago metric and the BTSC metric, the most important innovation was in the ability to identify those students in the lane CORE called “high school graduation ready”—those who would likely graduate high school but, in the context of California, have only 60 percent odds of meeting the A–G requirements.

Figure 2. Criteria for BTSC On-Track Lanes, Checklist Method

	Course taking	GPA	Attendance	Credits
Postsecondary competitive	Five select A–G courses with "B" or better	3.5+	95+ percent	N/A
Postsecondary promising	Four select A–G courses with "C" or better	3.0+	95+ percent	N/A
High school graduation ready	Both A–G math and English courses with "C" or better	2.0	>95 percent	At least 55 credits earned
Vulnerable	One A–G math or English course with "C" or better and one other high school ready criterion (GPA, attendance, or credits)			
At-risk	Do not meet criteria for "Vulnerable"			

Source: Adapted from a slide from a BTSC learning session presentation, CORE Districts, January 14, 2020.

Note. *Post-secondary competitive* was defined: “Will almost certainly graduate high school and meet A–G requirements—meaning a student was very likely to be A–G eligible to attend a 4-year California University or Cal State University.” *Post-secondary promising* was defined: “Will almost certainly graduate high school and have a 90% chance of meeting A–G requirements—meaning they are likely to get into college.” *High school graduation ready* was defined: “Will almost certainly graduate high school and absent additional support will have a 60% chance of graduating A–G eligible.” *Vulnerable* was defined: “Absent additional support will have an 80% chance of graduating and only a 20% chance of being A–G eligible.” *At-risk* was defined: “Absent additional support will have a 50% chance of graduating and only a 5% chance of being A–G eligible.”

Participants found this model easy to understand because it clearly delineated which of the four data points led to a student being placed into their on-track lane. A CORE staff member explained:

[I]t was very graspable and relatively easy to comprehend for our schools as they were engaging in the work. And I think it helped just to really clarify a lot of the

distinctions that were important for our aims. ... [It was also] a source of good information to progress monitor how kids were doing as they got through the first half of the year.

In fall 2019, CORE set an overall network aim and worked with schools to set school-level aims based on the BTSC on-track metric. As slides presented at the BTSC Fall Convening stated, the network aim was:

We will produce breakthrough improvement in our students' life trajectories by transforming the ninth-grade experience:

- We will decrease the percentage of students who are at-risk or vulnerable by at least 20 percentage points.
- We will increase the percentage of students who move up at least one on-track level by at least 20 percentage points.

This double-faceted aim incorporates both the lane students end up in and a measure of growth (i.e., students who "move up" lanes over time). This allows schools to get credit for improvement while also keeping a focus on getting as many students as possible to be on track. While there was some variation in the specific percentages in schools' aims, Cohort 1 schools set aims in fall 2019 that were generally in line with the overall network aim.

Even though educators found the metric easy to understand, over time they raised concerns about how some individual students were classified. For example, it was counterintuitive for educators that if a student had all As and Bs in five A–G classes but missed more than 4 days of school in the first semester, they would be classified as only "high school graduation ready" at the end of that semester. In fact, additional analyses by CORE (personal communication, June 2020) showed that attendance was very predictive of high school graduation but less predictive of students' meeting A–G requirements. Furthermore, the checklist did not take into account the fact that higher difficulty courses, like honors and AP classes, were highly predictive of students meeting the A–G requirements. In response to these concerns, CORE paused the use of the BTSC metric at the end of the 2019–20 school year and began working with Education Analytics to revise the metric based on predictive analytics work conducted using years of longitudinal data from the CORE districts.

CORE originally intended to have the new "compensatory points" version of the BTSC ninth-grade on-track metric available in October 2020. However, when CORE initially released a revised version of the metric in fall 2020, they quickly took an "equity pause" for the rest of the school year to address concerns about how particular student subgroups and schools landed in the new system. Throughout, CORE communicated with the network and district leaders about the original metric, the rationale for the new metric, and the revised metric and sought feedback

from multiple teachers and district leaders. The Compensatory Points metric was ready for use at the start of the 2021–22 school year, in time for the launch of Cohort 3.

CORE called the new approach the “compensatory points system” because it allowed strengths in any measure to lift a student’s on-track lane (Figure 3). CORE also renamed the lanes to indicate three levels of postsecondary likelihood (competitive, promising, and potential) and two levels of vulnerability (vulnerable and highly vulnerable).

Figure 3. Criteria for BTSC On-Track Lanes, Compensatory Points Method

9th Grade Model									
On-Track Lanes	From:	To:	Interpretation						
Post-Secondary Competitive	80	100	Will almost certainly graduate high school and graduate meeting A-G; most strongly prepared for success in college						
Post-Secondary Promising	70	79	Will almost certainly graduate high school; 9 in 10 chance of graduating A-G						
Post-Secondary Potential	50	69	Will almost certainly graduate high school; 3 in 5 chance of graduating A-G						
Vulnerable	30	49	4 in 5 chance of graduating; 1 in 5 chance of graduating A-G						
Highly Vulnerable	0	29	1 in 2 chance of graduating; 1 in 20 chance of graduating A-G						
A-G Subject	Course Level			Grade Point Multiplier					Points Possible per School Year
	Low	Medium	High	A	B	C	D	F	
History	1.0	1.0	2.0	x4	x3	x2	x1	x0	8
ELA	3.0	3.0	3.0						12
Math	2.0	2.0	3.5						14
Science	2.0	2.0	2.5						10
Language Other Than English (LOTE)	1.0		1.5						6
Visual Performing Arts	0.5		0.5						2
Other AG subjects (elective)	0.5		0.5						2
Estimated Maximum Points (A-G)									54
Other Factors									
Overall GPA	7.0	(weighting factor= 7.0) X Current Year Overall GPA							28
Attendance Rate	<80%	0	80% to 90%	6	90% to 95%	9	95% to 100%	13	13
Total Credits Earned	0-39	0	40-49	4	50-59	4.0	60+	5.0	5.0
Estimated Maximum Points (Other Factors)									46
Total Possible BTSC Points									100

Source: Reprinted from a CORE Board packet, CORE Districts, October 1, 2021.

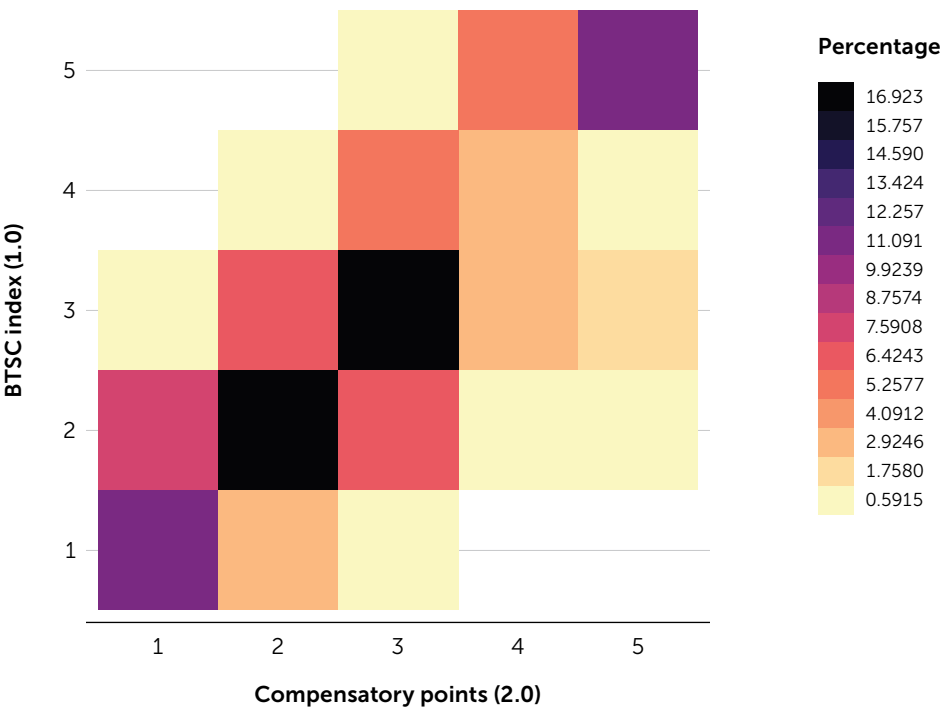
While the compensatory points method provided a more nuanced understanding of students’ on-trackness than typical binary measures, the technical complexity had a downside. It ultimately took CORE over a year to revise the metric, partially because CORE invested time to get feedback on early iterations of the compensatory points method, adjusting their approach in response to concerns raised by educators. The delay was amplified by the disruption to schooling from the pandemic—including the state pausing measurement of student progress through grades and standardized tests, which were the core components of the metric. Regardless of the reasons, schools and CORE were without a way to track a school’s or the network’s overall progress towards their aim for about 18 months.

After they unveiled the revised metric, CORE attempted to refocus the network on the aim by having schools set an aim each year. Additionally, some coaches prominently displayed each school’s aim on meeting agendas. Nonetheless, when teams described their progress in our interviews, the aim was not the standard against which they evaluated their work, as would be expected in a Breakthrough Series or improvement science network.

The “compensatory points” BTSC ninth-grade on-track metric produced different results and was not as transparent as the checklist, which led to initial challenges for schools in Cohorts 1 and 2.

Not surprisingly, given the changes in how the metric was calculated, the revised metric placed some students in different lanes than they would have been in using the original metric. Figure 4 is a heat map showing how students (in this case, students in BTSC schools during the 2022–23 school year) would have been classified on the BTSC Checklist (1.0) versus how they were classified on the BTSC compensatory points ninth-grade on-track metric (2.0). In the heat map, darker colors indicate more students. The bottom left to upper right diagonal shows students who would have been classified in the same lane regardless of the metric, which represents 60 percent of the students. The complement was also true: 40 percent of the students would have been placed in different lanes depending on whether CORE used the 1.0 or 2.0 metrics. PACE does not know which version of the metric more accurately predicts students’ academic trajectory.

Figure 4. BTSC Checklist (1.0) Versus Compensatory Points (2.0)



To our knowledge, CORE did not share an analysis like this with schools as part of the rollout of the revised metric to help educators understand the similarities and differences in the two approaches; they also did not present evidence of the improved validity of the model (if there is such evidence).

The change in how the metric measured students' trajectories was a challenge for some teams. A CORE coach explained how one of their teams questioned the move to the new metric:

I think we have huge implications ... a lot of this work is going to be very much questioned. ... I'm thinking of some of the [teams] that I talked to, they're very savvy and [saying], 'Well, we've been doing all of this work, this seemed to work for us and we understand the metric itself. Why are you now pausing and going back and redoing a metric that we've been using for the last 18 months? It's suspicious.' ... It's something now that [the coaches] have to go back in and re-sell. ... I think it's going to land for a lot of people, but ... for some of my schools it's ... 'We're not understanding what's going on here.'

While the transition was not smooth for all teams, the new model addressed some of the concerns about the original checklist, but at the cost of added complexity. For example, a student with 89 percent attendance was not automatically dropped down to the middle, yellow "high school graduation" lane (the name for the original yellow lane) on the compensatory points metric. Instead, that hypothetical student would have received 6 of 13 possible points for attendance and could still have attained higher BTSC lanes if they took challenging classes and got good grades. While this improved the face validity with educators for some types of students (e.g., those with good grades who just missed attendance benchmarks), the checklist made clear the discrete measures that affected a student's on-track lane, which teachers found helpful. In contrast, CORE chose to present the compensatory points system in an aggregated way to prevent educators from "gaming the system" by focusing on how to increase points rather than by more holistically improving student outcomes. This choice made it difficult for educators to see what specifically a student needed to improve to move up an on-track lane.

Other aspects of the revised metric were also not intuitive for educators. For example, the various weights assigned to course level and the fact that those levels were based on predictive analytics—not solely on how or if schools designated course levels—were confusing. Based on the predictive analytics, some students (e.g., newcomers taking sheltered English classes, students with IEPs in support classes) were essentially locked into lower ratings with less chance of "movement" if those classes were not A–G approved because of how their courses were weighted. The algorithmic calculation of course difficulty was particularly confusing for educators in schools that had made a policy decision to place all students in the same level of classes (i.e., not offering "honors" sections) for ninth grade.

In interviews, CORE program staff initially struggled to explain all the information included in the compensatory points model. As one stated, "members of our team have looked at this 37 times and still don't always fully understand it. ... I think we've explained it to two different professional improvement advisors. [It] took 45 minutes each time." The "black box" of the new compensatory points metric made it even more difficult for coaches to help participants to

understand what needed to change (e.g., courses taken, GPA) to improve an individual student's—much less a school's—on-track rate. Nonetheless, after the initial transition, teams in Cohorts 1 and 2 adapted to the compensatory points version of the metric. Cohort 3 schools knew only that version and never raised the same level of concern about the on-track measure.

The CORE on-track metric traded the benefit of a multifaceted measure for time and clarity.

Based on the analytics work used to develop it, we feel confident that both versions of the BTSC on-track metric are likely better at predicting how on track students were than any other single measure available to educators in BTSC schools. CORE strove to get the BTSC on-track metric out to schools as frequently as possible so that they could use it to monitor student progress regularly. They never fully achieved their goal.

The process of districts uploading their BTSC schools' data to CORE and then receiving data back from Education Analytics took a few weeks every time and required effort on the part of both educators and CORE staff. During 2022–23, for example, CORE aspired to release updated metric data on all students every 6 weeks (which was more frequently than in previous years). In contrast, the component parts of the BTSC metric were updated daily in schools' student information systems. As a result, by the time schools got the metric data, they were usually somewhat outdated, so educators sometimes relied on their schools' data instead of the metric data. This was particularly true in cases where a progress report came out in the time between when districts sent data to CORE and when schools received the on-track metric. Additionally, BTSC's decision to combine multiple measures in the metric sometimes obscured which aspects of student performance were the root cause of students' struggles. For example, teams needed to run additional analyses on their own to see what portion of students with one or more Ds or Fs also had attendance issues because the on-track metric combined these two different components into one measure. The component parts on their own, like grades and attendance, were often more informative for the types of interventions that would help a student or group of students improve.

Overall, interviews make us question why CORE did not rely on schools' existing data systems to monitor student progress on more regularly and test change ideas. One team member explained:

I have used [the hub dashboard]. [But w]e have a data dashboard for [our site]. I do 3-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.

This quotation exemplifies the experiences of many educators we interviewed who valued receiving on-track metric data for each student as they started ninth grade (by calculating the eighth-grade on-track metric for incoming ninth graders over the summer), at the end of the first semester, and at the end of the second semester. This cadence of data—which would be less burdensome for CORE to deliver—could still have supported monitoring progress at the network, school, and student levels.

During the BTSC on-track metric’s hiatus, CORE introduced the Developmental Relationships Survey (DRS), which required substantial effort to administer; educators appreciated it, but it was not used in a valid way.

In fall 2020, while the BTSC on-track metric was on hiatus, CORE decided to introduce the DRS, a student survey from the Search Institute with 20 multiple-choice items and three open-response items. The DRS can provide insights into relational dimensions of how students are experiencing a school, and using it as an annual diagnostic is in line with research from the What Works Clearinghouse’s (WWC) 2017 guide *Preventing Dropout in Secondary Schools* (which is conceptually adjacent, though not identical, to the goal of BTSC).

CORE decided that schools should administer the DRS to all ninth graders three times each year, using the results to track changes in grade-wide, subgroup (e.g., English learners, selected focal students), or individuals’ responses from the beginning to middle to end of the year. They asked teams to set aims around the resulting change in DRS scores as well as the D/F rate. CORE asked teams to administer the DRS in ways that yielded a high response rate. This effort was a heavy enough lift that some teams focused their PDSA cycles for some of 2020–21 around DRS administration.

When teams received data back from CORE, many appreciated the insights they got about students (which will be discussed more in the section on the relationships driver). From a purely technical perspective, however, the decision to track student “improvement” on the DRS as a component of the aim (or even just as a measure of outcomes for PDSA cycles) is problematic because, according to the Search Institute’s website at that time, the Search Institute had not validated the DRS for use to measure change over time. While it is common for “practical” measures (i.e., data collected from the normal operations of a classroom or school) to be used for PDSAs, those measures by definition require minimal effort. Putting substantial effort into collecting a measure that you plan to use repeatedly to measure change over time, even though it has not been validated for that use, is not a recommended practice.

Overall, we question how CORE used the DRS for three reasons: (a) there is no evidence that the DRS is valid for measuring change, so it should not be administered three times a year on the assumption that it is a valid measure of improvement; (b) it is also unclear if the DRS is related to students’ meeting A–G requirements, so it should not be used to track progress towards

network goals; and (c) getting a high response rate on the DRS took a lot of effort, leaving team members less bandwidth to test changes that might directly improve student outcomes. The way BTSC used the DRS three times a year was problematic, but you could nonetheless make a reasonable argument that the DRS provided information on student experience that was worth the effort of *annual* administration, in line with research (WWC, 2017).

In conclusion, CORE spent substantial energy in BTSC on developing innovative outcome measures to track progress towards the aim. Ultimately, we find that while CORE's motivation to create a more refined measure of on-trackness had merit, their iterative work to create a better outcome measure left the network without a measure on which to base their aim for three semesters. We also raise questions about the validity of the innovative measures they selected, due to differences in how students' on-trackness was measured in the different versions of the BTSC ninth-grade on-track metric and the Search Institute's published guidance on how to use each of its measures appropriately. Combined, these measurement issues lead us to question the trade-offs for the innovative measures that CORE selected relative to the costs in terms of school and CORE staff time as well as the validity of the resulting data.

We now pivot from the measurement infrastructure to CORE's process for recruiting schools.

Recruitment

IHI describes the process for recruiting organizations into a Breakthrough Collaborative thus:

Organizations elect to join a Collaborative through an application process, appointing multidisciplinary teams within the organization charged to learn from the Collaborative process, conduct small-scale tests of change, and help successful changes become standard practices. Senior leaders from participating organizations are expected to guide, support, and encourage the improvement teams, and to bear responsibility for the sustainability of the teams' effective changes. (IHI, 2003, p.4)

We use this description as a point of reference for understanding BTSC's recruitment process and the extent to which this aspect of the work mirrored a Breakthrough Collaborative.

Given the clear alignment between BTSC's goals and California high schools' need to improve A–G completion, CORE did not anticipate much difficulty in recruiting schools. CORE reached out to districts to identify potential BTSC schools and conducted a 1-day site visit to meet with key leaders at schools nominated by their districts. In the initial conception, CORE planned to have schools do a system analysis: examining historical data, interviewing students about their experiences, and making an inventory of their existing transition processes before

attending the initial network launch meeting. However, CORE struggled to recruit schools in a timely fashion, and most schools that joined the first cohort showed up at the launch with relatively little understanding of BTSC and without having investigated their own systems. Recruitment for Cohort 2 was smoother but did not include clear expectations for schools to do a system investigation or require a deep commitment from senior leaders to champion BTSC work. Recruitment for Cohort 3 was more varied because some schools recruited into a broader randomized controlled trial of NSI grantees.

The school recruitment process, especially for the first cohort, did not build sufficient buy-in from teams about why they needed to be in BTSC and the amount of effort they would need to put in to realize tangible improvements.

Our initial data collection showed that team members in Cohorts 1 and 2 often felt their district “voluntold” their school to join BTSC. Site administrators were typically not deeply invested in the work and did not feel responsible for its long-term success. Coming into the first orientation meeting for Cohort 1, team members did not necessarily have a clear understanding of BTSC or see it as a priority strategy for meeting school goals. One administrator explained:

I pretty much was voluntold. This is what we're doing. ... I think anything works if you have buy-in. And so, for ourselves, we definitely had to go through a buy-in process at first. ... From the very beginning, I definitely think there were a few folks on the team—and even for myself—there were questions around just the relationship and the nature of the consultancy [that BTSC provides] ... because I've worked with other consultants before, in terms of different types of school projects—different networks and organizations. And some have been tremendously wonderful and help us move our practice, and some not as much. ... We definitely had to go through a process around inviting [our CORE coach] in, really going through and looking as a team, what is ... BTSC? ... What is CORE?

As this comment shows, recruitment did not rapidly build the type of buy-in envisioned by IHI. Instead, coaches needed to gain time with members of school teams to build relationships and trust.

Perhaps the most important missing piece of the recruitment process was that, in most schools across both initial cohorts, principals did not commit to the level of support described by IHI in terms of guiding and encouraging the improvement teams or bearing responsibility for the success and sustainability of the work. Uneven support from school principals (who were rarely on teams) meant that school teams struggled to get even basic resources. In the first two cohorts of BTSC, schools struggled to establish a regular cadence of meetings for their Improvement teams (the teams assigned to work with CORE on BTSC) and their Freshman Success teams (the teams schools were supposed to create to monitor student progress and support students

who needed additional help). As a result, BTSC became an on-track network where participating schools did not necessarily have a team meeting regularly to monitor whether students were on track and refer struggling students to additional supports. Without administrators' commitment to championing BTSC, teachers not on the team were often unaware of the work and did not feel like they were expected to try ideas promoted by BTSC, which hindered the spread of the work even in later years.

One final data point highlights how the recruitment of Cohort 1 struggled to build buy-in to BTSC and how this improved over time: of the 14 schools from Cohort 1, six dropped out of BTSC; in contrast, only one school from Cohort 2 and no schools from Cohort 3 had dropped out through 2022–23.

As we will see in the next section, the lack of protected meeting time also fundamentally changed the work done in one of BTSC's most important drivers: the adult teaming driver.

CORE did not require BTSC schools to conduct a system analysis, denying CORE key information and removing a possible source for motivating school engagement.

In traditional improvement science—though less so in a Breakthrough Series—work begins with participating organizations analyzing their systems to understand current processes and performance shortfalls. BTSC's recruitment did not systematically include a substantial system analysis. Our data lifted up three major issues in the implementation of BTSC that could have been avoided by a well-conducted system analysis:

- Individual schools, and the network as a whole, did not understand what caused students to be off track in participating schools. Key questions that such an analysis could have addressed include, for example: How were schools already on-track monitoring? What data did schools have that could be useful for on-track work? What existing structures did schools have to support ninth graders when they first started to struggle? What proportion of off-track students had low attendance? How were Ds and Fs distributed across courses and individual teachers? This type of knowledge could have had implications for the change ideas CORE designed, which change ideas each school focused on, and schools' belief that BTSC could help them improve.
- BTSC was slow to leverage existing structures (e.g., students cohorted into houses) that could have provided opportunities for key activities like on-track monitoring.
- For the first couple of years of BTSC, the master scheduling driver focused on the idea of getting ninth graders scheduled in A–G classes instead of remedial classes because starting in remedial classes can drastically narrow the pathway to meeting A–G requirements. However, most participating districts had existing policies designed to ensure that all students' schedules enrolled them in sufficient A–G classes in ninth grade.

Our data suggest that there are multiple benefits of ensuring that schools do a system analysis, including building buy-in for the need for change and providing the information necessary to make good decisions about how to do improvement work. The vignette about Pine High School later in this report provides a counterexample of how a coach who had prior experience working with one of the BTSC schools—and who understood the school’s structures, processes, and culture—and an administrator who supported the work were able to build traction for BTSC.

In conclusion, the first round of recruitment did not engage site leaders, establish a consensus for teams about why their schools needed to be in BTSC, or explain how BTSC would help them make meaningful improvements; in turn, leaders did not bear responsibility for ensuring the success of BTSC teams. The result was 42 percent attrition from the first cohort. Later cohorts had minimal (6 percent for Cohort 2) to no attrition (for Cohort 3), but schools in later cohorts still did not consistently start with a shared understanding of their schools’ needs or a vision for how BTSC would help meet them. As a result, one key role for BTSC’s ongoing learning sessions and the processes of testing new ideas was to build school buy-in to the work. We turn next to the content of the change package to see how it attempted to meet this need.

Content of the Change Package

The main premise in the Breakthrough Series approach is that knowledge about how to improve outcomes in an area exists; if organizations use the Model for Improvement to test those evidence-based practices and learn how to implement them in their context, they can then spread those practices in their organization and reliably improve outcomes. Breakthrough Series are run by a hub (in this case, CORE) whose role is to articulate a theory based on existing research that explains how organizations can improve. One main lever for this in a Breakthrough Series is the faculty, including a chair with deep knowledge of the content area. The faculty create a “change package” (i.e., specific interventions that research says organizations should put in place to improve outcomes) based on research that helps organizations operationalize the research base. The faculty also create an associated measurement plan so that participating organizations can measure their implementation of research-supported ideas in their local context and see if their implementation is leading to improved outcomes (IHI, 2003). This section describes the drivers and change package in BTSC. It also shows that the theory embodied by CORE’s change package evolved in ways that were not aligned with the research base and the core tenets of continuous improvement and a Breakthrough Series—namely, that systems are perfectly designed to get the results they seek to achieve.

CORE did not find guidance from existing research and experts' knowledge to develop a change package based on the type of evidence that undergirds typical Breakthrough Collaboratives.

CORE did not identify a consistent “faculty” for BTSC, as envisioned by the IHI Breakthrough model. Instead, they undertook the task of synthesizing research and expert knowledge during their planning year (2017–18) on their own, doing an evidence scan to identify the key “drivers” (i.e., the main foci of their work where they believed changes that schools made would yield improvement towards the aim). CORE read research and interviewed researchers and leaders with expert knowledge about improving on-track rates. CORE engaged with research on early-warning indicators (see, for example, Balfanz & Byrne’s 2019 analysis), a long line of work from the Chicago Consortium on improving secondary school outcomes (see, for example, Easton et al., 2017), and the work of the University of Chicago Network for College Success (2023), which provided strong support for ninth grade as a pivotal year and examples of how a program with similar goals worked in another context.

The WWC issued a practice guide in 2017 called *Preventing Dropout in Secondary Schools*. This guide provides the type of clear causal evidence that is ideal for a Breakthrough Series, though the outcome of focus is only a partial match for ninth-grade on-track. CORE did not rely on that guide as part of its examination of research, and we recognize that CORE would have had to have been cautious in extrapolating from this guide, given their focus on ninth-grade on-track. However, because of the strength of causal evidence that the guide provides and the fact that—as a practice guide—it synthesizes causal studies and recommends specific practices, we nonetheless use it as a reference point as we examine what CORE selected for the change package.

Notably, existing research and experts that CORE consulted suggested many and varied concepts—which differed in terms of the nature and extent of their research backing—that were thought to support improved ninth-grade on-track rates. The wide array of potential drivers is perhaps best illustrated by the fact that at the expert convening CORE held in October 2018, 24 ideas got support from attendees as potentially impactful. Of these, the five most frequently identified (by dot voting) as impactful for improving ninth-grade on-track were:

- building school-site capacity around data interpretation and action based on data,
- conducting targeted academic interventions for off-track students,
- ensuring that a master schedule is set up to allow all students to enroll in A–G courses,
- establishing relationship-building practices for school staff to use with students, and
- providing school-site professional learning and reflection around identity, bias, and cultural responsiveness.

Unlike some of the examples that IHI Breakthrough Series used (e.g., specific patient discharge protocols, the degrees a bed should be elevated to reduce lung infections during hospitalizations), CORE’s process for interrogating the research did not result in a small set of

specific changes or protocols with *causal* evidence of success across multiple contexts—that is, clear proof that implementing a change leads to quantifiable improvements in outcomes—as the IHI model assumes. Said differently, CORE’s search turned up ideas that had a degree of evidence and support from experts but that did not meet the threshold for causal evidence of improved outcomes that typically undergirds a Breakthrough Collaborative. CORE also chose not to have the support of a faculty (as described by IHI) with deep expertise in relevant causal research to operationalize the existing research, with its flaws and shortcomings, into a change package that schools could attempt to implement.

CORE selected a set of five drivers that they believed would be impactful based on their scan of research and expert opinion.

1. **Freshman Success teaming:** The design of this driver followed closely the work of the University of Chicago Network for College Success (2023). That model includes an Improvement team, which leads the improvement science work in ninth-grade on-track (and, in BTSC, would interface directly with CORE), and a Freshman Success team, where adults come together to have asset-based conversations about students who are struggling and plan interventions. As we will describe later, CORE shifted to a model led by a single team and added specific meeting protocols designed to structure conversations about struggling students.
2. **Strong student-adult relationships:** One of the design principles of BTSC’s change package when the first cohort launched was “Every student knows: ‘You belong here. We care about you.’” Over time, this driver evolved based on the Search Institute’s framework and corresponding survey for understanding the dimensions of developmental relationships with adults that help students succeed (the DRS).
3. **Supportive transition from eighth to ninth grade:** For example, improve the outreach of high schools to feeder middle schools and improve students’ orientation to high school to build their familiarity with the campus and their connectedness to adults and peers.
4. **Master scheduling practice:** CORE originally conceived of this driver as focusing on making sure all students had access to A–G courses in ninth grade, based on research and expert guidance. This driver evolved to a focus on cohorting students and teacher common planning time when CORE learned that the districts and schools in BTSC already placed students in A–G-appropriate courses in ninth grade.
5. **Grading practices:** Based substantially on Joe Feldman’s work detailed in the book *Grading for Equity* (2018) and Doug Reeves’s work on improving grading practice, this driver presented research on the inequities stemming from traditional grading approaches and suggested strategies such as allowing retakes, creating a class period for students to make up missing assignments before grading periods closed, and eliminating zeros for missing assignments (i.e., creating a 50 percent floor, where the lowest grade possible for an assignment—even if it was not turned in—was 50 percent).

This list was partially aligned with the recommendations of CORE’s expert convening, with some deviations based on CORE’s beliefs about what would work best for their schools during the time frame of the grant. CORE’s beliefs about what schools could accomplish in 3 years placed instruction improvements and academic interventions to support struggling students outside the scope of BTSC. CORE recognized that they would need to further develop and refine a change package through the work of the network using evidence that participating schools accumulated around specific change ideas. To do this well, CORE would need to gather data about the effects of change ideas on students as teams tested them in schools, which is the focus of a later section. The remainder of this section describes the evolution of the drivers and change package.

CORE changed from thinking of the drivers sequentially to trying to get schools working in all five at once; along with challenges that schools faced in getting meeting time for BTSC, this led to changes in the desired structure of BTSC teams.

Prior to delving into specific drivers, it is important to understand how CORE thought about the way teams would engage with the drivers over time. Initially, CORE conceived of the drivers as sequential. They launched Cohort 1 informing schools about all five but starting work on transitions and adult teaming; in 2019–20, they started work with Cohorts 1 and 2 on relationships as well. After 2 years, CORE realized that if the drivers were addressed sequentially, progress would be too slow to get through all the drivers, so they created a new structure to facilitate work on all five drivers simultaneously, which we refer to as the “five-driver model.” To understand the shift in team structure, we need to return to how schools came to be in BTSC.

Initially, CORE conceived of BTSC as having two main teams: the Improvement team, which would interact with CORE and test change ideas, and the Freshman Success team, which would focus on on-track monitoring and having asset-based conversations to figure out how to support struggling students. (These structures are described in more detail in the next subsection.) But few schools invested in regular protected meeting time for both BTSC teams, which affected the nature of BTSC. Entering the 2020–21 year (when schools were doing hybrid or online education because of COVID-19), BTSC streamlined the two-team structure for BTSC with the five-driver model. Each school was asked to identify a team lead and a “champion” for each of the drivers. (The team lead could be a champion for one of the drivers.) Champions were typically identified because they had roles or interests aligned with the work in that driver. CORE assigned each of their coaches to lead one of the drivers, presenting content at learning sessions to the driver leads, producing a change package that explained the work of the driver, and recommending tasks to be accomplished, which was known as the “Key Actions Checklist.” Coaches were also encouraged to provide support to driver champions across the network in the work of that particular driver (in addition to continuing to support their assigned schools), though most coaches did not make this a regular part of their practice.

This new five-driver model had clear advantages: There was now only one team that needed to meet regularly for BTSC, the identified champions owned leadership of a piece of work that they likely had an affinity for, and work would ideally proceed on multiple drivers at the same time. Disadvantages, described later in this report, included the loss of some critical team roles (especially grade-wide on-track monitoring) and the potential for siloed work as five individuals led work on five aspects of BTSC. The revised five-driver structure was in place by the time Cohort 3 attended its launch in fall 2021.

Next, we discuss the five drivers in turn, starting with the two that were most central to BTSC.

The adult teaming driver evolved from a focus on on-track monitoring and supports to a focus on structuring conversations about students designed to change adult mindsets.

At the inception of BTSC, CORE envisioned that there would be two teams at every school doing the work of ninth-grade on-track (see Table 2).

Table 2. Team Structure Proposed at BTSC Launch

Improvement team	Freshmen Success team
<p>What is it? Leadership team of teachers, counselors, administrators, and relevant support staff.</p> <p>Major responsibilities:</p> <ul style="list-style-type: none">• Manage the improvement work:<ul style="list-style-type: none">• oversee the vision and structure for the improvement work happening on site and• reflect on progress of improvement efforts and capture and spread successful practices that emerge.• Monitor and support students:<ul style="list-style-type: none">• regularly review data, responding to changes in course performance and attendance and• connect students to supportive interventions that exist beyond the classroom.	<p>What is it? Grade-level teams of teachers, counselors, and support staff who all work with a shared cohort of students.</p> <p>Major responsibilities:</p> <ul style="list-style-type: none">• Monitor a shared cohort of students.• Communicate regularly to:<ul style="list-style-type: none">• compare individual students’ course performance in different subjects;• share intervention strategies that have been successful for particular students; and• reflect on the impact of interventions tried.

Note. Adapted from a BTSC Change Package slide, CORE Districts, April 2019.

The structure and roles of these teams were created based on IHI’s Breakthrough Success team model (for the Improvement team) and research about ninth-grade on-track work in Chicago. In explaining the Freshman Success team in a presentation at the February 2021 learning session, CORE shared a quote from Emily Krone Phillips’s book *The Make-or-Break Year* (2019), which describes successful work in Chicago to improve ninth-grade on-track:

Freshman Success Team meetings were a **core part of the work** designed to keep freshmen on track to graduate at Tilden. During these meetings Holmes [the team lead] would lead the team through exercises to **parse the data** on freshmen grades and absences. The group would consider **broad trends**, such as which courses the freshmen tended to fail most frequently, as well as **specific cases** of students who were struggling. (Philips, 2019, emphasis by CORE).

CORE modeled the BTSC Freshmen Success team on this approach, which had evidence of success, and brought a leader from Network for College Success to speak to all BTSC teams at one of the learning sessions. The BTSC change package in spring 2019 shows that CORE anticipated that it would be providing new types of data on students to schools; schools would establish regular routines of looking at trends in those data, use presumably preexisting tiered intervention systems (because interventions were not in the change package) to support students who needed additional help, and test ideas from the change package to improve ninth-grade on-track rates. However, few principals established and protected meeting time for either the Freshmen Success team or the Improvement team. As a result, the work of routinely examining on-track data for broad trends to identify areas for intervention, examining existing practices to see what is working for students, and pairing students in need with tiered interventions did not materialize in most schools, and examination of data stayed at the level of teachers testing change ideas in their individual classrooms.

Because schools did not prioritize embedding Freshman Success teams into their existing structures, BTSC's approach to adult teaming moved further from the research base on ninth-grade on-track. In the process of changing what it asked schools to do around freshman success, CORE deemphasized some critical functions of on-track work and shifted communications from requesting a two-team structure to support ninth graders to asking schools to do specific activities. When BTSC switched to the five-driver model, it put the work of meeting to discuss students under the adult teaming driver. The Key Actions Checklist introduced during 2021–22 for the adult teaming driver included a specific protocol for meetings known as the KidTalk Protocol, which helped structure communications about individual students. (That was the original work of the Freshman Success team.) Notably, it did not include the functions of the Improvement team, such as reviewing what Philips (2019) described as “broad trends” in data across the entire ninth grade and connecting students to support interventions outside the classroom. BTSC materials explained the KidTalk Protocol as:

Designed to facilitate ninth-grade team conversations about an individual student's on-track status and school experience. This strengths-based protocol asks team members to reflect on the student's strengths and positive connections to school prior to brainstorming and prioritizing growth areas. The protocol also requires team members to commit to actionable next steps to support the student's success in school. Team members should plan to revisit the student within a few weeks to review the implementation and outcomes of changes developed by the team.

The materials went on to explain that the adult teaming champion should work with the BTSC team to select students who were off track to discuss using the protocol, though materials did not provide selection criteria.

The KidTalk Protocol ideally requires all teachers of a given student to come together to have a conversation about the student, but finding a time for all relevant teachers to meet and do the KidTalk Protocol proved to be a challenge. Many schools in BTSC did not have a regularly scheduled time when the requisite teachers were all free, unless students and teachers were cohorted into “houses” or “academies” within the school. And when those structures did exist, shared meeting time had preexisting routines and other topics competing for time. While some schools were never really able to test the KidTalk Protocol, others attempted it with only some teachers present; at least one solicited input from teachers virtually. When teams were able to meet, they consistently found that the KidTalk Protocol was relatively time intensive for each selected student, and as a result, they could only get through a few students each year.

In spite of the difficulty of fitting it in, this protocol was a top priority for BTSC, and many schools attempted some version of the KidTalk Protocol, with substantial local variation in the criteria for selecting focal students. A CORE staff member explained how the protocol was designed to work:

Because you are focusing on grades, academics, you're focusing on kids that are in your classroom every day that you have influence over. You're focusing on them in an asset-based way. [It's a] huge shift for how teachers ... talk about students with each other. ... [H]opefully you would discover some insights about kids. Like realize like, "Oh, I'm the only class they have an F in. They're doing well in every other class. The issue might be me and my relationship with this kid." Hopefully have some of your assumptions challenged in a way that helps you be more curious about other kids too. ... I think there's an idea that, in my mind, some of that personal work that teachers do in the context of those conversations might affect their orientation towards other students who are struggling in their class. ... [However] I think in terms of having a concrete strategy for changing structures and practices on a macro level at a school, there isn't a strong connection right now in how we're using KidTalk to get there.

CORE believed the KidTalk intervention would help a small set of students directly, shift adults' beliefs about students, and—through the change in adult mindsets—ultimately affect a larger set of students. Another CORE staff member explained the hope that as teams built norms around seeing student assets, BTSC could leverage that into having teams explore which other students had a similar profile and determine how to support them as well, but that process was not specified in the protocol.

Teams typically believed the protocol was designed to help the selected students individually, not as an intervention that primarily targeted adults' beliefs about students. Team members sometimes explained that they were working to figure out how to spread the KidTalk Protocol to more students:

We've got 82 [focal] students and we want all of them to get a KidTalk. I don't know the particulars of the next PDSA cycles for [our team lead], but I believe that's one of her goals, is to figure out, how do we make this a little bigger? Where do we go?

As this quotation shows, many schools thought of the KidTalk Protocol as an intervention to support struggling students, and their focus was on expanding it to all those in need. However, given the limited time devoted to team meetings and the high numbers of students who had at least one D or F in the schools (typically, more than 50 percent of ninth graders received at least one D or F), it was infeasible to expand the use of the KidTalk Protocol to reach all students in need.

Teams' misunderstanding of the purpose of the KidTalk Protocol influenced the way they used the protocol, which CORE recognized but did not consistently correct. A BTSC coach said:

[W]hat we noticed is that the kids that the teachers brought up are kids that really needed support. ... [T]hey were more Tier 3 than Tier 2. And we're trying to focus more on Tier 2. But we felt like we needed to meet the teachers where they were at. ... [A]nd so we continued with that. ... [I]f that's where their mindset was of like, we need to focus on these kids, because they felt that urgency.

This quotation shows two fundamental challenges that came up with BTSC's change package, especially in the adult teaming driver:

- BTSC leaders needed to decide what in the change package they wanted to ask schools to be "tight" on (i.e., to try to do in a very specific way) versus what they wanted to be "loose" on (i.e., enable discretion for how schools adapted something to the local context). Every time that type of decision arose, BTSC had to balance the potential pushback they would get from schools for trying to keep something "tight" versus the potential lower fidelity to how the change package operationalized CORE's theory of action (if they chose to be "loose").
- BTSC leaders needed to figure out if and how to communicate the rationale behind change ideas, which was challenging when the root problems CORE was trying to address were the educators themselves and their mindsets.

In the case of KidTalk, BTSC chose to be relatively "loose" and, in the process, did not clarify for teams how to do the protocol in ways that instantiated BTSC's theory about how it worked.

In fall 2022–23, the adult teaming driver added a second protocol called Grade Book Review, which returned the central idea of on-track monitoring to the adult teaming driver. It involved a team of teachers coming together 2–3 weeks into the marking period to identify students currently getting Ds or Fs, agreeing to test a change that they believed would lead to better student grades, and then having a follow-up meeting at least once later in the marking period to reflect on how well the change idea worked based on student progress. The first schools piloted Grade Book Review during winter 2022–23, so its use was not sufficiently widespread for us to have a clear sense of its effectiveness, but it provides evidence of CORE’s ongoing development of BTSC in ways that hold promise.

There is strong causal evidence from research to support a range of strategies for supporting secondary student engagement and student relationships; however, the BTSC change package did not focus on those approaches or have teams to test changes in schools’ systems for building relationships.

CORE’s 2018 expert convening elevated the idea of spreading adult–student relationship-building strategies into a central position within BTSC. Later, the isolation of the pandemic brought the importance of students’ relationships to adults and peers and their connectedness to school into stark relief. CORE made relationships a central part of their on-track work, both by elevating the eighth- to ninth-grade transition as a key time for building the relationships among students, adults, and their school and by devoting an entire driver to adult–student relationships.

As a reference point to understand the relationships driver, we turn to the WWC practice guide *Preventing Dropout in Secondary Schools* (WWC, 2017). This guide should be used with caution when looking for a research base to inform BTSC because using it requires extrapolating from the outcome of dropout prevention to A–G completion. Given the dearth of the type of causal evidence for BTSC compared to what typically undergirds a Breakthrough Series, however, the findings are worth considering. The WWC practice guide lifts up evidence around the importance of relationships for secondary school success. Its third recommendation, which has “strong evidence” (defined on p. 4 as “consistent” causal evidence that a practice “improve[s] student outcomes for a diverse population of students”), focuses on engaging students and improving their capacity to manage challenges. The approaches described here have the level of causal evidence that typically supports a Breakthrough Series: multiple randomized, controlled trials with valid outcome measures for a similar population and across a range of contexts.

More specifically, the WWC practice guide recommends that schools “engage students by offering curricula and programs that connect schoolwork with college and career success and that improve students’ capacity to manage challenges in and out of school.” The guide specifies three approaches:

1. Directly connect schoolwork to students' options after high school.
2. Provide curricula and programs that help students build supportive relationships and teach students how to manage challenges.
3. Regularly assess student engagement to identify areas for improvement, and target interventions to students who are not meaningfully engaged. (WWC, 2017, pp. 29, 33, and 37)

Additionally, the WWC practice guide recommends the following practices that have "moderate evidence":

1. Provide intensive, individualized support to students who have fallen off track and face significant challenges to success.
2. For schools with many at-risk students, create small, personalized communities to facilitate monitoring and support. (WWC, 2017, pp. 3–4)

For each of these practices, the WWC practice guide provides additional explanation of the practice and examples of what it looks like, along with information about the studies that provide the causal evidence in support of the practice. Because these studies did not assess the outcome of improving A–G rates, it would be especially important to test if these approaches also moved the BTSC ninth-grade on-track metric, but it is logical to think that they might because some of the factors that contribute to students staying in secondary school are also related to A–G completion. Overall, the WWC practice guide provides general support for BTSC's decision to center relationships and provides specific examples of approaches for relationships and student engagement that have been shown to be effective at reducing dropout rates for secondary students. Some of these recommendations—most specifically, those with "strong" causal evidence—fall outside the scope that CORE set for BTSC because they focus on courses, curriculum, and programs, but the third, fourth, and fifth recommendations above all seem potentially compatible with BTSC.

In the years that PACE studied, three main activities and change ideas dominated the work in the relationships driver: (a) administration of the Search Institute's DRS, (b) ideas from the Search Institute's *The Relationship Builder's Guidebook* (Pekel et al., 2021), and (c) a specific change idea known as the "2x10."

The Developmental Relationships Survey (DRS). In 2020–21, while schools were providing mostly virtual education, the focal work of the relationships driver was administering the DRS three times each year to all ninth-grade students. Teams invested substantial energy on getting a high response rate, with some school's BTSC teams testing change ideas to increase students' response rates. Schools came up with a range of plans to try to get data from all students, such as using a class all ninth graders took (e.g., English language arts) to administer the survey three times a year.

The fact that some schools ran improvement test cycles to improve DRS response rates is especially notable because, at that time in BTSC's trajectory, most teams were testing only one or two change ideas at a time. Administering the DRS more effectively is not actually a change to the system that could improve student outcomes in and of itself; as a result, better response rates could only improve student outcomes if the resulting data were used to make systemic changes in students' experiences. This is an important distinction because it means that the only way to understand the potential impact of the DRS on on-track rates is to look past the question of administration and response rates to how the data were used and if those uses, in turn, led to changes that improved student outcomes.

CORE worked hard not only to support the administration of the DRS but also to build Tableau dashboards to help educators see and analyze the resulting data. Educators generally appreciated the insights that the DRS gave them about students. A team lead explained that the fact that the DRS quantified the quality of student relationships was a real value-add for educators at their school: "I've never seen anything like that before, and I'm like, 'Wow, that's really cool,' because it does make such a huge difference but often it's not captured, it's just anecdotal." Additionally, using data from the DRS to identify students who need additional support is potentially powerful, and educators found the information that they received from the DRS to be actionable. One relationships champion explained:

The relationship data survey has been very helpful. ... [W]e each will choose a number of students who are scoring low on that Relationship Survey and try to help them feel more connected to school. I think that's fun. That's a good thing to do.

This quotation was typical of team members who appreciated the information that the DRS provided.

The DRS can be used in-line with WWC practice guide recommendations around annual surveys that identify causes of low student engagement, strengths and weaknesses across the school in how it supported student relationships, and groups of students who are at risk. However, administering the DRS three times each year was not supported by either prior research or information the Search Institute provided about valid uses of this tool. Additionally, bringing in an outside measure like the DRS raises questions about sustainability because it is not clear if or how schools will be able to continue administration and get data into a dashboard or some other format for analysis after the end of the grant. As a result, while we recognize that schools appreciated the DRS, we recommend that it be administered only annually in-line with Search Institute information about validity and the research base on the annual use of such tools, and we question whether the effort exerted to administer it three times each year was worth the opportunity cost (e.g., testing change ideas that might directly benefit students and spreading implementation if there was evidence that they improved student outcomes).

The Relationship Builder's Guidebook. *The Relationship Builder's Guidebook: Activities and Approaches to Enhance Developmental Relationships* (Pekel et al., 2021) was introduced to BTSC in winter 2020 when a staff member from the Search Institute was a guest speaker at the BTSC winter learning session. The guidebook provides more than 200 pages of information about developmental relationships and activities to get to know students and explore the developmental relationships framework, along with activities for each of the five dimensions of the framework. CORE made the guidebook available to all schools and encouraged relationships champions to select change ideas from the guidebook to test themselves and to get their colleagues to join in testing those or other ideas. Some of the ideas in the guidebook have strong evidence behind them, such as greeting students when they enter a classroom (which the WWC practice guide lists as an example of an “everyday strategy that teachers can use to foster student engagement”). While many driver champions tested ideas from the guidebook, we do not know of cases where an idea was tested, found to be successful, and ultimately spread to all ninth-grade teachers. Instead, the more common scenario was an almost scattershot testing of different ideas by a couple of teachers in some of the BTSC schools. Across the network writ large, CORE did not facilitate the consolidation of knowledge based on results from multiple schools’ tests about which strategies appeared to be most effective and how to implement them consistently. As a result, while the guidebook had reasonable ideas, and at least some (and maybe all) of them were well supported by the research base, we do not see evidence of this work leading to system change in participating schools; therefore, we do not have reason to believe that it would likely have a marked effect on the network or student’s on-track rates.

The 2x10. During the 2021–22 school year, a single change idea became most prominent in the relationships driver: the 2x10. This change idea was based on the understanding that even teachers who believe in relationships can get mired in deficit thinking and may not have effective strategies to form relationships with the large numbers of students they teach daily. The goal is to change the way that teachers view and interact with students as well as to build teachers’ self-efficacy.

To do the 2x10, a teacher selects a student (likely one who is struggling or with whom they do not feel like they have established a good relationship) and they take 2 minutes a day for 10 days to have conversations with the student about nonacademic topics. Over the course of the 10 days, the teacher is supposed to gain a better understanding of the student, and the student hopefully recognizes the teacher’s intentionality in building a relationship with them, building mutual trust. Some educators who tried this idea found it to be very powerful. One BTSC team member explained:

All of us have the anecdotal evidence that whenever we get to do our 2x10s—the 2 minutes for 10 days with ... a student who’s having some issues—it makes a difference. ... Yes, we would love to scale it out.

However, many teachers found it challenging to find a given student and devote their full attention to a nonacademic conversation with that student for 2 minutes a day for 10 days straight, so they ended up lowering the number of meetings. Additionally, keeping the goal of improving the on-track rate in mind, this intervention is a one-to-one intervention that would be hard to scale to all students who might need it in traditional high school settings with large student–teacher ratios and a significant number of students who are struggling in one or more classes. The experience of interacting with struggling students about nonacademic things reportedly had an impact on the beliefs of some of the teachers who tried it, and (as the previous quotation illustrates) it provided new insights about the focal student. But ultimately—like the KidTalk Protocol—this was an approach for changing adults’ mindsets, not an approach to changing system processes, which is what continuous improvement methodologies are designed for. This was most problematic because the members of the BTSC team—and especially the relationships drivers who were most likely to test these ideas—were often selected because their personal beliefs and roles already led them to center student–adult relationships.

One team member responsible for leading relationships work at their school explained their understanding of both the importance of relationships and where the change package for this driver fell short:

I think the relationships work is misguided, because people misunderstand the goal. ... [T]he goal of building strong relationships ... is to make it clear that even though I am a distant authority character that works for an institution, I'm a person and you're a person and that we can relate to each other in additional ways beyond authority. Building those relationships is a really fun part of the work. ... [T]he BTSC work as narrowly defined, is like, "Let's build stronger relationships so that we can accomplish something else." And where I'm at and where [my administrator] is at is ... [y]ou all [at CORE] have been talking about relationships for a long time. When is the "so that we can do something else" going to be talked about? So the 2x10 is very valuable, in addition to what comes after you build a strong relationship. When is the time to talk to the kid about their content strengths and their content skills gaps?

This team member raised a nuanced concern. They worked hard to build good relationships with their students and recognized the importance of change ideas like the 2x10. But they also noted that the 2x10 in and of itself did not provide students with the learning experiences they needed in school to become on track for lifelong success. What this comment did not address is the lack of systemic approach to supporting *all* teachers—not those who already find building relationships “really fun”—to build strong relationships. Overall, we find that while research supports CORE’s decision to work on student engagement and adult–student relationships, the way they went about that work diverged from both the research on practices responsible for improved outcomes elsewhere and continuous improvement methodologies’ focus on system change.

Based on prior research, CORE anticipated that a key lever in improving on-track rates would be getting students into college-ready classes. But when CORE learned that districts in BTSC already had policies that placed ninth graders into a core set of A–G classes, they improved their change package by shifting focus in this driver to cohorting students, which is aligned with research.

The master scheduling driver was initially conceptualized around the fact that in some districts in California, students who do not meet specific requirements in middle school are typically placed in ninth-grade classes that do not meet A–G requirements, making it very difficult for them to ever meet A–G requirements (especially if they get a D or F in one or more classes and get behind because, if they start in remedial classes, they do not have any room in a 4-year graduation timeline to spare). What CORE realized at the start of BTSC is that the districts whose schools participated in BTSC already had existing policies and practices that gave students course schedules with sufficient A–G courses. The challenge in BTSC schools was not enrolling students in A–G courses in ninth grade but supporting students to succeed in the A–G classes in which they were enrolled.

BTSC did not really start work on the master scheduling driver until they changed to the five-driver model during 2020–21. During that year, they reconceptualized the driver based on what they had learned about the actual scheduling challenges facing BTSC—most importantly, the challenges in scheduling Freshman Success team meetings and getting all of a student’s teachers to meet for a KidTalk Protocol. At this point, the driver shifted focus to cohorting students and getting common planning time for teachers so that they could better monitor and support students, a strategy that is supported by the WWC practice guide. By 2022–23, very few schools (arguably only one) had restructured ninth grade due to BTSC to create the cohorts recommended by research, though in a later section, we profile the story of Pine High School, which found positive effects from testing this strategy. More commonly, some BTSC schools had already organized their ninth graders into houses prior to starting BTSC. In those cases, BTSC had an inconsistent track record in inserting their routines around identifying and supporting focal students into the adult collaborative spaces that those preexisting house structures created.

While it is notoriously difficult to change grading practices, BTSC was successful with this driver in some schools in which it picked up momentum because teachers were motivated when the changes they introduced improved student outcomes, which was clearly visible in this driver.

The question about whether the content of the change package is aligned with strong evidence for improving student outcomes is different for the grading driver than for the four other BTSC drivers. California’s A–G policy means that students are eligible to apply to the UC and CSU systems if they get a C or higher in a specified series of courses. As a result, there is a clear causal link between students receiving Cs or better in all the ninth-grade A–G classes and

those students making progress towards meeting A–G requirements. For that reason, this section does not specifically cover research around grading practices and the A–G policy in California.

The number of students getting a D or F in one or more classes in ninth grade was clearly problematic when BTSC launched. (Recall that CORE presented data showing that 66% of African American and Hispanic ninth graders in the CORE districts failed one or more classes.) Under California Education Code, while school districts set grading policy (Cal. Education Code § 49067), teachers have the final authority to determine students' grades (Cal. Education Code § 49066). With grading decisions (though not grading policies) codified as being within teachers' authority, CORE recognized from the beginning that working to get teachers to make changes to their grading practice would be politically sensitive in some BTSC schools. Nonetheless, based on Feldman's work (2018) showing that grading practices are typically idiosyncratic, demotivating for many students, and poor representations of student mastery, CORE introduced work in the grading driver in 2020–21. Feldman's work argues for grading practices that are transparent and reflect student mastery of key content at the time a course grade is assigned, as opposed to typical grading practices that he sees as punishing students for noncompliance (e.g., turning assignments in late or not at all) or meeting learning goals on a later timeline than the teacher intended (e.g., retaining or averaging in a low grade on an early assessment taken prior to a student demonstrating mastery even if the student later masters the content). Fully implementing Feldman's ideas requires transparency about students' learning goals and levels of mastery, regular feedback for students on their progress, and other systems that support students in learning "soft skills" (e.g., following directions, completing tasks on time). In short, full implementation of Feldman's approach would require modifications to Tier 1 instruction and feedback in many classrooms.

The grading driver had two major assets in BTSC work. The first was COVID-19, which upended grading in California in spring 2020, followed by a failure crisis during 2020–21 as students who had not learned as much as usual during the prior year struggled (and received commensurately lower grades), especially in an online learning environment (Hough & Chavez, 2022). Solely online instruction during fall 2020 created a scenario where many teachers and school districts were more likely to consider changes to grading than they might have otherwise. The second main asset for this work is that some individual teachers in many BTSC schools were already reconsidering assessment and grading systems in their classrooms, including substantial moves towards more innovative mastery grading and other rubric-driven approaches. Such teachers were often selected to join the BTSC team as the grading driver champions.

In many schools, grading work was largely confined to the driver champion and one or two close allies. Broader resistance, as had been anticipated by CORE, manifested in many schools. One grading champion explained how she was exploring grading and had plans to try to spread it as some of her colleagues retired in upcoming years:

I try to implement at least one practice from this [Grading for Equity] book. ... And I'm just trying to go through trial and error and then bring back to my department what is working for me—what could work for them. Every Monday we meet for a PLC [professional learning community] and we try to get on board as a department because, ideally, what [School name] wants to do is get every single department grading similarly and then, from there, expand to a school-wide grading system.

But, the unfortunate part is some of our teachers are a little bit less open-minded. They have a grading system that works for them, and it works well, but it's not exactly what the research has shown to be effective in equity. ... My department, in particular, has a lot of teachers who will be retiring within the next 5 years. ... [M]aybe [we can get] these newer teachers who don't have a grading practice yet, to incorporate these suggestions from the book into their grading practice. ... That is my goal here for at least the next couple years.

The situation this teacher described was typical in many (but not all) schools, where the work of the grading driver developed a devoted following but had few mechanisms to spread in ways that made these changes a consistent practice across departments throughout ninth grade (or beyond).

CORE, on its part, promoted the grading changes suggested by Feldman that required the least overhaul of teachers' core instruction: allowing students to retake exams, providing a class period for students to make up missing work, not giving zeros for missing assignments, and "minimum grading" (i.e., not assigning any grades lower than a certain threshold, typically 50 percent, so that students do not slip out of mathematical range of passing a class a class due to poor performance on a given assignment). Most schools ultimately engaged in testing one or more of these change ideas in one or more teachers' classrooms.

Interview data suggest that part of what ultimately supported most schools to test change ideas around grading was the way this driver—more so than any other driver in BTSC—engaged the Model for Improvement, providing teachers with a clear connection between the change and an improvement in student outcomes. We discuss that in the next section.

Taken as a whole, the content of the change package is reasonable but is not connected to the research base to the same extent as a typical Breakthrough Series.

The ways the change package evolved from inception through the 2022–23 school year highlight CORE's ongoing learning, illustrate some of the challenges faced in the network, and affect the potential impacts of BTSC on student outcomes. The BTSC change package had roots in existing research on improving outcomes for secondary students; at the same time, it was less connected to a strong causal research base than the Breakthrough Series Collaborative

model intends. Specific improvements to curricula and instruction, even those that have strong research support, were scoped out of BTSC from the start. Taken as a whole, the change package—especially as it evolved—was missing core components of a research-based approach to ninth-grade on-track, including monitoring student progress and identifying additional supports for struggling students. Additionally, while the change package assumed that schools already had interventions to support students who were struggling, interviews did not substantiate that there were high-quality interventions available in all schools that students could be directed to via on-track monitoring.

In place of typical on-track structures and broader system changes, the first three BTSC drivers were all ultimately designed to support changes in adult mindsets around seeing student assets and helping adults build relationships with individual students. However, many of the educators on BTSC teams already seemed to know their students, cared about those students, and had strategies for building relationships with them prior to engaging in BTSC. For example, a quotation from a relationships driver lead earlier in this section described building relationships with students as a “fun” part of the job; another educator discussed the importance of leading with “love” to get even struggling students to take the risks necessary to succeed academically; and many others’ interviews and conversations at BTSC events provide evidence of a deep interest in and rich knowledge of their students’ lives as well as their academic strengths and weakness. It is possible that educators overestimated the quality of their relationships with students. Even if that was the case, the educators engaged in BTSC were both the ones whom the change package was most likely to reach and those who seemed to need mindset shifts the least.

In conclusion, CORE attempted to build a change package that mirrored a typical Breakthrough Series in its connection to evidence. However, health care research provides more clear causal evidence around interventions—due to its norms around randomized controlled trials—than is typical for interventions in education. Additionally, some decisions CORE made about what to prioritize in the change package did not closely adhere to either the existing research base or the foundational tenets of continuous improvement—namely, continuous improvement’s focus on changing system processes instead of individuals, which is based on the understanding that the outcomes we achieve are the natural result of the systems we have enacted. Furthermore, BTSC schools could not implement some of the elements around adult-teaming structures that had a stronger researcher base because district and school administrators who had decided to engage in BTSC (or had been told by their district that their school would participate in BTSC) did not provide the meeting time and other supports necessary to implement the most basic features of BTSC; this, in turn, led to CORE moving away from some key components of their original change package. The result was that the change package included ideas that were reasonable but that were not typically ideas about how to change system processes based on strong evidence from causal research.

Nonetheless, there was a plausible path from CORE's change package to the aim of increasing ninth-grade on-track. Continuous improvement has a built-in mechanism for figuring out which ideas are most efficacious and which should be adapted or abandoned: the Model for Improvement. The next section presents evidence on whether BTSC helped schools use this central approach of continuous improvement effectively.

How Do We Know if a Change Is an Improvement?

As described in the introduction to this report, the Model for Improvement is the central process by which schools should have been able to take ideas from the change package, test them in their system, and measure the outcomes to decide if they should spread, adapt, or reject the change. Part of CORE's role as the network hub under the IHI Breakthrough Collaborative model was supporting schools in assessing whether the changes they were trying were actually improvements. Measurement plays such a central role in improvement science that one of its core principles is "we cannot improve at scale what we cannot measure" (Bryk et al., 2015, p. 87). IHI explains the central role of measurement in a Breakthrough Series:

During Action Periods between the Learning Sessions, teams test and implement changes in their local settings—and collect data to measure the impact of the changes. They submit monthly progress reports for the entire Collaborative to review. ... The aim is to build collaboration and support the organizations as they try out new ideas. ... All teams are required to maintain run charts tracking their system measures over time and key faculty members review each team's monthly report to assess the overall progress of the Collaborative. (IHI, 2004, pp. 6–8)

In describing Measurement for Improvement on its website, IHI named the importance of many sequential "small tests of significant changes," which use "just enough data to learn." Continuous improvement has coined the term *practical measures* to explain the attributes of measures that provide just enough data to learn. Two key aspects of the definition are particularly relevant for understanding the types of measures needed for BTSC. Yeager et al. (2013) explain:

[I]mprovement efforts require **direct measurement of intermediary targets** ... in order to evaluate ideas for improvement and inform their continued refinement. For example, is a student's motivation and grit actually improving in places where a change has been introduced? ... [M]ost significant from a practical perspective, [such measures] need to be **engineered to embed within the constraints of everyday school practice**. (p. 12)

Yeager et al. go on to note that practical measures are ideal for addressing whether a change is an improvement. This section uses these ideas about measurement in a Breakthrough Series and in improvement science more generally as a reference point for

describing CORE's efforts to support schools in measuring whether the changes they were testing were an improvement.

CORE focused their measurement efforts in BTSC on innovative measures and tests of small changes rather than supporting schools to identify practical measures that would support small tests of significant changes.

Earlier sections described where CORE invested most heavily in measurement for BTSC: the BTSC on-track metric and the DRS. Additionally, CORE did not invest in supporting schools to identify the "practical" measures of implementation and intermediary targets that could have helped them to run the Model for Improvement. This task of developing a measurement system is traditionally the responsibility of a faculty in the IHI Breakthrough model, but BTSC did not have a faculty to do this work. When coaches tried to get schools to conduct and document PDSA cycles, CORE's BTSC on-track metric and the DRS were the main measures at their disposal. While these measures had strengths in diagnosing a system and measuring outcomes, they were ill suited as direct measures of "intermediate targets" because of the time lag in getting educators the BTSC on-track metric, the fact that the BTSC on-track metric combines multiple data points (course placement, grades, attendance) into one measure, and the DRS's lack of demonstrated validity as a measure of improvement.

Additionally, some of the ideas being tested did not match IHI's guidance that "small tests of *significant* [emphasis added] changes accelerate improvement" (IHI, 2003). Especially in the initial years before CORE developed the Key Actions Checklist, schools selected ideas to test relatively idiosyncratically. Some of the ideas selected were not inherently bad—for example, one school's pilot of multiple strategies to get parents to sign up for a school communication app—but they were also not supported by evidence of improved outcomes (as suggested by the IHI Breakthrough model) or sufficiently powerfully connected to the goal of improving on-track rates to merit consuming the BTSC team's time. A strong set of measures of interim targets could have made it apparent that these ideas should be quickly abandoned and replaced with more robust ideas that had a better chance of improving student outcomes.

But even once the Key Actions Checklist improved teams' focus on CORE's drivers, the change package included ideas designed to change adults' beliefs about students (e.g., the 2x10, the KidTalk Protocol), which are not typically measured in schools and for which CORE did not provide "practical" measures. These change ideas were also not conceived of as changes in the processes of schools, which is the type of change that continuous improvement is best designed to support. This was especially true for the changes in the relationships and adult teaming drivers. Across most drivers, the result was a combination of small tests of ideas focused on changing individuals instead of systems and a lack of measures that could detect whether those changes affected student outcomes. As a result, even when schools attempted to test ideas and document their process on PDSA forms, they were unable to gather strong evidence of whether their changes were an improvement.

The driver where schools had the most success with the Model for Improvement was the grading driver, where the practical measure of improvement was clear and readily available.

While CORE did not identify or develop measures of improvement associated with most change ideas or drivers, the real-time data on student grades that were maintained as part of all schools' existing student information systems were an obvious and ideal "practical" measure. The text box describes how a coach supported the entire ninth grade in one school to assess the efficacy of a range of possible change ideas for reducing the number of students getting Ds or Fs.

How One School Tested Grading Changes

At one school, a CORE coach used existing ninth-grade meeting structures to get teachers to try a range of approaches to help students improve their end-of-semester grades. The coach went to existing meetings and asked all teachers to identify one student who currently had a D or F in their classes and to select a strategy to try to help the student improve their grade. Teachers each picked the strategy they thought would work best, ranging from allowing students to retake tests to creating a make-up period near the end of the semester where students could complete and turn in missing assignments for full credit. The coach introduced a tailor-made tracker to help teachers log their intervention ideas, students' grades before the intervention, and their grades at the end of the semester. At the start of the next semester, they met again and looked at which interventions most consistently led to students' grades improving. Building in a period for students to make up missing work emerged as clearly the most impactful for enabling students to improve their grades to a C or higher, and additional teachers pledged to try that approach the following semester.

This test is noteworthy for a few reasons. The first is that it required minimal effort to run: There was no special data collection because it relied on "practical measures"—in this case, existing data on student grades. The second is that teachers ran a "small test"—it was only one student per teacher—of a "significant change"—making changes to the grading process in a teacher's classroom. There was also clarity about how they would know if the change was an improvement (the student who had a D or F would end the semester with a C or better) and a measure of that outcome. Finally, because student grades are a major determinant of whether they are on track, if the improvements were scaled, that would likely lead to improved outcomes.

In the example in the text box, the fact that the coach was able to leverage existing team structures to test change ideas about the process of grading and every teacher had some degree of buy-in to the results (because they participated in the tests) led an expanded set of teachers (including those not on the BTSC team) to bring new practices into their classrooms that they saw worked for their students.

Overall, however, the teams and CORE itself were not clear about the connections between changes they tested and either interim improvements or overall progress in improving the on-track rate.

In our data collection during 2022–23, we observed teams’ discussions about their change ideas and evidence of improvement as they worked at learning sessions and when we asked team members in interviews to reflect on the progress they had made towards improving on-track rates in BTSC. Comments typically reflected an inability to provide evidence that the changes they tested had improved student on-track rates or that the ideas could be scaled in ways that would improve on-track rates. Team members sometimes questioned whether their efforts were having an impact on students, or they rested their claims about the benefits of BTSC on stories of individual students who improved rather than broader evidence of success. For example, one team member explained:

[I]f you look at the numbers in overall, it’s hard to tell [if BTSC has led to improved on-track rates]. But when you look at individual kids, and you look at those stories of those individual students, you see it. And—I can list off the top of my head right now four at the very least that have benefitted. ... [O]ne particular girl, I won’t say her name, but [she] started off the straight Fs [and] was a tardy/truancy issue. ... [T]hat was ninth grade, and then [by] senior year, where she was on the [sport] team, A–G compliant, straight As, and then was able to go to a college. ... [S]o that transition might have happened on its own, [but] chances are that typically that you wouldn’t see that. ... They would maybe [have] graduated, [but] that’s it.

This individual teacher was very active and supportive of BTSC. Their comments include examples of students who did better than the teacher thought was typical for a ninth grader. However, they provide no evidence linking individual changes tested to interim outcomes; they also cannot link changes tested and spread to improvements in the on-track rate in the ninth grade overall. Instead, they offer a type of anecdotal evidence that is typical of more traditional attempts at reform than the type of evidence-driven approaches of improvement science and Breakthrough Collaboratives.

A CORE coach also reflected on the challenge that the change package was loosely based on research, and CORE did not have a fully developed system for establishing the efficacy of change-package components:

I feel less than satisfied with the fleshed-outness of everything. ... [I]n my dream world, things would just be ... a little tighter. We would just control the “fidelity.” I say “fidelity” because these are not actual really pressure-tested changes. These are not ... research based. Given that they’re not that ... I would love to have people participating in ... a more rigorous data collection.

As described earlier, CORE did not have the type of research base that typically supports a Breakthrough Collaborative. As this coach noted, the ideal way to respond to this situation would be to have a rigorous measurement approach that supports teams to test whether the changes they are trying are an improvement and a network structure that supports teams in learning about effective approaches from one another. However, the relatively low incidence of deep and consistent application of the Model for Improvement meant that schools and CORE alike did not get clear signals about which approaches they should spread and adopt widely, which approaches they should adapt until they could be spread and adopted widely, and which approaches they should abandon. Schools also did not get the interim evidence of effectiveness of any change that might spur broader support of improvements by educators and administrators outside the BTSC teams. In short, BTSC schools did not consistently apply the principles of continuous improvement to their BTSC work, so they did not get clear signals about which changes were likely to lead to improved student outcomes.

The next section describes the extent to which teams spread approaches based on whatever evidence they had in ways that might ultimately benefit student outcomes.

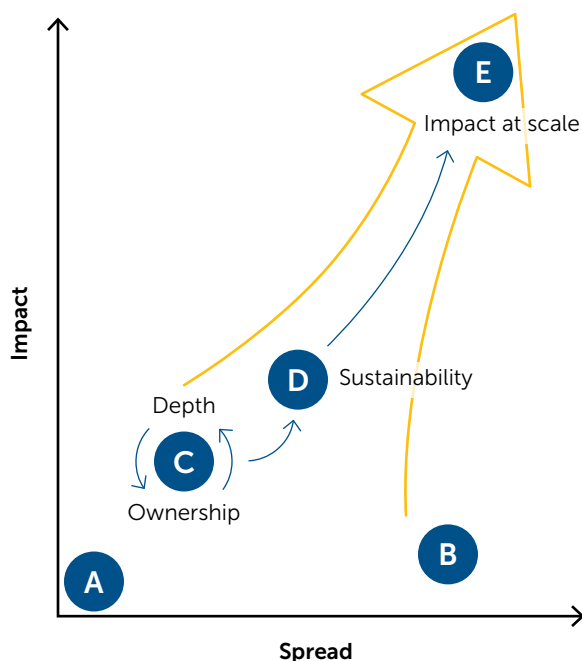
School Engagement With BTSC

So far, this report has examined the implementation of BTSC relative to models of continuous improvement and research on improving student on-track rates. We now pivot to examining evidence about the outcomes of BTSC in participating schools. This section focuses on the proximal outcome of *traction*, as defined by Gallagher et al. (2022). This framework overlays Coburn’s 2003 analysis of the features of reform implementation that lead to sustained, positive impacts on student outcomes at scale and continuous improvement’s approach to system reform. We build on Coburn’s definitions by explaining what each facet of scale looks like in improvement networks (Gallagher et al., 2022), which is what CORE would hope to support during the first few years of BTSC. We propose that *spread* in improvement networks begins when adults beyond just the members of the Improvement team test and (if evidence shows that a change is an improvement) spread the new approach. Given the centrality of the Model for Improvement and the use of data, we note that an additional feature of work that has sufficient *depth* is that there is evidence that the change is an improvement, and ultimately, there are shifts in educator beliefs. When there is sufficient *ownership* of the work, we find that typically, a champion is committed to a particular change or that work in a particular driver leads to

improvements for students and ultimately spreads outside the team to the broader staff. Finally, to achieve long-term *sustainability*, the work needs to move from being an additional thing that adults could choose to do to something that is embedded in structures and processes in the school after BTSC ends, thus promoting widespread and ongoing use.

Figure 5 illustrates the idea that, over time, if a school moved from initiating BTSC work (Zone A) to building depth and ownership without much spread beyond the BTSC team, the overall impact on student outcomes would likely be small (Zone C); additionally, the impact would be small if what spread lacked depth and ownership (i.e., changes were superficial and tests did not show improved outcomes; Zone B). To achieve impact at scale, all aspects of Coburn’s definition have to be initiated—that is, school-owned, deep work needs to be spread and integrated into key systems and processes (which requires leadership commitment) to be sustained over time (Zones D and ultimately E).

Figure 5. Traction Framework for Reforms Affecting Student Outcomes at Scale



We use this theory to examine the nature and extent to which schools engaged in BTSC and the likelihood that their BTSC work would affect student outcomes. Schools building traction with BTSC work would be a direct outcome of BTSC that in turn could lead to the ultimate outcome of improved student on-track rates.

CORE Struggled to Get Traction

This section uses the idea of traction as a leading indicator of whether BTSC is making inroads in participating schools that could enable BTSC to produce desired student outcomes. Examining traction as well as student outcomes is important because continuous improvement typically believes in starting small before spreading, so we might expect a time lag between small tests leading to system changes and improved student outcomes. Additionally, sometimes student outcomes do improve in the short-term, but if changes did not actually reach all dimensions of scale, then they could fail to be sustained after the grant period.

Especially during the first year with Cohort 1, CORE struggled to get schools to set the conditions necessary for BTSC, and as a result, the work lacked the momentum necessary for a successful Breakthrough Series.

As a network hub, CORE has consistently tracked schools' engagement with BTSC, though their tools have evolved as CORE itself continued to learn how better to monitor progress within the network. For Cohort 1, many schools did not provide the basic enabling conditions that could have helped BTSC gain momentum. Common challenges included schools not having regular times for educators to meet to do the work of ninth-grade on-track; BTSC teams and leads not making consistent times to meet with CORE coaches; and leaders (school and district) not showing clear support for BTSC and helping their staff understand how BTSC aligned with school priorities. In fact, a community-wide progress summary sent out to all BTSC schools in February 2020 (about a year after Cohort 1 launched) noted that one of the biggest challenges in the network was “**consistency** both in holding regular meetings and trying change ideas” (emphasis in the original). This is critical given that meeting and testing change ideas are the fundamental preconditions for success in an improvement collaborative.

Not surprisingly with such little school engagement, early reports on progress in Cohort 1 showed that relatively little on-track work was happening in schools. Of the 13 schools doing BTSC in February 2020, three did not submit any monthly progress information, and other data showed inconsistent engagement in BTSC in these schools. Table 3 shows that the remaining schools were trying a broad range of ideas around monitoring student progress, supporting focal students (with check-ins, behavioral interventions, and academic interventions), and preparing for summer bridge. Almost 1 year after the first cohort launched, most schools were working in one driver at a time (which was aligned with CORE's initial conception of the work) on testing a relatively idiosyncratic set of change ideas.

Table 3. School-by-School Highlights of Change Ideas Being Tested Reported by CORE

Past month's progress	Planned next steps
Counselors making positive phone calls home for 20 students, and surveying incoming 8th grade students when they visited campus	Separate surveys by feeder middle school and use the results to plan summer bridge
Examining student data for accuracy and considering running PDSAs prior to broad implementation of new ideas	Collect data on the current PDSA on getting students to tutoring
Using a universal checklist to build consistency when adults check in with students	Use a parent phone call in regards to 11 students needing more intensive support
Presenting the Developmental Relationship Framework to school staff	Get teachers to pick one student with whom to try one change associated with one element of the Framework
Send notification to students about tutoring to encourage attendance	Debrief PDSA about whether notifications improve tutoring attendance
Tested 1:1 counselor check-ins with target students; acknowledge on-track and "moving up" students with a certificate at an assembly; invite ninth grade teachers to select 4 target students to attempt (unspecified) interventions with; learned that afternoon sections of algebra have a higher D/F rate	Engage students more around the target student list
Held an all-day retreat for Freshman House leads; discussion of student progress monitoring revealed divergent approaches across houses	Create a hybrid intervention tracker using the best approaches from all houses
Successfully tried using lunch detention as a deterrent to tardiness	Offer incentives to students who are on time
Learned that behavioral interventions produced minimal improvement for focus students; developed an intervention tracker for focal students and selected students	Shifting to academic interventions and using the tracker to collect data on efficacy
7th period class offered for student needing additional support is showing success; recruiting teachers for Summer Bridge	Develop a script for mentors to use when meeting students

Source: BTSC community-wide progress summary provided by the CORE Districts, February 2020.

This list is notable because many of the ideas being tested are not very tightly connected to the change package and, as a result, do not represent a coherent, research-based theory of what needs to be changed to lead to improved on-track rates. During the 2019–20 and 2020–21 school years, most BTSC schools remained in Zone A of the traction framework in Figure 5.

Over time, the changes CORE made to their approach supported schools in doing more of the things BTSC was trying to get them to do.

During 2020–21, CORE was trying to get schools to do work in the relationships driver and at least one other driver (e.g., transitions, adult teaming). BTSC made two major changes in 2021–22 and 2022–23 designed to increase the pace of BTSC work in schools: the five-driver model and its accompanying Key Actions Checklist.

As explained previously, CORE introduced the five-driver model in 2021–22. Most schools were successful in naming champions for most of the drivers and getting work started in multiple drivers (taking into account that the transitions and master scheduling drivers were both seasonal because they aligned with particular activities that occurred only once a year). This quotation from a team member in fall 2019 describes a school team that is excited about BTSC but hasn't yet figured out how to convert the enthusiasm from team members into BTSC work:

I think we have great individuals on the team, [who are] open-minded and want to do the work. ... We're all excited about looking at a whole grade level. That's another success. And then we always come ready to the meetings with a lot of ideas. Now, the next step is getting those ideas off the paper and making it a reality.

As this quotation shows, in 2019 this team was interested in doing work on BTSC but had not really been able to get that work going. The mix of excitement about BTSC and lack of clear vision for how they would successfully engage with the work was typical for participating schools at that time. By fall 2021, that team had identified champions for all five drivers, and the team was meeting every few weeks and either planning or testing change ideas in some of the drivers. A team member explained:

[E]ach time we're talking about, what are you working on this week? Are there any struggles or any things that you need? And then we brainstorm, are there any other people that you could bring into that? So, that's what we try to do.

The change expressed in these two quotations from the same team exemplifies the positive results of the changes that BTSC made in the interim. Another quotation from the same team in fall 2021, however, exemplifies one of the challenges teams found as they were using the five-driver model: "We just got to figure out, how does it all fit together? And so, that's where we are. Each person's doing their own thing, but we're also trying to come together and have a cohesiveness." While the five-driver model was not a panacea, it did help most schools move from working largely on just one driver at a time to having two or more drivers active (depending on the time of year).

Along with the five-driver model, in 2021–22 CORE introduced the Key Actions Checklist to build cohesiveness across the drivers, which previously had had change packages structured uniquely by the developer of each change package. The Key Actions Checklist was a template completed by the CORE coaches in charge of each driver to specify the tasks that should be completed in a driver between learning sessions. The Key Actions Checklist was summarized in a 1-page document (see Figure 6 for an example), which provided a concise overview of the work that the team lead and each champion would ideally be leading over a specified period.

Figure 6. Sample 1-Page Overview of Key Actions for April–June 2022

DRIVER CHAMPION KEY ACTIONS: APRIL-JUNE 2022						
BTSC Team Lead	Continue to lead BTSC Team Meetings and support Driver Champions	Monitor 9th grade on-track data and hold a data conversation	Hold a learning consolidation to take stock of change ideas tested to date	Hold a Mid-Year Pulse Check to help your team reflect on its routines & culture	Support access and use of Relationships Data	Share BTSC Team progress with Principal
ADULT TEAMING CHAMPION	Create focal list of Students for KidTalk	Hold regular meetings using asset based KidTalk Protocol	Monitor next-steps and changes to determine if they resulted in improvement	Hold a KidTalk Reflection Meeting at the end of the MP (incls submitting data) to improve your KidTalk process		
8TH-9TH TRANSITION CHAMPION	Consolidate Learning and Prototype “Transition Measures”	Identify Focal Groups of Students Based on 8th Grade Data	Identify Focal Change Ideas for Improving Sense of Belonging for Incoming 9th Grade Students	Co-create Plan-Do-Study-Act with Transition Point Person		
STUDENT-ADULT RELATIONSHIPS CHAMPION	Administer Spring Relationships Survey	Learning Consolidation: Use the Spring Survey results to: (1) identify success cases, (2) the changes that led to those successes, and (3) systems and structures		Reflect on your role and how your work supported relationship-building practices across your school		
GRADING PRACTICES CHAMPION	Schedule time in BTSC Team Meeting to share learning of Grading Practices this year and implications for next year.		Work with allies and principal to determine where, when and how D/F rates will be monitored in the 2022-2023 school year		Complete learning consolidation by reflecting on end of semester grades. Prepare to share at start of school.	
MASTER SCHEDULING CHAMPION	BE CURIOUS!	Review Current State: Leaks - Pre-emptive problems - Repeaters - policy for non a-g courses	Teacher Team Measures: LING - Index (D/F rates, DR survey, Attendance)	Drop? Any movement Chances <ul style="list-style-type: none">Teacher Common Prep	Drop? Any movement Chances <ul style="list-style-type: none">Student Cohort	

Source. CORE Districts, 2022.

This shift to the Key Actions Checklist and its concise, 1-page summary represented a major shift in tone for BTSC because it was relatively directive about what teams should be doing and testing. While initially CORE took a tone of *requesting* schools' engagement and responding to lower engagement by trying to ask for what they thought schools would deliver (e.g., dropping the Freshman Success team from the change package when they realized schools were not making time for that structure), the Key Actions Checklist *set expectations* for what strong engagement in BTSC would look like. Importantly, the Key Actions Checklist was used as guidance, not a set of steps to be checked off; schools and driver champions could customize their work to meet local needs and context, but the clarity supported coaches and schools alike in their work. The Key Actions Checklist helped narrow the set of ideas that most teams tested and, in doing so, got more teams testing ideas that CORE believed would lead to improvement. The five-driver model and the Key Actions Checklist together led schools to do more BTSC work, pushing most schools out of Zone A.

BTSC did not support schools to change the systems that were producing the outcomes.

The central assumption of continuous improvement is that systems, not individuals, produce outcomes. Therefore, to use continuous improvement successfully to change the outcomes, you need to change the structures and processes that make up a system. However, there were three ways that BTSC was unsuccessful in getting most participating schools to undertake systems reform: (a) BTSC did not engage a broad coalition of support in many schools that could have enabled them to spread ideas and insert BTSC into key processes and structures; (b) many changes in the adult teaming and relationships drivers tried to change individuals, instead of changing systems; and (c) in most drivers, BTSC did not produce evidence that changes were an improvement, so the warrant for spreading changes was largely anecdotal. As a result, even though the BTSC team members did more BTSC work after the advent of the five-driver model, the work still did not get broader traction in most schools.

At the start of BTSC, CORE did not develop deep buy-in from district or school leaders to BTSC. While districts "voluntold" schools to participate, the districts did not follow that up with an expectation that principals would make BTSC a school priority, ensuring that their ninth-grade team was committed to testing and spreading the ideas offered by BTSC. In many schools, teams were dominated by those who were willing and had available time—often, newer teachers who were interested in reform. At the extreme, one school's entire team was composed of counselors, case workers, and staff from a nonprofit organization that provided a range of student supports, with neither teachers nor administrators regularly participating. Teams that did not include influential teachers and did not have administrative support faced challenges in spreading testing to those not on the team and in getting BTSC embedded into the central structures and processes that can institutionalize reform. And, as one team member explained, if BTSC remained an isolated initiative, it was an uphill battle for BTSC to make a difference for ninth-grade on-track:

It is really hard [for the BTSC team] to just be a separate entity. I feel like ... all the ninth-grade teachers have to be a part of it. It's just, if we want to make a difference for our ninth graders, then this has to be something that is just not a separate five, six people on the side situation. And I feel like it should just be a team of ninth-grade teachers that are working, are doing this work. So, it's hard to have little pieces here and there.

This sentiment exemplifies the challenges that many schools had during the first 4 years of BTSC in getting beyond Zone C of the traction framework—that is, in getting traction with the work beyond a small group. Vignettes in the next section, which discusses schools that got deeper engagement in BTSC, show what a difference it made when broad coalitions within schools supported BTSC and helped expand ideas being tested past the BTSC team and into the broader ninth-grade work in those schools.

The second major reason that BTSC did not change schools' systems is that key change ideas, especially in the relationships and the adult teaming drivers, focused on changing individual adults rather than the systems and processes that would affect an entire grade level. The belief undergirding those drivers was reasonable—some educators' mindsets impede their ability to meet all of their students' needs—but the change ideas did not create new processes and structures to support those adults in improving their knowledge and skills and changing their behaviors in substantial ways. A core tenet of continuous improvement is that individuals' beliefs and behaviors are a result of the systems they are in, so to change those beliefs and behaviors, you need to change the systems, not the individuals. But the change packages presented ideas designed to change adult mindsets without a system to support all adults engaging in those activities and improving student outcomes. As a result, BTSC did not fully leverage the strengths of continuous improvement.

The final issue that may have impeded scaling BTSC is the way CORE chose to think about the processes of testing and implementing new practices. One of the major problems with reform in education is the way schools "implement" reform by changing structures but never actually changing how adults perform their central roles of teaching, counseling, and supporting students (Tyack & Cuban, 1995). Preventing rapid (and typically superficial) implementation was a central focus of BTSC, where both learning sessions and internal tracking documents focused on whether a team was testing ideas at a scale that was "small enough to learn" rather than making widespread but superficial changes. However, BTSC attended insufficiently to another main threat for the success of new initiatives: not having a path to scale. On multiple teams, the team members were engaging in work that they believed was valuable, but BTSC materials and events did not focus on how to help schools move from these small tests of changes to learning to spreading implementation. The measurement system that CORE presented made finding evidence of improvement from early tests—which is a crucial component of getting a broad set of teachers to try a new idea—very difficult in all but the grading driver (because, outside of

grading, there were no “practical” measures to show that the changes the BTSC team members tried were actually an improvement). CORE’s approach attempted to help schools avoid Zone B of the traction framework (broad spread of superficial ideas), but the lack of purposeful strategies for spread and insufficient evidence of effectiveness of tested ideas left many schools in Zone C (narrow adoption of new practices) instead.

Over time, BTSC started elevating the idea that BTSC team members should find ways to present information at grade-level or staff meetings to raise awareness of BTSC work. In 2022–23, CORE started tracking the state of testing and spread (planning; small-, medium-, or widespread testing; implementation; or abandonment) within the network. While these steps might not be sufficient to support most schools to scale new practices (given the nature of the change ideas and the lack of data from most tests of new ideas), they do suggest that CORE became aware that they needed to focus more on spreading changes if they wanted to affect enough students to have the potential to improve overall on-track rates.

Learning About Building Traction From Positive Outliers

While the preceding section described general challenges for CORE’s efforts to get BTSC schools to scale substantive changes, in this section we explore the stories of three “positive outliers,” schools at which we see that BTSC work gained traction towards scale—depth, ownership, spread, and sustainability—via specific efforts by both coaches and school team members. These sites exemplify some of the challenges and opportunities for beginning and then scaling an initiative of this kind. We also consider preliminary evidence about student outcomes in these schools (all of which we refer to by pseudonyms to protect confidentiality).

Pine High School. Pine High School has about 2,000 students, with an almost equal split in demographics between White and Hispanic students. Pine is a comprehensive high school that has an honors program students can apply for before the start of ninth grade if, according to the school website, they are “confident of their math and science aptitude.” Students who apply are not guaranteed acceptance. As could be expected, interviews suggest that the honors program attracts a disproportionate number of the higher achieving students in the ninth-grade class whose families are informed about the program and believe their students would be successful in honors science and social studies while taking an additional 10-credit technology class (on top of the typical course load) in their ninth-grade year. BTSC focused on students in the regular pathway. A strong BTSC team that included teacher leaders and a supportive administrator was able to pilot and then expand a new structure with embedded supports that improved ninth-grade on-track rates for students who were not in the accelerated program.

Pine’s BTSC coach, who was employed by Pine’s COE, was already working with Pine before BTSC started. As a result, he came to his role coaching BTSC with an understanding of Pine’s structures and initiatives and with relational trust established. As he explained, his prior experience with Pine was very helpful in launching BTSC:

I did have a preexisting relationship with the district and, more, with the principal. ... [B]ecause of that, I think his willingness to gather a team and his trust in our endorsement of this work set us up for success right from the beginning. One of the things that I think is different in a positive way about the way our county coaches are engaged with these teams [as opposed to the CORE coaches is] ... the preexisting relationships that we had in most of these schools—the knowledge we have of the larger system that they operate in. I think it helps us to be responsive to where this BTSC work sits in that system and the implications it has for some other parts of the system.

In addition to the preexisting relational trust, this coach was aware that a set of teachers had already done a book study on *Grading for Equity* prior to BTSC, and he understood the honors program and that the classes outside it had a higher concentration of students who might need extra support. This knowledge was key to how BTSC unfolded at Pine.

At the same time, the coach recognized that BTSC was a small part of his job (he estimates 10–15 percent). As with the other BTSC coaches who worked for COEs, there were some drawbacks in this role of not working for CORE. For example, COE coaches often could not attend internal CORE coaching meetings and were typically less well versed in the details of the change package and other processes that CORE was promoting. Additionally, he did not have the time to be as hands-on as some of the coaches who worked for CORE. Recognizing both his limited time and the strong leadership skills of BTSC team members, the coach focused on rapidly shifting the onus of responsibility—and, by extension, ownership—over to site-level team members. He reflects that “the work is not so contingent on [my support] that it can’t happen if I’m not there. ... They’ve also taken on ownership of the work in a way that I have to worry less about than I did a couple of years ago.” One asset that helped the team assume leadership was an assistant principal who listened to teachers, empowered them to test changes and spread ideas to their colleagues, and had the authority to make changes to some of the structures and processes at Pine.

Pine’s work focused on a small test of a significant change—namely, grouping a set of ninth-grade students. The BTSC team grouped three teachers (English, math, and science) and arranged the master schedule so that a “cohort” of students took classes with these teachers. These teachers had previously participated in the *Grading for Equity* book study, and they agreed to a uniform method of grading: not deducting points if student work was late and implementing a 50 percent floor on assignment grades. Because they shared students, the teachers could coordinate things like due dates on significant projects, discuss students across subjects, and create spreadsheets to monitor the progress of all students in the cohort closely. When the teachers identified students who had poor attendance or low grades, they used the KidTalk Protocol to structure their communication about how to support those students better, offering interventions and monitoring whether those interventions helped students improve.

Unfortunately, the school's data system was not configured in a way that made it easy to run a test of the effects of this change. So the team took the student IDs of all cohorted students and manually set up a comparison between those students and the rest of the ninth grade. A team member explained that after the first year:

We saw that our kids that were in a cohort ... group versus our other freshmen had fewer Ds and Fs in A–G classes; they earned more credits than the average freshmen, and ... the GPA was higher than the average ... freshman not in a cohort.

As they saw positive results, administrators encouraged team members to “tell their story” of grading reform and share their data about effectiveness with colleagues in their departments, to complement the administration’s message supporting grading reform. As BTSC team members shared the approaches that they tried in the cohort and the evidence they had about positive effects on students, some members of the science department and Integrated Math 1 teachers agreed to test grading reform during the spring semester. The administration later credited the expanded test with an overall reduction of about 15 percent in the number of ninth graders who received a D or F in Integrated Math that spring.

After testing the cohort model with three teachers and one cohort of students, Pine expanded the cohort model to include two cohorts of students and three additional teachers the following year. The BTSC team was highly strategic in spreading out the original three cohort teachers, positioning one as the “lead” of the new cohort and making sure the original team effectively mentored new cohort teachers. In and across the two cohorts, teachers collaborated on additional tests of meaningful changes like grade floors, competency-based grading, and Tier 3 interventions guided by information gathered during the KidTalk Protocol. With six teachers and two student cohorts of 100–120 students each, about 50 percent of the school’s ninth graders were in one of the two BTSC cohorts in 2022–23. Additionally, ongoing examination of student progress showed that students in the two cohorts had a 10 percent higher on-track rate than the other half of the ninth grade, which included the students in the honors track.

Pine’s use of cohorting proved effective, both in terms of teacher collaboration and student progress. At the same time, one teacher notes, the model needed refinement because of how it interacted with the honors track in the master schedule. For example, cohorted students were only able to enroll in Earth Science and College Prep English and they could not enroll in honors English or higher level science. The same teacher who raised concerns about the current cohort model said that she believed all students in the ninth grade should be cohorted to fairly distribute students, make it possible for all students to take honors English and more rigorous science in ninth grade, and equally serve the whole grade level. Such an expansion would continue the work of both spread and sustainability. In addition to considering expanding the cohorting further at Pine, the principal began talking about the approach in district meetings, prompting other schools in the district to consider cohorting; when the assistant principal was offered a position as principal at another school, he planned to take the idea to his new school as well.

Pine's experience in BTSC was unusual in that it started in Zone C because preexisting relationships meant that the work already had buy-in from a small set of staff and could pilot important changes because of administrative support. However, this case also highlights several aspects of the work that seemed important in other schools that experienced them as well.

The first was the use of data—which is a fundamental element of continuous improvement—alongside the personal experiences of educators who tried a change to build a warrant for spreading several ideas that they tested beyond the pilot stage. Even though the team needed to invest extra effort to compare their students' achievements to the rest of the ninth grade, those data were critical to spreading the change to other teachers in the grade, creating a second cohort the year following the initial pilot, and even enabling Pine's administration to present their success to other schools in the district so that they could consider the changes as well. When data indicated that the changes helped students, Pine was able to expand to about half of the ninth grade over the roughly 3 years of BTSC.

Additionally, the role of an engaged administrator was critical at Pine. It is not possible to leverage a school's structures or processes (e.g., the creation of the master schedule) without administrative support. Administrators do not need to be team leads, but they need to be sufficiently engaged to authorize changes in policy and practice and to provide necessary resources to support the work.

Finally, Pine attributes its success to its ability to embed the grading changes and KidTalk Protocol suggested by CORE into the cohorting structure. This structure helped teachers get to know a small set of students better than they would have otherwise. The teachers regularly communicated about the progress of the entire group of students, provided them with a coherent set of experiences, and offered students who needed it additional support. This retains the core features of the Freshman Success team that was originally central to CORE's change package but was dropped when too few schools were willing to commit to implementing it. It is also an example of how the KidTalk Protocol, when adapted by embedding it into a process, could become something that was available to support an entire group of students in addition to its potential to change adults' mindsets.

Magnolia High School. Magnolia is a modestly sized high school serving primarily underserved students (68 percent Hispanic, 19 percent African American, and 7 percent Asian American/Pacific Islander). Magnolia is organized with a ninth-grade team and theme-based "academies" for 10th through 12th grades. Unlike Pine, Magnolia did not begin with a foundation of relational trust or the kind of systemic understanding that facilitated the school team's having quick ownership of the work and a broad base of support for small tests of substantial changes. Instead, the initiation of BTSC at Magnolia was stymied by several contextual factors, and the work was slow to gain traction.

When BTSC first began, there were concerns that the school might be closed or consolidated due to low enrollment. Interviews also revealed that Magnolia teachers felt that they were “voluntold” to join BTSC without even much advance notice that they would need to leave their campus for 2 days to attend the BTSC launch meeting. Initially, teachers expressed frustration that BTSC did not align with their preexisting work or goals for the ninth-grade experience at Magnolia. A CORE staff member explained how these factors combined to make launching the BTSC work challenging at Magnolia:

I think they didn't have the best introduction to this work. ... They had some teacher leaders, a case manager, and a counselor [at the launch], but they all had been told to come ... the day before. So they're already pretty salty about being pulled from their classroom with zero warning, which is super legit. And then I think throughout the day they were feeling frustrated that they weren't getting a more concrete training of what they're supposed to do, and they thought ... they had heard all of this before. They're like, "We know that the research says the ninth-grade year matters, so tell us something we can do about it." They were really frustrated, and they also, almost the entire team, didn't come back [to the BTSC launch] for day 2.

In a context where organizational survival felt like a pressing concern and teachers perceived BTSC as an unwanted additional initiative, most were reluctant to become involved or take up the drivers suggested by the initiative.

This opening dynamic made it hard for the CORE coach to institute regular meetings with the team. During the first 2 years, the principal offered passive support but provided no active assistance in the form of dedicated meeting time or staff access; instead, an assistant principal was named as a BTSC team co-lead at the site. The people who regularly attended BTSC team meetings were the assistant principal and support staff members (e.g., counselors and case workers), but few teachers. Key teacher leaders, who led a standing ninth-grade team meeting that determined the course of ninth-grade operations, did not initially participate in BTSC. These teacher leaders had substantial organizational leadership responsibility and successfully played a gatekeeper role, keeping BTSC as a side initiative and rarely granting BTSC space on the ninth-grade meeting agenda. Regardless, CORE's coach regularly came to the school, met with whoever was willing to meet with her, and used her skills in continuous improvement to add value to their work. She also tried to attend the ninth-grade meeting any time she could to learn more about the school and have people get to know her.

Through her persistence and skills, CORE's BTSC coach built institutional knowledge and relational trust through work with administrators and staff members. In 2021–22, the assistant principal and BTSC co-lead transitioned to being the site principal. She left the BTSC team because she no longer had time to be the lead but continued to champion the work, facilitating the spread of BTSC into the ninth grade and getting the teacher leaders—and former

gatekeepers—to assume ownership of BTSC. Over time, trust between teachers and the coach developed recursively, with the CORE coach better understanding the politics and programmatic structures of the school and Magnolia teachers better understanding the logic, drivers, and tools offered by BTSC. This cycle built a strong sense of co-ownership of the work.

Once the ninth-grade teacher leaders trusted their coach, they found ways to make parts of BTSC coherent with their ongoing work to improve ninth grade. In particular, these teachers saw that BTSC could have direct impacts on their work around student intervention, a realization that created an opening for greater collaboration. The ninth-grade team began acting as a unified group, conducting small tests of significant changes and giving the work increasing depth. The first major change idea tested was an intervention to address Magnolia’s issue of student (non) attendance. While the ninth-grade team had an existing communication process called “student intervention time,” this had been a one-way conversation, which consisted largely of teachers reporting issues and problem behavior to a counselor. The ninth-grade team instead tried using the KidTalk Protocol at one instance of the meeting to have a more asset-based discussion. Through testing the protocol, teachers developed a whole new way of talking about their “frequent flyers” (students who regularly raised concerns). By the next meeting, one of the teacher leaders had taken ownership and adapted the protocol, and, as the CORE coach described:

[The teacher leader] always puts her own spin on things, so she customized the protocol [and] customized the [student] tracker. She always improves everything. ... We talked about how [the original “student intervention time” process] was very much like teachers just unloading problems on the case manager as opposed to just thinking about things that they could do in their class. ... The depth of things that came up [with the revised protocol] and the solutions that came up were totally different than a no-fly list or a suspension. It was just a really powerful conversation. ... People talked about the difference of how it felt—like there was more agency in saying things that I could do in my class than just telling the case manager what to do. That was the KidTalk revolution at Magnolia.

The idea of adapting change ideas to work better in local contexts is central to continuous improvement and was exemplified by this process, which built depth and ownership of the KidTalk Protocol. Embedding KidTalk into a broader process meant it could reach any ninth-grade student whom adults thought it would benefit, which is the goal of spread.

In addition to the success of the Frequent Flyer/KidTalk test, the ninth-grade team reconceptualized another of their preexisting structures: the end-of-semester makeup time called “power hour.” Magnolia historically had “power hour” as a chance for students to do or turn in missing work. The team pulled ideas from the BTSC relationships driver and tested inserting them into this existing structure. One team member explained:

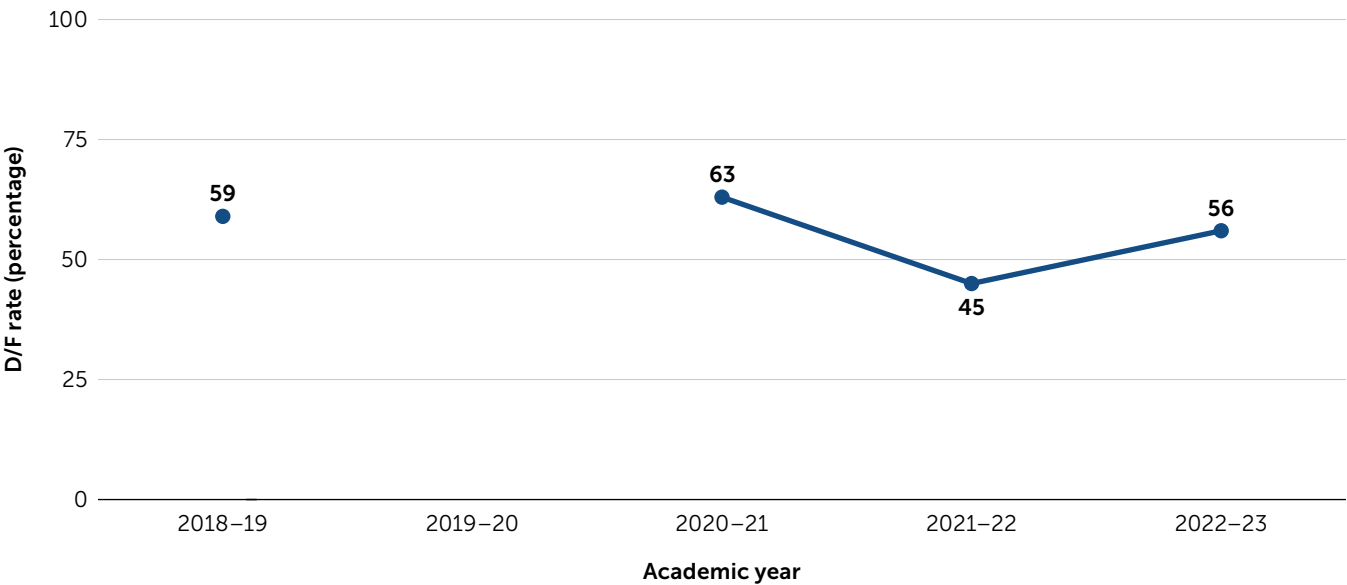
It's become a whole big thing this year where they're framing it as a relationship-building thing and they're trying to get teachers to really connect with kids during it and both use the academic as an excuse to have personal conversations and also have the personal as a reason to get kids to come do academics.

In this example, leaders at Magnolia adapted concepts from BTSC's change package that elsewhere were small-scale interventions affecting only small groups of students. Like at Pine, when these ideas were pulled into grade-wide processes and structures, they served as a vehicle for improving the way the schools engaged with their students. Finally, the team began investigating the variation in grading processes and the proportion of Ds and Fs given out by ninth-grade teachers, with the intention of ultimately agreeing on shared practices that would make it easier for students to understand and meet teacher expectations across their classes.

Magnolia's progression through the various zones of traction exemplifies a pattern of BTSC initiation that other school sites experienced, but they were able to turn that initial experience around. During Year 1 of implementation, Magnolia remained in Zone A, with little ownership or depth to the work; the CORE coach struggled to get the team constituted, and the work did not progress past a superficial level. During Year 2 of the initiative, work hovered between Zones A and C. A few support staff received personalized support from the coach on parts of their roles, but BTSC did not have access to the ninth-grade house or leadership structure, so it was impossible to conduct impactful tests or institute widespread changes if they succeeded. It was only in Year 3, when the former BTSC co-lead was promoted from assistant principal to principal and the ninth-grade team leaders became engaged with BTSC, that a path to scaling BTSC work to the whole of the ninth grade opened. As noted in the discussion of Pine, using the house structure (or cohorting students) became a powerful tool to facilitate implementation of the drivers at Magnolia. This spread and scope put the school into Zone D by the fourth year of the initiative.

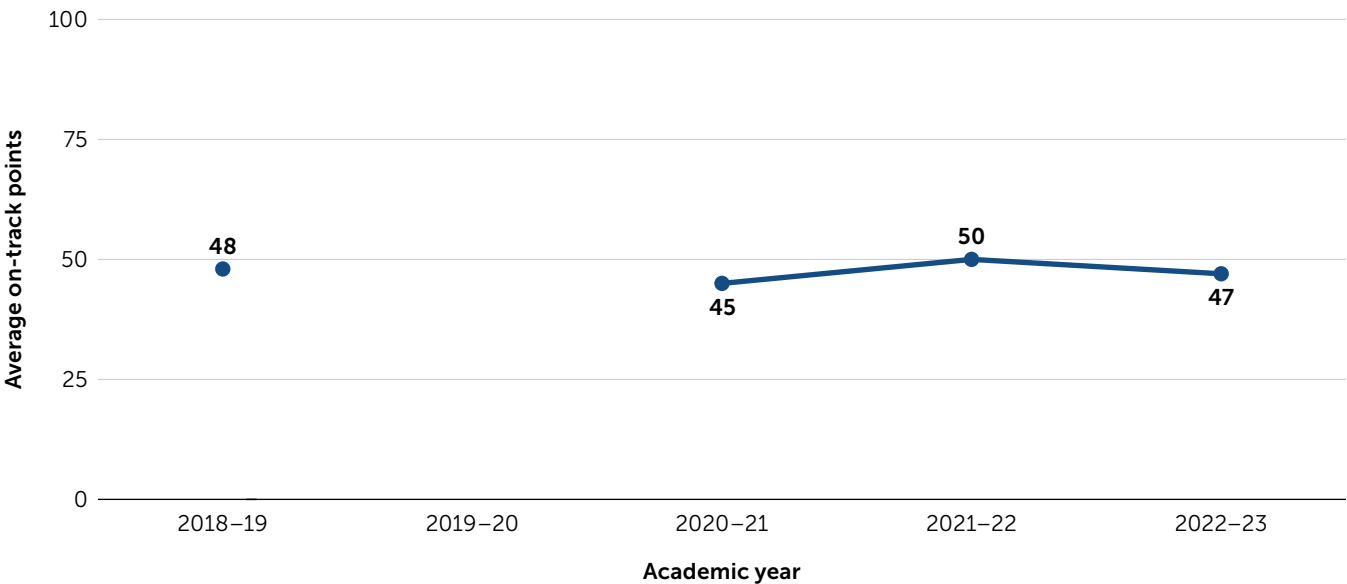
Looking at student outcomes over this period, we see that the year Magnolia joined BTSC (in April 2019, making 2018–19 essentially a baseline, or pre-intervention, year for student outcomes), 59 percent of ninth graders at Magnolia received at least one D or F in the fall and/or spring semester (Figure 7). Mirroring statewide patterns, the D/F rate increased in 2020–21, when the school offered mostly remote learning. In 2021–22, there was a sharp decline in the D/F rate, but it increased to 56 percent in 2022–23.

Figure 7. D/F Rate at Magnolia, 2018–19 Through 2022–23



Not surprisingly, because student grades play an important role in calculating on-track points, we see some similarities in the trends over time in ninth-grade on-track points at Magnolia (though the trend lines are inverted because, generally speaking, better student grades lead to declining D/F rates and improving on-track points; Figure 8).

Figure 8. Average Ninth-Grade On-Track Points at Magnolia, 2018–19 Through 2022–23



These trends should not be interpreted as causal, meaning we are not claiming that BTSC is responsible for any improvement we might or might not see. First, these data are for different cohorts of students, and interviews suggest that the nature of the student population may have shifted over the years that Magnolia was in BTSC; any variation from a population shift would only compound the typical year-to-year variation in metrics like these due to the fact that every year different students enter ninth grade. Perhaps more important, BTSC was never expected to be the only way a school was changing: staff changed year to year, and there were always other simultaneous attempts to support students better. Last, these data may reflect broader changes in the population. Overall, these data show why staff were excited about their success the preceding year when we interviewed them in spring 2023. Given the staff's focus on results and willingness to adapt ideas to support student success, they have the potential to modify their existing approaches or try new ones to attain the results they want consistently in the future.

Aspen High School. Aspen is the smallest of the three high schools profiled in these vignettes, with only about 80 students in the freshman class and a teaching staff of about 20 across Grades 9–12. Aspen was only in its second year of operation (with students enrolled in ninth and 10th grades and a plan to grow by one grade each year) when they joined BTSC's second cohort. Despite the school's being open for only 2 years, many members of the original staff and teachers had already turned over, and the young (in both overall experience and tenure at the school) teaching staff's primary focus was on building their classroom management and pedagogical skills, which BTSC was not designed to support.

As at Magnolia, the CORE coach at Aspen initially had to build relational trust and contend with a disengaged administration. During the first year, the principal was on the BTSC team but also named two co-leads, who did not feel like they had been given the authority to lead the work. They also struggled to find a regular meeting time for their Freshman Success team. They tried two different meeting times. One conflicted with a grade-level meeting that half the team was required to attend; the other was outside the contract day, and some team members had additional responsibilities that conflicted with that option. As a result, the team struggled to gain traction during the first year.

In Year 2 of the initiative, a dynamic teacher who had strong support from the principal moved from co-lead to sole lead of the team. This lead was able to institute a robust participatory model in which team members made decisions collaboratively and shared responsibilities like taking notes. The lead felt deep ownership of the work, a sentiment that she passed along to team members by folding collaborative analysis of data that resulted in joint decisions into the team meetings. When the five-driver model was introduced, it helped catalyze work at this small school. As one team member explained: "We have a small school ... [and] we were essentially able to give each person a driver title or role. ... It's been really great because everyone has a role." In addition, as the school added a new grade in their second and third years in BTSC, a teacher from the original BTSC team moved from the ninth grade to the 11th grade, which helped spread BTSC ideas beyond the ninth grade.

While at Pine the coach's prior knowledge of the system helped to quickly "fit" BTSC to the institution and at Magnolia the coach had to spend time learning how to integrate the initiative to the school, a different process of accommodation occurred at Aspen. Because Aspen was such a new institution with only nascent institutional practices, there was not yet an established "how we do things here." Because the new team lead felt ownership of BTSC work, they were essentially able to make the BTSC meeting the meeting where all important ninth-grade business was done.

As ownership and sustainability grew, so too did the depth of the changes. Combining work from the adult teaming, relationships, and grading drivers, Aspen used the meetings for regular monitoring of student progress. This monitoring was paired with a three-tiered intervention model that provided targeted Tier 2 and Tier 3 supports for some students (e.g., coordinating with families or connecting students to adult mentors) and Tier 1 opportunities for tutoring and making up work for all students. This combined monitoring and support system provided students with many opportunities for connection and academic assistance. And because the school had only 20 teachers, it also spread BTSC work to most of Aspen's staff, with strong consistency of implementation in ninth grade and some spread of basic practices evident in later grades. Over time, because of the small teaching staff and the embeddedness of BTSC, the initiative has become how the school—not just the ninth grade—does things.

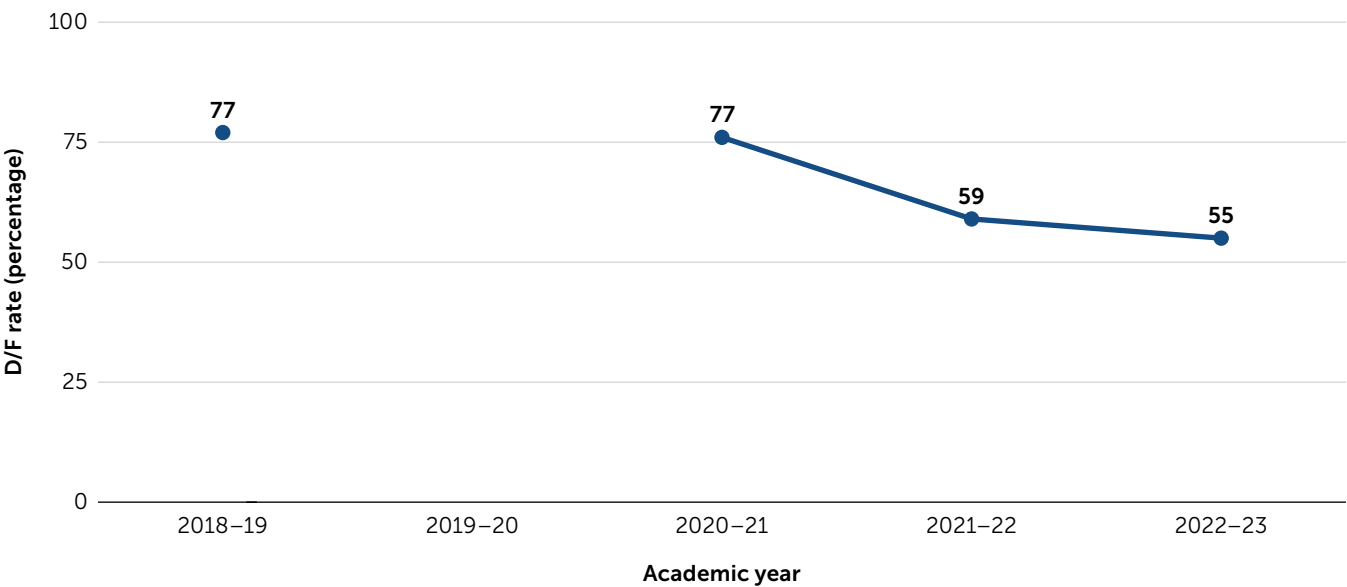
In fall 2021, Aspen got a new principal, who came in very supportive of BTSC because she was involved with the program at her prior school, where she was the assistant principal. She was a background champion of the work, continuing to elevate the BTSC team lead's leadership beyond BTSC to other elements of the school. In spring 2022, during Aspen's third year of BTSC, the team's lead started work on her administrator credential and went on leave in the spring. While her temporary absence was somewhat magnified by the size of her role in such a small school, the lead used her absence to focus on building and transferring leadership to other members of the team. She said:

My principal has always been telling me ... "You are most likely not going to be here, if you go the admin route. So, at some point, you're going to need to pass the torch." And ... I understand that ... I mean, I'm always trying to share what I've been taught from CORE and from BTSC. So hence why I'm trying to expand it out through other grade levels.

Here we see the team lead attending to issues of leadership transition, spread, and sustainability, considering both how BTSC will persist in her absence and how its work might become more deeply embedded in the regular processes of the school. The story of Aspen primarily highlights the importance of leadership, the fact that fewer people need to be involved to reach a tipping point of involvement in a small school environment, and that when ideas are embedded into structures and processes, that can facilitate their spread and sustainability. It also exemplifies how BTSC provided valuable teacher leadership opportunities.

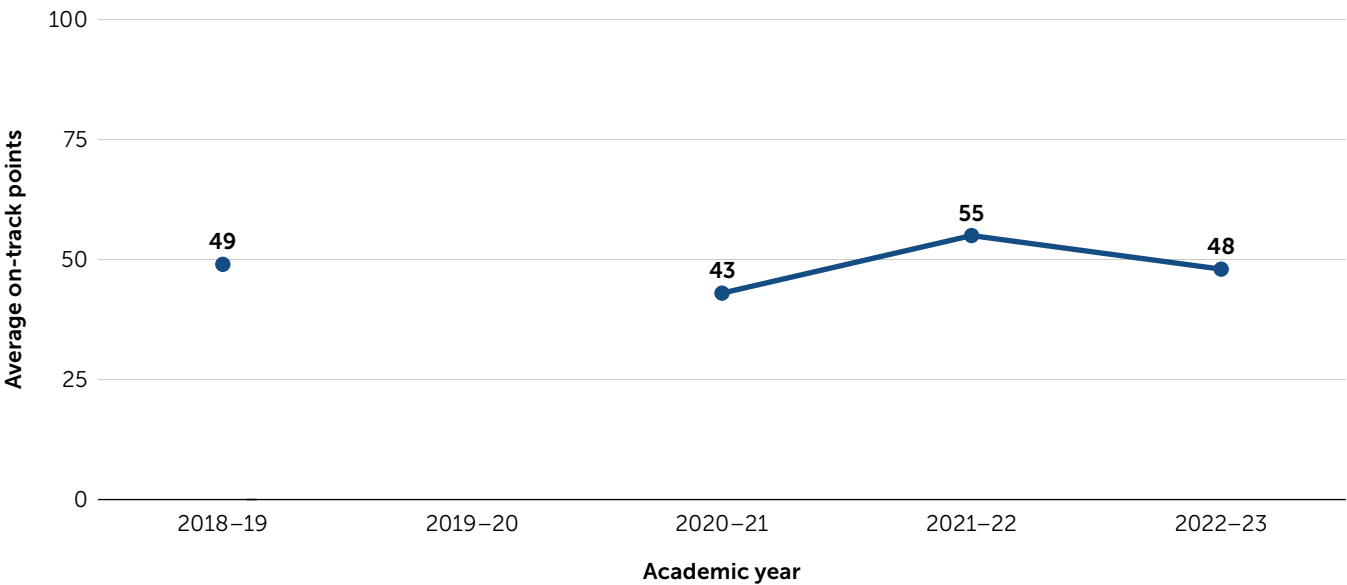
Examining student outcomes over this period, we see trends that match those described in interviews (Figure 9). Both when Aspen entered BTSC in spring 2019 and in the year of hybrid learning, 77 percent of students received at least one D or F over the course of the school year. The D/F rate dropped substantially the following year and dipped even further in 2022–23.

Figure 9. D/F Rate at Aspen, 2018–19 Through 2022–23



The trend in on-track points is slightly different from the trend in D/F rate (Figure 10). At baseline in 2018–19, the average ninth grader at Aspen had 49 BTSC ninth-grade on-track points. In 2020–21, the average number of on-track points was lower, likely an indicator of students’ struggles in the aftermath of the pandemic. Inverse to the D/F rate, there was a notable increase in on-track points from 2020–21 to 2021–22, but while the number of students with one or more D/F was even lower in 2022–23 than in 2021–22, the average number of on-track metric points dipped a bit from 2021–22 to 2022–23.

Figure 10. Average Ninth-Grade On-Track Points at Aspen, 2018–19 Through 2022–23



While this may initially seem counterintuitive given that grades are a major contributing factor to on-track points, there are a couple of possible explanations. First, other factors that are aggregated by the BTSC on-track points, such as course difficulty or attendance, could account for the divergent trends. Given the way differences in the D/F rate versus on-track metrics are calculated, however, it is more likely that even though the proportion of students getting at least one D/F rate declined over those last 2 years, the number of Ds and Fs received by those students who had a D/F in at least one class went up. This would lead to a decrease in the number of points students received for the individual classes they were enrolled in, for their overall GPA, and possibly for credits earned, which in turn would lower their on-track points enough to bring the overall average down. We reiterate the caution about not interpreting these data as showing the effectiveness of BTSC because it is cohort data and, as Table 4 shows, the number of students changed substantially from year to year at both Aspen and Magnolia, which provides clear evidence that the cohorts are not identical to each other.

Table 4. Year-to-Year Variation in Cohort Size for Aspen and Magnolia High Schools

School	Number of students				
	2018–19	2019–20	2020–21	2021–22	2022–23
Aspen	155	Hold harmless	92	64	112
Magnolia	201	Hold harmless	271	318	312

Conclusion. The vignettes about Pine, Magnolia, and Aspen High Schools offer examples of how BTSC gained traction in three schools. At each school, relational trust and shared understanding were essential for building ownership among school site personnel and for beginning the work with any sort of depth. It was only when team members at each school felt empowered to take on BTSC and “make it their own” that small tests of significant changes could progress in earnest. School leaders—both administrators and teacher leaders—played critical roles in BTSC at all three schools. Even when administrators were not actively involved in BTSC, their support was critical.

Additionally, the central change ideas that these teams tried differed sharply from those that were the focus of work in many BTSC schools. While many BTSC teams tested ideas that affected one student at a time, BTSC teams in all three of these schools adapted BTSC change ideas in ways that integrated them into their processes for progress monitoring and supporting an entire grade level or cohort of students (with the plan to expand to the entire grade). The initial tests were still “small enough to learn,” with teams typically trying a new protocol or process at one meeting and then adapting it for later meetings until they figured out how to integrate the change successfully into their ongoing structures and processes.

Student Outcomes

In addition to collecting qualitative data to understand how CORE designed and implemented BTSC and how much traction it got in participating schools, PACE collected and analyzed data on student outcomes. BTSC was implemented during the tumultuous time of the COVID-19 pandemic, when most districts with schools in BTSC largely shut down for more than 2 months during spring 2020 and then offered almost solely virtual instruction during the 2020–21 school year. When in-person school resumed in 2021–22, the system had received a nasty shock and everyone struggled to recover. While it is often true that the best way to understand progress of an initiative is to examine progress relative to what happened in similar places that were not participating (i.e., the counterfactual), that is even more true under circumstances where there are sharp changes in trends across an entire population, like those caused by the COVID-19 pandemic.

This section answers the question: What is the effect of school participation in BTSC on ninth graders’ BTSC on-track metric points in 2022–23 (the final year for which we have data)? The outcome measure is the number of points—on a scale from 0 to 100—on CORE’s BTSC On-Track Metric 2.0. We also include analyses of the D/F rate for 2022–23 and on-track metric points for 2021–22 to examine whether the findings are consistent across an additional measure and a different year. The remainder of this section describes the comparison group and how we selected it, presents descriptive data and trends for BTSC and comparison schools, and then presents a rationale for and results from our regression analyses.

Selecting a Comparison Group and the Resulting Analytic Sample

Our analysis of the outcomes of BTSC on ninth-grade on-track focuses on the 29 BTSC schools that are in the CORE districts and were participating in BTSC in 2022–23. This removes 11 BTSC schools from the analytic sample for two reasons. We decided to exclude the five schools outside the CORE districts for both conceptual and pragmatic reasons. Conceptually, the schools outside the CORE districts received ongoing coaching and guidance that was at times quite different from that received by schools supported by CORE coaches. In our interviews with external coaches, who were employed by their COEs, we learned that the vast majority of those coach’s roles were outside of BTSC and, as a result, they had some to little knowledge of the change ideas being promoted by BTSC. This should not be interpreted as a judgment of those coaches. In some cases, they brought substantial expertise and preexisting relationships with their schools from ongoing support roles like helping those schools with *differentiated assistance* (a California term for required technical assistance for struggling schools), as was illustrated by the previous discussion of Pine High School. Even when what these coaches provided appeared to be high quality, it is unclear the extent to which these schools got “BTSC” as designed, as opposed to a hybrid of BTSC with strong features of other programs. Pragmatically, doing the analysis and selecting appropriate matched high schools would have required a substantial amount of effort to gain access to those districts’ data. Given that the intervention those schools received was unlikely to be a good test of the efficacy of BTSC as it was intended, we made the decision to focus our analysis solely on schools in the CORE districts.

For any year of outcome data, we included only schools participating in BTSC during that year in the analytic sample. For 2022–23, this excludes the seven schools that dropped out of BTSC prior to that time. We recognize the value of leaving schools that attrit in an analysis, as is the norm in the context of a randomized controlled trial, when the research question focuses on the effects of being assigned to a treatment group. Since our research question is about the impact of BTSC on participating schools, the best approach is to remove schools when they stop participating.

Possible comparison schools were chosen out of a pool of schools from six CORE districts, which all participated in BTSC and had student-level data for the years of interest for our study. Schools eligible to be selected into the comparison group had the following characteristics:

- were public high schools,
- remained open for the duration of BTSC, and
- had not ever participated in BTSC.

There were 103 potential comparison schools that met these criteria, which we could match to 36 schools in the CORE districts that ever participated in BTSC. We selected the comparison group for the full set of schools that were ever in BTSC to enable a constant sample with which we could analyze outcomes for 2022–23 and earlier years (to see if the results were consistent across years).

To find matches for BTSC schools, we sought a sample that balanced organizational characteristics and the population of incoming ninth-grade students. At an organizational level, the ideal match would be a school of similar size, of similar governance (e.g., traditional and public), and in the same district. The districts participating in BTSC vary in size (and, therefore, the number of potential comparison schools vary), and in some districts, the majority of traditional public high schools participated in BTSC. As a result, it was clear that we could not find appropriate matches for all BTSC schools within their own districts. Relatedly, in all but two of the districts (which were large and had a diverse set of schools not participating in BTSC), finding a within-district match might have come at some degree of cost to other characteristics we wanted to have match (e.g., size and student characteristics) because of the relatively high proportion of high schools participating.

To optimize for a comparison group that contained as many same-district matches as possible without overly compromising other school characteristics, we turned to the *What Works Clearinghouse Procedures and Standards Handbook* (2022) for guidance on baseline equivalence. WWC considers samples equivalent if, on key baseline variables, the effect size for their difference is less than 0.25 (with appropriate adjustments for the clustering of students within schools). Given the importance of having close matches and the general similarities among districts in the study (all large, urban districts in California), we set a slightly more conservative bar of an effect size of less than 0.2 for the sample differences in means at baseline for each of the following variables (for each of which we created a 3-year average pre-BTSC):

- ninth-grade size,
- ninth-grade percentage Hispanic,
- ninth-grade percentage African American,
- ninth-grade percentage free or reduced-price lunch (FRPL),
- incoming ninth graders' eighth-grade math SBAC, and
- incoming ninth graders' eighth-grade English language arts (ELA) SBAC.

We began by selecting a set of matches that were constrained by district, using optimal matching and not allowing any comparison school to match with more than one treatment school (i.e., no replacement). Because of the proportion of BTSC schools in some districts, solely using a within-district approach led to 10 BTSC schools having no comparison. In districts where there were insufficient comparison schools available for a within-district match, we allowed the schools with the worst within-district match to match with a school from the broader set of

data until all schools had an initial match, with a preference for within-district matches where possible.³ We then assessed the effect size for the differences in each of the variables above for the sample of treatment and comparison schools. When the overall sample equivalence was less than 0.2 for each measure, our matching process was complete.

Table 5 shows the mean values for the matching variables, and Table 6 shows the distribution of BTSC and comparison schools across the CORE districts.

Table 5. Values of Matching Variables for BTSC and Comparison Sample

Schools	Average <i>N</i> of ninth graders	Percentage FRPL	SBAC math	SBAC ELA	Percentage African American	Percentage Hispanic
BTSC	444	77	2,482	2,511	13	70
Comparison	491	77	2,490	2,518	12	71
Effect size for difference	-0.18	0.02	-0.17	-0.16	0.06	-0.02

These results met our criteria for baseline equivalence.

Table 6. Distribution of BTSC and Comparison Schools Across the CORE Districts for the Analytic Sample

District	<i>N</i> BTSC schools	<i>N</i> comparison schools
District 1	3	1
District 2	4	2
District 3	4	5
District 4	5	0
District 5	6	2
District 6	7	19

³ We operationalized “insufficient” comparison schools available for a within-district match in two ways. First, in some districts, more than half the schools participated in BTSC, making it impossible to have unique, within-district matches for all schools. In that case, we matched as many schools within the district as possible, retaining the best of the within-district matches and selecting the best nearest-neighbor matches from the potential comparison group for the remaining schools. Using this approach yielded a potential comparison group where the effect sizes for the difference between treatment and comparison group exceeded the WWC-recommended threshold of 0.25 for two variables (eighth-grade test scores in ELA and math). We then removed “low-quality” within-district matches to reach the overall threshold of less than 0.25 for the effect size of difference between the treatment and comparison groups on each matching variable. The resulting sample had an effect size of less than 0.2 for each matching variable.

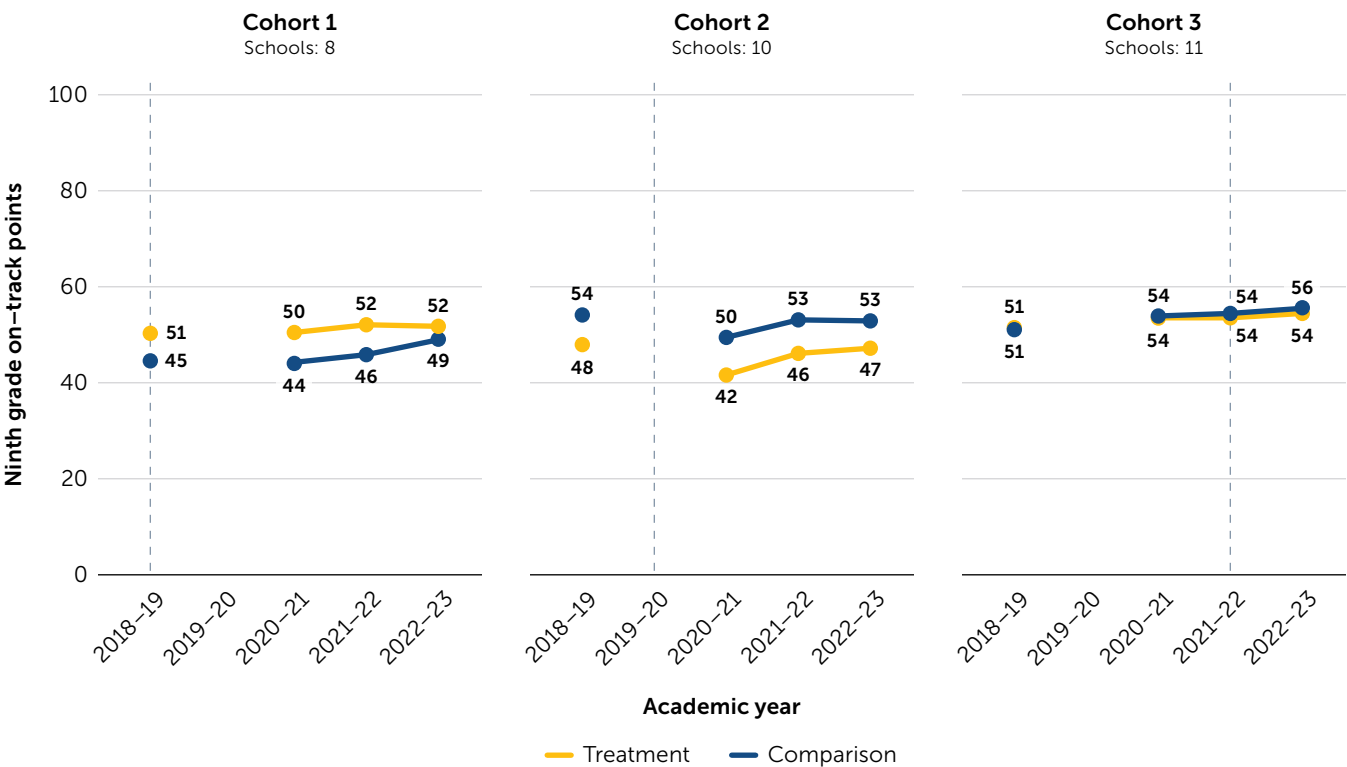
The district that had the highest proportion of its high schools participating in BTSC (District 4) did not have any comparable within-district matches, and District 6 provided many more matches than participating schools. While the imbalance in districts was not ideal, it enabled us to find matching schools that served a similar population of incoming students, which provides an appropriate counterfactual.

Descriptive Statistics on BTSC and Comparison Schools

Appendix A presents means and standard deviations for the covariates used in the regression analyses that examined the effect of BTSC on ninth-grade BTSC on-track metric points for 2021–22 and 2022–23 (the 2 years for which we estimated outcomes).

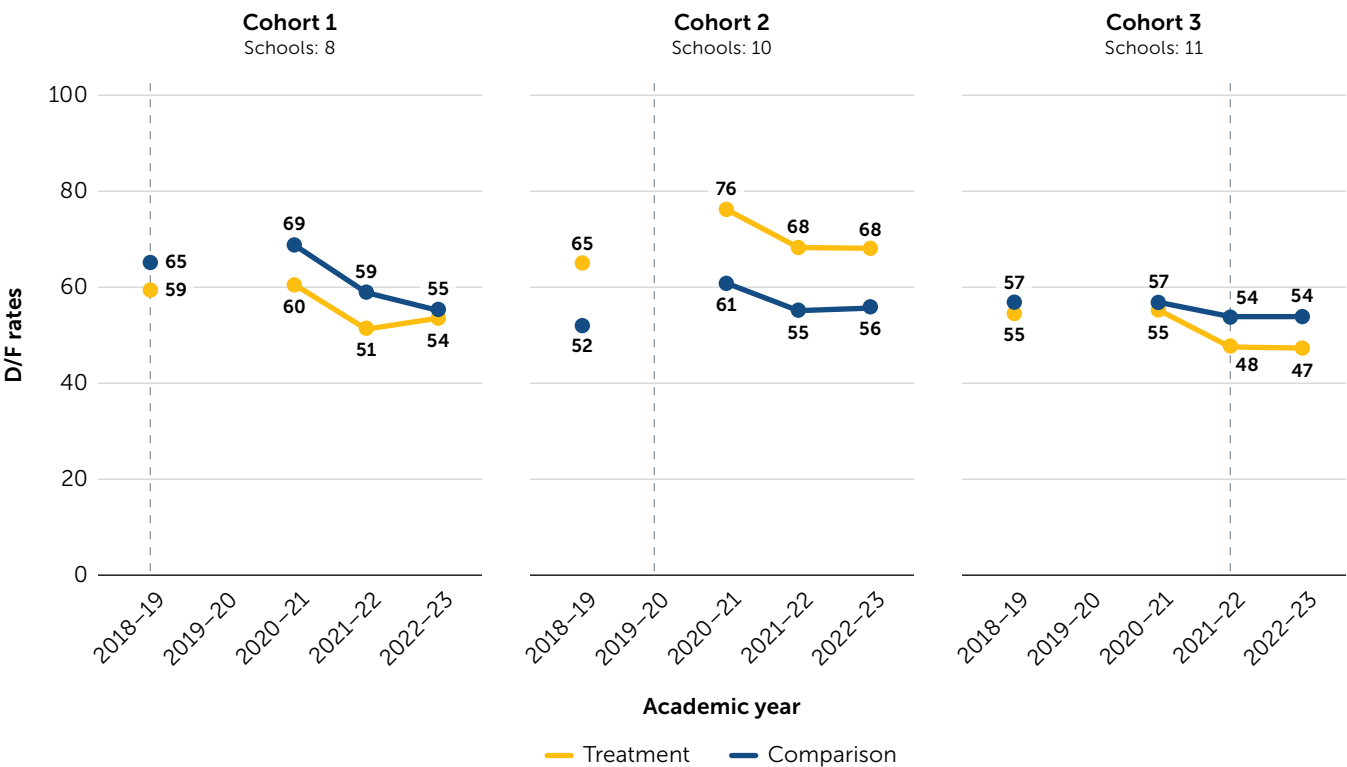
We next present trends for BTSC and comparison schools of two outcomes of interest for our 2022–23 analytic sample for the years from 2018–19 through 2022–23. We did not calculate outcomes for 2019–20 due to missing data and the way that the state’s “hold harmless” policy affected grades. Figure 11 shows the average student on-track points in all BTSC schools and the average student on-track points in all comparison schools for all three cohorts of BTSC schools and their comparisons. The averages for these graphs are calculated by summing all possible BTSC points for students in BTSC and comparison schools and dividing by the total number of students with BTSC points. The dotted vertical line indicates when the cohorts joined BTSC.

Figure 11. Average BTSC On-Track Metric Points in BTSC Schools Versus Comparison Schools



We also examined the trends in the D/F rate for BTSC and comparison schools during the same period because grades were the focus of one of the five drivers and might be more sensitive to changes from BTSC. Figure 12 shows the D/F rate for ninth graders (defined as the proportion of ninth-grade students who received one or more Ds or Fs in A–G courses) in the three BTSC cohorts. The D/F rates shown in Figure 12 are calculated by summing the total number of students who received a D or F in an A–G course divided by the number of students who took an A–G course.

Figure 12. Average School D/F Rates in BTSC Schools Versus Comparison Schools



It is important to realize that because these data represent different cohorts of students and because there typically is some degree of variation in metrics like the BTSC on-track metric or the D/F rate year upon year, small changes should not be seen as evidence of improvement (or backsliding on apparent improvement). Even if there were improvement, this would not provide evidence that BTSC *caused* the improvement because of the many factors unaccounted for in this analysis. The more reasonable inference to make from both these graphs is that there appears to be a slow but steady recovery in student outcomes after the pandemic, which is a comparable trend to those documented in other outcome measures across the state (Hough & Chavez, 2022).

Results of the Regression Analysis

The goal of the regression analysis was to estimate the impact of BTSC on student outcomes relative to a comparison group and adjusted for differences in student characteristics. The ideal way to do this type of analysis would be with random assignment, which guarantees that there is no bias in the estimate of the difference between a treatment and a comparison group. We carefully selected a comparison group as the best way to estimate the counterfactual under the constraints of our design. Following best practices in the field, we selected a confirmatory regression model in advance, which we then used to test the null hypothesis “there is no difference between BTSC and comparison schools” in ninth-grade students’ on-track metric points in 2022–23. We also ran a series of exploratory analyses that examined a range of ways of specifying the model in terms of the covariate adjustments (for reasons described below) and that examined two additional but related outcome measures: (a) ninth-grade students’ on-track metric points in 2021–22, and (b) D/F rates in 2022–23. If we saw similar results in the exploratory analyses as we did in the confirmatory analysis, we would know that the results of the confirmatory analysis were robust to model specification.

It was important to run multiple model specifications because of some less-than-ideal characteristics of some of the variables in our models (as is typical in these types of secondary data sets). Conceptually, we were seeking a set of covariates that supplement the school-matching process by enabling us to adjust statistically for differences in incoming student performance and other characteristics. The best variable in that regard was the BTSC eighth-grade on-track points, but due to gaps in the data necessary to calculate that measure, 20 percent of students were missing that variable, and in 10 of the 58 schools in the analytic sample for 2022–23, more than 50 percent of students were missing that variable. Appendix A shows the potential covariates, explains variables we created to address some of the issues with missing data, and presents the percentage of students missing each variable in 2022–23 along with the correlation between the variable and BTSC ninth-grade on-track metric points.

After examining the potential covariates, we selected a confirmatory model that accounted for the nesting of students within schools and included student-level covariates for:

- eighth-grade BTSC on-track metric points;
- race/ethnicity (a composite categorical variable indicating if a student indicated that they were Hispanic and, if not, what race they identified as);
- English learner (EL) status (a categorical variable showing if students had ever been identified as EL and, if so, their assessed English proficiency); and
- parental education (a categorical variable for the highest level of parental educational attainment from high school through postgraduate, including if the student declined to state parental education).

The model also included an indicator for whether a school was in BTSC and the cohort of the BTSC or comparison school (with matched comparison schools assigned to the same cohort as their BTSC pair). The independent variable in the confirmatory analysis was ninth-grade BTSC on-track points. The resulting model confirmatory was:

$$90T_{ij} = \beta_{00} + \beta_{01}BTSC_j + \beta_{02}Cohort_j + \beta_{03}80T_{ij} + \beta_{04}EthRace_{ij} + \beta_{05}EL_{ij} + \beta_{06}ParentEd_{ij} + u_{0j} + e_{ij}$$

For this model, if the coefficient for BTSC was positive and statistically significant, we would state that there was a positive difference between BTSC and the comparison schools, and we would infer that BTSC was positively related to improved BTSC ninth-grade on-track metric points; if the coefficient for BTSC was negative and statistically significant, we would state that there was a negative difference between BTSC and the comparison schools, and we would infer that BTSC was related to lower BTSC ninth-grade on-track metric points; and if the coefficient for BTSC was not statistically significant, we would not be able to claim any effect (positive or negative) of BTSC on student outcomes. Table 7 presents the results of the model.

Table 7. Results of the Confirmatory Model

Variable	Coefficient	P value
Ninth-grade on-track points		
Treatment	-0.532	0.575
Cohort		
1	(base)	
2	-1.484	0.218
3	0.204	0.863
Eighth-grade on-track points	0.827***	0.000
White	(base)	
Asian	-0.465	0.387
American Indian	-7.144**	0.003
Pacific Islander	0.108	0.937
Filipino	-0.694	0.463

Variable	Coefficient	P value
African American	-2.549***	0.000
Multiracial	-1.161	0.103
Hispanic	-2.033***	0.000
EL status		
EL	-0.346	0.315
EO	(base)	
IFEP	1.325*	0.021
RFEP	1.739***	0.000
Parental highest education		
Graduate degree or higher	2.565***	0.000
College graduate	1.775***	0.000
Some college or associate's degree	1.104**	0.003
High school graduate	(base)	
Not a high school graduate	-1.185***	0.000
Decline to state	0.36	0.230
Constant	11.280***	0.000
Ins1_1_1		
Constant	1.236***	
Insig_e		
Constant	2.631***	
Statistics		
N	19,401	
Converged	1	

Note. EL = English learner; EO = English only; IFEP = Initial Fluent English Proficient; RFEP = Reclassified Fluent English Proficient.

*** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.10$.

Based on these results, we find no evidence of an effect of BTSC (positive or negative) on ninth-grade on-track metric points.

To test whether our results were robust for alternative model specification, we ran models with different versions of the covariates and found consistent results. We also ran exploratory models examining the D/F rate in 2022–23 and BTSC ninth-grade on-track metric points for 2021–22. The results are presented in full in Appendix A. Across all model specifications and all three outcome measures, we find no evidence of an effect of BTSC on student outcomes.

Summary and Conclusions

CORE attempted to take the idea of an IHI Breakthrough Collaborative and apply it to education with the Breakthrough Series Collaborative to improve ninth-grade on-track rates in low-performing, urban high schools. Here we provide several key lessons from BTSC that could inform ongoing development of BTSC and other efforts to use a Breakthrough Series model to improve student outcomes.

After closely tracking this initiative for its first 4 years, PACE finds that:

1. BTSC diverged fundamentally from an IHI Breakthrough Series Collaborative because it did not rest on strong causal evidence and an associated set of measures for schools to use in tracking whether their application of evidence-based changes was generating the expected improvements in student outcomes. We hypothesize that while the weaker available evidence base in education is partially responsible, the greater cause is that CORE never found someone to fill the role of a faculty chair for BTSC. While CORE staff did consult experts, the overall design of BTSC was not driven by someone with the extent of knowledge of ninth-grade on-track research and its application, which IHI names as undergirding a Breakthrough Series. This had two major implications. First, a strong faculty chair would have ensured that the change package was based on rigorous research that showed a clear gap between what we know works for improving ninth-grade on-track and typical practice. Additionally, the faculty chair would have created a measurement plan to assess whether the changes that teams tried were an improvement. These divergences are fundamental enough that we do not believe that BTSC, as it was enacted, is a strong test of the potential for the Breakthrough Series approach in education.
2. CORE's recruitment process also did not closely mirror a Breakthrough Series approach. It did not successfully engage leaders of participating schools as champions of the work who were committed to supporting teams' testing and then spreading changes in their systems. This is somewhat understandable in the context of the short timeline of the grant and the context of multiple demands on secondary schools. But it was

nonetheless deeply problematic for the effectiveness of BTSC because it led to deviations from central components of research-based on-track work and created substantial barriers to scale in some schools.

3. While CORE used some aspects of improvement science in BTSC, many schools consistently fell short of some of the core tenets of improvement science (and even continuous improvement more broadly) in that they did not consistently: (a) examine the current system, (b) test changes to system structures and processes, (c) gather evidence from those tests to see if changes were an improvement, and (d) spread changes in the system based on evidence that the changes were an improvement. Additionally, CORE's search for an improved ninth-grade on-track measure meant the network had no measure to use in an aim for 2 years. CORE did not anchor the network around the aim by regularly examining network progress and using the measure to identify positive outliers for the network to learn from. As a result, we also do not believe that BTSC is a test of the potential for improvement science to improve student outcomes.
4. CORE's change package focused heavily on ideas around changing adult mindsets. While we do not disagree with CORE's stance that some educators could better support students if they had more asset-based mindsets, improvement science is designed to support system changes under the assumption that individuals' mindsets are themselves an outcome of the systems that individuals are in. In schools where BTSC gained traction, BTSC teams adapted mindset-change ideas into system changes—for example, embedding the KidTalk Protocol in regular progress-monitoring meetings held by a small group of teachers teaching a cohort of students. Putting the idea into the system made it possible for the teachers to track all students' progress and have asset-based conversations about all struggling students. When those same change ideas did not get embedded in systems, they affected only the few adults who chose to take them on and a few of their selected students.
5. Because BTSC focused its measurement effort on innovative outcome measures, it did not provide schools sufficient support for tracking whether the changes they were trying were an improvement. Absent evidence that changes affected students, teams had only anecdotal evidence to use when they wanted to convince others to spread practices supported by BTSC.
6. Few schools achieved traction with BTSC. Those that did had engaged leaders who ensured that there was time for BTSC teams to meet and who supported adapting existing structures or processes—or creating new ones—that embedded central ideas from BTSC into the ninth-grade system. These schools typically monitored ninth-grade students' progress, tried to connect students with supportive interventions if they needed help, and monitored data to see if the things they were trying made a difference in student outcomes.

7. We do not find evidence that BTSC had any impact on student outcomes on average. This is not unexpected given the changes tried and the amount of traction in the average BTSC school. There is mixed evidence about whether some individual schools that tested system changes and spread them were able to have a positive effect on student outcomes, which suggests that BTSC's approach could be adapted to improve its efficacy.

As we look beyond this evaluation to the broader education sector and other attempts to improve student outcomes, we are left with two closing thoughts.

You can have anything, but you can't have everything. This saying is colloquial, but the central lesson is one that education seems to struggle to learn. All parties in BTSC—CORE, their participating districts, and the participating schools—underestimated the opportunity costs of doing BTSC the way they did it. Their decisions are worth revisiting so as not to repeat them in the future.

Districts and schools often acted as though BTSC were “free,” joining in hopes that it would improve on-track rates with only the donation of a few educators' time. This did not work in most schools. It also undervalued those educators' time, which they could have used for doing other things on their campuses, investing in instruction, or fortifying their resilience during the particularly challenging time our study took place. Engaged team members were typically very dedicated to their students and put their efforts—as best as they could, given conflicting demands for their time—into BTSC. The districts and schools that joined BTSC without providing the fundamental conditions of collaborative meeting time and prioritization of this effort made BTSC an uphill battle where individual educators toiled hard while lacking the strong administrator buy-in typically needed to make important changes to school systems. This is yet another example in education where schools, especially those that are struggling, feel pressured to take on more initiatives than they can do well. The result is a diffuse impact, overworked educators, and changes that are unlikely to be sustained in most schools. Districts and schools need to stop underestimating the costs of engaging in too many diffuse initiatives and instead focus their efforts on sustaining only a few top-priority strategies until those come to fruition.

On CORE's side, the biggest opportunity-cost questions arose around measurement. Highly precise measures are critically important when they are being used for accountability and when the consequences of error are high for the individuals involved. Measurement for learning and improvement is very different from measurement for accountability. The result of CORE's investment were (a) an expanded view (beyond the basic Chicago estimate of on-track based on students' Ds or Fs or low credit accumulation) of the number of students who might not be on track for meeting A–G requirements and (b) a quantifiable measurement of students' developmental relationships (which complemented what teachers knew from their personal relationships with students). The BTSC metric is likely more precise than just flagging students

based on grades and credit accumulation, and educators definitely appreciated the insights gained from the DRS. But how did BTSC capitalize on the improved precision in identifying students who might not be on track? How did schools improve systems for supporting students who were not experiencing strong developmental relationships? Our data do not provide evidence of consistent benefits when we answer these questions, but our data do illuminate the trade-offs. The cost of this precision and quantification was that CORE did not have resources to invest into building schools' capacity to use the data that already existed in school information systems for grades, attendance, and discipline. These in-school data could have given real-time information to schools about student progress. CORE also could have taken the data capacity they invested in innovative measures and dashboards and directed it towards building educator capacity to identify practical measures to assess whether changes they were testing were an improvement.

The story of BTSC is not over. CORE continued working with the BTSC schools described in this report during the 2023–24 school year, and they may keep working with some of them, especially those in Cohort 3, after that. They also have funding to serve schools that were in the control group for a randomized controlled trial that included BTSC's Cohort 3 schools and have recruited additional schools that will combine to form at least one more cohort. Every year of BTSC was more successful (at least in terms of engaging educators in the work) than prior years, and CORE improved their own internal systems for running BTSC each year. For example, by the end of 2022–23, CORE started tracking which ideas teams were testing and which they spread. CORE could use these data about school implementation paired with data about D and F rates to refine their change package and the way their coaches work with schools. It is our hope that, in addition to maintaining the strengths of the BTSC approach that we tried to highlight in this report, CORE will attend to aspects of the program that could be improved. If they do both things, CORE has an important opportunity to help some of California's districts learn how to improve student outcomes in their secondary schools.

References

- Balfanz, R., & Byrne, V. (2019). Early warning indicators and intervention systems: State of the field. In J. A. Fredericks, A. L. Reschly, & S. L. Christenson (Eds.), *Handbook of student engagement interventions* (pp. 45–55). Academic Press. doi.org/10.1016/B978-0-12-813413-9.00004-8
- 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.
- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3–12. journals.sagepub.com/doi/10.3102/0013189X032006003
- CORE Districts. (2018, October 3). *Breakthrough success community expert convening pre-reading memo*. docs.google.com/document/d/1HNb8VFCZLgaht_CPSnACSBnnCnyYSnX5nWkfMtPpPiE
- Easton, J. Q., Johnson, E., & Sartain, L. (2017, September). *The predictive power of ninth-grade GPA* [Research report]. University of Chicago Consortium on School Research. consortium.uchicago.edu/sites/default/files/2018-10/Predictive%20Power%20of%20Ninth-Grade-Sept%202017-Consortium.pdf
- Feldman, J. (2018). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin Press.
- 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-with-continuous-improvement
- Gallegos, E., & Willis, D. J. (2024, February 12). *Most California high school seniors shut out of even applying to the state's universities*. EdSource. edsources.org/2024/most-california-high-school-seniors-shut-out-of-even-applying-to-the-states-universities/705635#:~:text=
- Grunow, A., & Park, S. (2019, April). *Towards a common vision of continuous improvement*. [Paper presentation]. American Educational Research Association Annual Meeting, Toronto, Canada.
- Hough, H. J., & Chavez, B. (2022, October). *California test scores show the devastating impact of the pandemic on student learning* [Commentary]. Policy Analysis for California Education. edpolicyinca.org/newsroom/california-test-scores-show-devastating-impact-pandemic-student-learning
- Institute for Healthcare Improvement. (2003). *The Breakthrough Series: IHI's collaborative model for achieving breakthrough improvement* [IHI Innovation Series white paper]. ihi.org/resources/white-papers/breakthrough-series-ihis-collaborative-model-achieving-breakthrough
- Johnson, S. (2021, February 9). *California teachers grapple with grading nearly a year after initial school closures*. EdSource. edsources.org/2021/california-teachers-grapple-with-grading-nearly-a-year-after-initial-school-closures/648376
- Pekel, K., Sethi, J., Muhammad, F. Z., Eisenberg, C. & Bergstrom, E. (2021). *The relationship builder's guidebook: Activities and approaches to enhance developmental relationships*. Search Institute.
- Phillips, E. K. (2019). *The make-or-break year: Solving the dropout crisis one ninth grader at a time*. The New Press.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Harvard University Press.
- University of Chicago Network for College Success. (n.d.). *Understanding research and applying data*. ncs.uchicago.edu/sites/ncs.uchicago.edu/files/uploads/components/NCS_FOT_Toolkit_URAD.pdf
- University of Chicago Network for College Success. (2023). *Ninth-grade success and college readiness*. ncs.uchicago.edu/content/ninth-grade-success-and-college-readiness
- What Works Clearinghouse. (2017, September). *Preventing dropout in secondary schools* (NCEE 2017-4028). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. ies.ed.gov/ncee/WWC/Docs/PracticeGuide/www_dropout_092617.pdf
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. ies.ed.gov/ncee/WWC/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf
- Yeager, D., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Practical measurement* [unpublished manuscript]. Carnegie Foundation for the Advancement of Teaching. carnegiefoundation.org/wp-content/uploads/2013/12/Practical_Measurement.pdf

Appendix A

Table A1. Codebook: Definitions of Variables Used in Report Regressions and Analysis

Variable	Description
Ninth-grade on-track points	BTSC metric that ranges from 0 to 100 for ninth graders
Treatment	A flag that is equal to 1 for schools in BTSC and 0 for the comparison schools
Cohort	Possible Values: 1, 2, 3. Cohort 1 launched in April 2019. Cohort 2 launched in April 2020. Cohort 3 launched in September 2021.
Eighth-grade on-track points	BTSC metric that ranges from 0 to 100 for eighth graders
Has 8OT flag	A flag that is equal to 1 for students with eighth-grade on-track points and 0 for students not having this score
Race/ethnicity	<p>If a student is Hispanic, then this variable is Hispanic; otherwise, use the available student race.</p> <p>Possible Values: White, Asian, American Indian, Pacific Islander, Filipino, African American, Multiracial, Hispanic.</p> <p>The primary reason for this approach is the observation that many Hispanic students did not indicate a specific race. To address this gap and ensure consistent classification, this variable prioritizes Hispanic identity when available.</p>
EL status	Possible values: English Learner (EL), English Only (EO), Initially Fluent English Proficient (IFEP), Reclassified Fluent English Proficient (RFEP), or to be determined (TBD)
Parental highest education	Possible values: Graduate Degree or Higher, College Graduate, Some College or AA, High School Graduate, Not a High School Graduate, Decline to State
Enrolled in US <3 years	Student enrolled in a U.S. school for less than 3 years
Hispanic flag	Student indicator for whether or not Hispanic
Race	Possible values: White, Asian, American Indian, Pacific Islander, Filipino, African American, Multiracial

Table A2. Covariates for Regression Analyses

Year	Covariate	Subcategory	Treatment	Comparison
2022–23	Eighth-grade OT points		52.365 (21.161)	54.303 (22.642)
2022–23	Flag for eighth-grade OT points		0.821 (0.383)	0.782 (0.413)
2022–23	EL classification	English learner	0.254 (0.435)	0.167 (0.373)
2022–23	EL classification	English or ASL only	0.380 (0.485)	0.438 (0.496)
2022–23	EL classification	Initial Fluent English Proficient	0.023 (0.150)	0.046 (0.209)
2022–23	EL classification	Reclassified Fluent English Proficient	0.343 (0.475)	0.349 (0.477)
2022–23	EL classification	To be determined	0.001 (0.028)	0.000 (0.015)
2022–23	Enrolled in US less than 3 years		0.113 (0.317)	0.106 (0.308)
2022–23	Parent education	College graduate	0.089 (0.285)	0.116 (0.320)
2022–23	Parent education	Decline to state	0.377 (0.485)	0.269 (0.444)
2022–23	Parent education	Graduate degree or higher	0.051 (0.220)	0.060 (0.238)
2022–23	Parent education	High school graduate	0.173 (0.378)	0.233 (0.423)
2022–23	Parent education	Not a high school graduate	0.204 (0.403)	0.184 (0.387)
2022–23	Parent education	Some college or AA	0.105 (0.307)	0.138 (0.345)
2022–23	Hispanic		0.719 (0.450)	0.687 (0.464)
2022–23	Race only	African American	0.107 (0.309)	0.120 (0.325)
2022–23	Race only	American Indian	0.194 (0.396)	0.065 (0.247)
2022–23	Race only	Asian	0.139 (0.346)	0.098 (0.297)
2022–23	Race only	Filipino	0.014 (0.119)	0.022 (0.146)
2022–23	Race only	Multiracial	0.051 (0.220)	0.045 (0.207)

Year	Covariate	Subcategory	Treatment	Comparison
2022–23	Race only	Pacific Islander	0.013 (0.111)	0.009 (0.097)
2022–23	Race only	White	0.482 (0.500)	0.641 (0.480)
2022–23	Race/ethnicity	African American	0.072 (0.259)	0.092 (0.289)
2022–23	Race/ethnicity	American Indian	0.002 (0.048)	0.001 (0.036)
2022–23	Race/ethnicity	Asian	0.098 (0.297)	0.076 (0.265)
2022–23	Race/ethnicity	Filipino	0.009 (0.097)	0.017 (0.128)
2022–23	Race/ethnicity	Hispanic	0.722 (0.448)	0.688 (0.463)
2022–23	Race/ethnicity	Multiracial	0.025 (0.156)	0.028 (0.164)
2022–23	Race/ethnicity	Pacific Islander	0.007 (0.081)	0.006 (0.076)
2022–23	Race/ethnicity	White	0.065 (0.247)	0.092 (0.289)
2021–22	Eighth-grade OT points		50.798 (26.054)	51.857 (26.270)
2021–22	Flag for eighth-grade OT points		0.818 (0.386)	0.787 (0.409)
2021–22	EL classification	English learner	0.261 (0.439)	0.179 (0.384)
2021–22	EL classification	English or ASL only	0.380 (0.485)	0.418 (0.493)
2021–22	EL classification	Initial Fluent English Proficient	0.025 (0.155)	0.046 (0.209)
2021–22	EL classification	Reclassified Fluent English Proficient	0.333 (0.471)	0.356 (0.479)
2021–22	EL classification	To be determined	0.002 (0.042)	0.001 (0.033)
2021–22	Enrolled in US less than 3 years		0.107 (0.310)	0.116 (0.320)
2021–22	Parent education	College graduate	0.093 (0.290)	0.112 (0.316)
2021–22	Parent education	Decline to state	0.324 (0.468)	0.284 (0.451)
2021–22	Parent education	Graduate degree or higher	0.067 (0.251)	0.058 (0.234)

Year	Covariate	Subcategory	Treatment	Comparison
2021–22	Parent education	High school graduate	0.180 (0.384)	0.228 (0.419)
2021–22	Parent education	Not a high school graduate	0.216 (0.412)	0.183 (0.387)
2021–22	Parent education	Some college or AA	0.119 (0.324)	0.134 (0.341)
2021–22	Hispanic		0.710 (0.454)	0.699 (0.459)
2021–22	Race only	African American	0.116 (0.321)	0.117 (0.322)
2021–22	Race only	American Indian	0.181 (0.385)	0.056 (0.231)
2021–22	Race only	Asian	0.139 (0.346)	0.098 (0.297)
2021–22	Race only	Filipino	0.017 (0.129)	0.019 (0.136)
2021–22	Race only	Multiracial	0.056 (0.229)	0.036 (0.185)
2021–22	Race only	Pacific Islander	0.012 (0.107)	0.010 (0.098)
2021–22	Race only	White	0.480 (0.500)	0.664 (0.472)
2021–22	Race/ethnicity	African American	0.078 (0.269)	0.089 (0.285)
2021–22	Race/ethnicity	American Indian	0.002 (0.044)	0.002 (0.045)
2021–22	Race/ethnicity	Asian	0.101 (0.301)	0.078 (0.267)
2021–22	Race/ethnicity	Filipino	0.011 (0.102)	0.013 (0.115)
2021–22	Race/ethnicity	Hispanic	0.712 (0.453)	0.701 (0.458)
2021–22	Race/ethnicity	Multiracial	0.021 (0.143)	0.021 (0.142)
2021–22	Race/ethnicity	Pacific Islander	0.007 (0.082)	0.006 (0.076)
2021–22	Race/ethnicity	White	0.069 (0.253)	0.091 (0.287)

Table A3. Logistic Model Regression Results

Table A3 presents D or F regression results for the random effects logistic model where the dependent variable is a flag for whether a D or F was obtained in an A–G course.

Variable	Logistic model	Variable	Logistic model
D or F flag		EL status	
Treatment	-0.026	EL	-0.114
Cohort		EO	(base)
1	(base)	IFEP	0.219*
2	0.647***	RFEP	0.172***
3	0.111	Parental highest education	
Eighth-grade on-track points	-0.085***	Graduate degree or higher	0.000
White	(base)	Graduate degree or higher	-0.448***
Asian	-0.087	College graduate	-0.350***
American Indian	1.652**	Some college or AA	-0.035
Pacific Islander	0.034	High school graduate	(base)
Filipino	-0.037	Not a high school graduate	0.078
African American	0.372***	Decline to state	-0.053
Multiracial	0.235	Constant	4.249***
Hispanic	0.317***	var(_cons[school_code])	0.252***
		Statistics	
		N	19,402

Note. *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.10$.

Table A4. All Regression Output

Table A4 includes all models with their coefficients and standard errors below each coefficient in italicized gray font.

Variable	Confirmatory Model	(1) "Base Model"	(2) With 8OT Flag	(3) Hispanic Flag	(4) Hispanic, Race	(5) Similar to model (2) excl. undisclosed	(6) Only 8OT	(7) Confirmatory 2021-22	(8) Base Model 2021-22	(9) With 8OT 2021-22
9th Grade On-Track										
Treatment		-0.532 <i>0.948</i>	-0.186 <i>1.183</i>	-0.297 <i>1.195</i>	-0.561 <i>0.93</i>	-0.827 <i>0.926</i>	-0.999 <i>0.958</i>	-0.686 <i>0.899</i>	0.318 <i>1.205</i>	0.686 <i>1.505</i>
Cohort										
1		(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
2		-1.484 <i>1.204</i>	0.93 <i>1.505</i>	0.925 <i>1.521</i>	-1.531 <i>1.182</i>	-1.422 <i>1.181</i>	-1.632 <i>1.225</i>	-1.771 <i>1.143</i>	-0.738 <i>1.534</i>	1.56 <i>1.919</i>
3		0.204 <i>1.182</i>	3.391 <i>1.476</i>	3.385 <i>1.491</i>	0.281 <i>1.16</i>	0.168 <i>1.148</i>	0.562 <i>1.196</i>	0.08 <i>1.121</i>	-0.52 <i>1.508</i>	2.749 <i>1.883</i>
8th Grade On-Track Points		0.827 <i>0.005</i>			0.83 <i>0.005</i>	0.813 <i>0.006</i>	0.814 <i>0.006</i>	0.853 <i>0.005</i>	0.648 <i>0.005</i>	
White		(base)	(base)	(base)			(base)		(base)	(base)
Asian		-0.465 <i>0.538</i>	7.618 <i>0.734</i>	7.572 <i>0.734</i>			-0.47 <i>0.641</i>	2.379 <i>0.552</i>	6.796 <i>0.717</i>	6.794 <i>0.716</i>
American Indian		-7.144 <i>2.385</i>	-13.668 <i>3.195</i>	-13.474 <i>3.194</i>			-4.028 <i>3.204</i>	-2.848 <i>3.012</i>	-8.253 <i>3.012</i>	-8.079 <i>3.009</i>
Pacific Islander		0.108 <i>1.356</i>	-4.255 <i>1.817</i>	-4.189 <i>1.816</i>			1.536 <i>1.684</i>	-4.257 <i>1.333</i>	-7.394 <i>1.768</i>	-7.439 <i>1.767</i>
Filipino		-0.694 <i>0.946</i>	7.248 <i>1.308</i>	7.096 <i>1.307</i>			-0.828 <i>1.061</i>	1.106 <i>1.027</i>	4.934 <i>1.341</i>	4.875 <i>1.34</i>
African American		-2.549 <i>0.538</i>	-11.314 <i>0.705</i>	-11.239 <i>0.705</i>			-2.602 <i>0.639</i>	-5.252 <i>0.55</i>	-14.897 <i>0.692</i>	-14.835 <i>0.691</i>
Multiracial		-1.161 <i>0.711</i>	-1.917 <i>0.967</i>	-1.93 <i>0.967</i>			-1.42 <i>0.825</i>	-1.022 <i>0.831</i>	-2.544 <i>1.046</i>	-2.405 <i>1.046</i>
Hispanic		-2.033 <i>0.426</i>	-7.512 <i>0.566</i>	-7.554 <i>0.566</i>			-1.585 <i>0.496</i>	-3.217 <i>0.441</i>	-9.648 <i>0.559</i>	-9.619 <i>0.558</i>
EL Status										
EL		-0.346 <i>0.345</i>	-9.549 <i>0.46</i>	-9.397 <i>0.461</i>	-0.14 <i>0.461</i>	-0.359 <i>0.411</i>	-0.761 <i>0.439</i>	-2.093 <i>0.357</i>	-8.57 <i>0.454</i>	-8.432 <i>0.454</i>
EO		(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
IFEP		1.325 <i>0.574</i>	8.398 <i>0.769</i>	8.344 <i>0.769</i>	1.393 <i>0.573</i>	0.777 <i>0.616</i>	0.789 <i>0.671</i>	2.935 <i>0.605</i>	7.764 <i>0.752</i>	7.776 <i>0.751</i>
RFEF		1.739 <i>0.27</i>	6.248 <i>0.373</i>	6.185 <i>0.373</i>	1.889 <i>0.261</i>	1.769 <i>0.31</i>	1.133 <i>0.323</i>	3.715 <i>0.284</i>	6.965 <i>0.374</i>	6.871 <i>0.374</i>
Parental Highest Education										
Graduate Degree or Higher		2.565 <i>0.491</i>	10.634 <i>0.667</i>	10.616 <i>0.667</i>	2.651 <i>0.488</i>	2.61 <i>0.534</i>	2.738 <i>0.501</i>	2.437 <i>0.5</i>	9.009 <i>0.631</i>	9.134 <i>0.631</i>
College Graduate		1.775 <i>0.398</i>	7.033 <i>0.538</i>	7.054 <i>0.538</i>	1.861 <i>0.397</i>	1.657 <i>0.426</i>	1.908 <i>0.406</i>	1.651 <i>0.413</i>	6.812 <i>0.531</i>	6.863 <i>0.531</i>
Some College or AA		1.104 <i>0.365</i>	2.889 <i>0.498</i>	2.875 <i>0.498</i>	1.07 <i>0.366</i>	1.135 <i>0.394</i>	1.066 <i>0.368</i>	1.038 <i>0.374</i>	2.707 <i>0.483</i>	2.763 <i>0.483</i>
High School Graduate		(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Not a High School Grad		-1.185 <i>0.329</i>	-2.181 <i>0.439</i>	-2.218 <i>0.439</i>	-1.21 <i>0.329</i>	-0.789 <i>0.359</i>	-1.184 <i>0.331</i>	-1.437 <i>0.336</i>	-2.445 <i>0.425</i>	-2.464 <i>0.425</i>
Decline to State		0.36 <i>0.3</i>	1.309 <i>0.402</i>	1.26 <i>0.402</i>	0.297 <i>0.301</i>	0.431 <i>0.335</i>		0.366 <i>0.312</i>	0.636 <i>0.395</i>	0.594 <i>0.395</i>
EL Status										
TBD			-38.597 <i>6.271</i>	-37.933 <i>6.269</i>				28.62 <i>14.547</i>	-32.664 <i>3.322</i>	-32.004 <i>3.321</i>
Enrolled in US < 3 Years			1.477 <i>0.506</i>	2.435 <i>0.537</i>	0.613 <i>0.515</i>	0.715 <i>0.599</i>	0.392 <i>0.671</i>		1.774 <i>0.5</i>	2.999 <i>0.539</i>
Has 8OT Flag				2.151 <i>0.406</i>						2.426 <i>0.398</i>
Hispanic					-0.949 <i>0.262</i>	-1.515 <i>0.372</i>				
White					(base)					
Asian						0.031 <i>0.472</i>				
American Indian						-0.259 <i>0.471</i>				
Pacific Islander						-0.463 <i>1.148</i>				
Filipino						-0.266 <i>0.86</i>				
African American						-2.638 <i>0.47</i>				
Multiracial						-1.869 <i>0.56</i>				
Constant		11.28 <i>1.137</i>	54.204 <i>1.39</i>	52.518 <i>1.438</i>	9.893 <i>1.067</i>	11.738 <i>1.105</i>	11.948 <i>1.176</i>	9.277 <i>0.993</i>	20.909 <i>1.385</i>	54.563 <i>1.71</i>
Ins1_1_1										
Constant		1.236 <i>0.104</i>	1.46 <i>0.101</i>	1.472 <i>0.101</i>	1.216 <i>0.104</i>	1.198 <i>0.108</i>	1.213 <i>0.109</i>	1.18 <i>0.104</i>	1.494 <i>0.099</i>	1.721 <i>0.097</i>
Insig_e										
Constant		2.631 <i>0.005</i>	3.027 <i>0.005</i>	3.026 <i>0.005</i>	2.63 <i>0.005</i>	2.602 <i>0.006</i>	2.623 <i>0.006</i>	2.637 <i>0.005</i>	2.675 <i>0.005</i>	3.022 <i>0.005</i>
Statistics										
N		19,401	23,613	23,613	19,217	14,983	13,245	19,489	19,430	24,184
ll		-78,656.03	-105,056.34	-105,042.34	-77,882.71	-60,316.10	-53,606.40	-79,122.91	-79,635.24	-107,499.41
chi2		31,418.93	3,862.96	3,894.22	31,132.74	25,479.69	21,374.77	31,043.24	27,467.38	4,286.53
aic		157,356.07	210,158.68	210,132.68	155,799.42	120,678.20	107,256.80	158,259.81	159,316.48	215,009.68
bic		157,529.28	210,344.28	210,326.35	155,933.10	120,853.33	107,421.61	158,314.96	159,497.60	215,203.92

Author Biographies

H. Alix Gallagher is director of strategic partnerships at PACE, where she leads PACE’s work around system reform to support improved student outcomes.

Benjamin W. Cottingham is associate director of strategic partnerships at PACE. His work focuses on the development of education systems and conditions that support the improvement of student outcomes across California.

Leah Faw is a qualitative researcher at PACE with a focus on continuous improvement. Her research has considered homeschooling, parental choice, school choice, and school improvement.

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

Inquiry: info@edpolicyinca.org
Media: press@edpolicyinca.org

Office: 650.576.8484

Follow us on social media:

edpolicyinca.org

