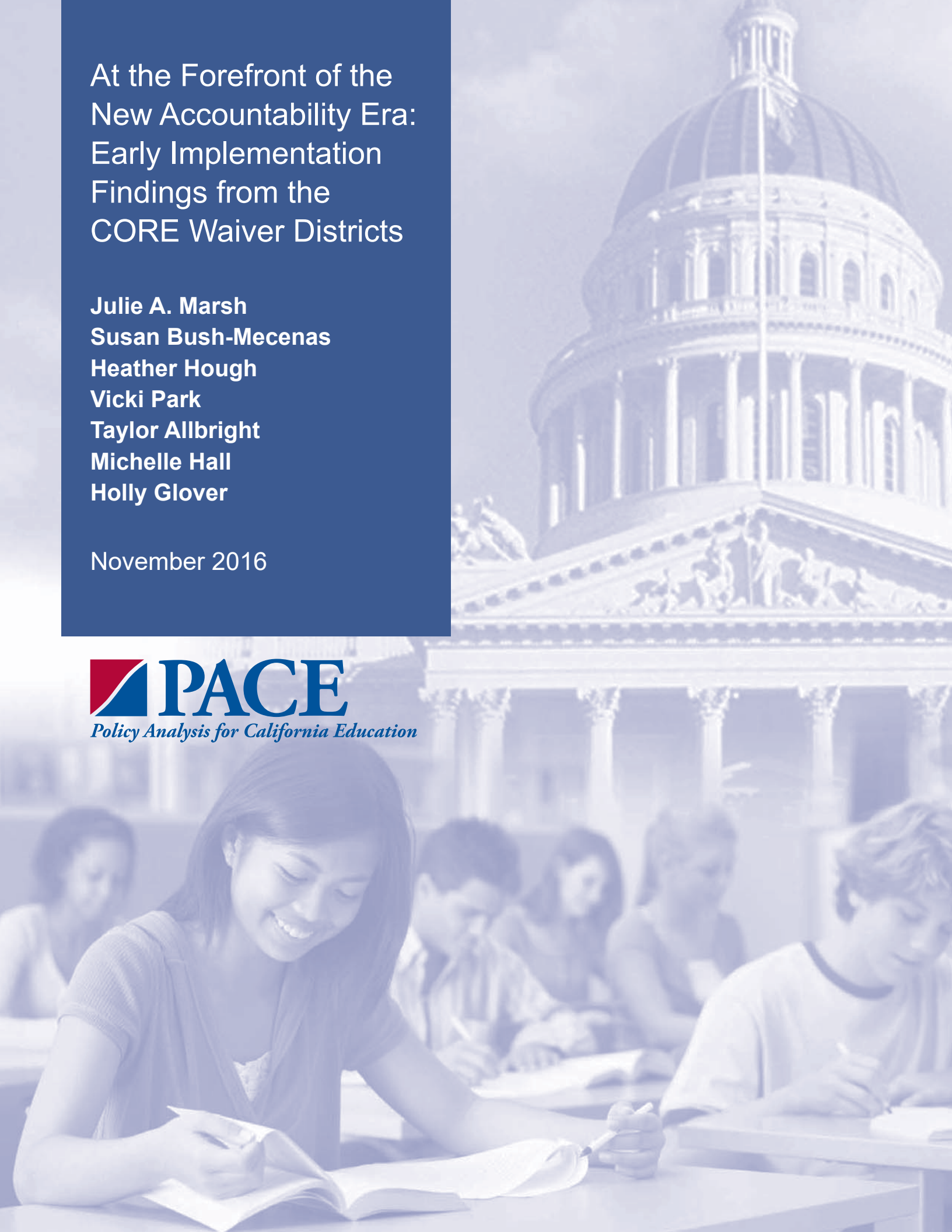
At the Forefront of the New Accountability Era: Early Implementation Findings from the CORE Waiver Districts

Julie A. Marsh
Susan Bush-Mecenas
Heather Hough
Vicki Park
Taylor Allbright
Michelle Hall
Holly Glover

November 2016



PACE
Policy Analysis for California Education

PACE

At the Forefront of the New Accountability Era:

Early Implementation Findings from the CORE Waiver Districts

Julie A. Marsh (University of Southern California), Susan Bush-Mecenas (University of Southern California), Heather J. Hough (Stanford University), Vicki Park (Uuniversity of Utah), Taylor Allbright (University of Southern California), Michelle Hall (University of Southern California), Holly Glover (Stanford University)

Acknowledgements

In October 2015, Policy Analysis for California Education (PACE) launched a research partnership with the California Office to Reform Education (CORE) districts. This research partnership is focused on producing research that informs continuous improvement in the CORE districts and policy in California and beyond.

We would like to thank the many individuals who contributed to this report. First, we are grateful to the generous sponsor of this research, the S. D. Bechtel, Jr. Foundation. We also thank all of the leaders and administrators in CORE and the CORE waiver districts for their support throughout this project, along with the many school principals and facilitators who participated in the research activities and shared their valuable time and insights with us. This project would not have been completed without the assistance of many colleagues at the USC Rossier School of Education and Stanford University.

Abstract

California and the nation are at the crossroads of a major shift in school accountability policy. At the state level, California's Local Control and Accountability Plan (LCAP) encourages the use of multiple measures of school performance used locally to support continuous improvement and strategic resource allocation. Similarly, the federal Every Student Succeeds Act (ESSA) reinforces this local control, requiring more comprehensive assessment of school performance and a less prescriptive, local approach to school support. These changes represent a major cultural shift for California schools and districts. Ahead of the curve, six California districts, known as the CORE waiver districts, have implemented an innovative measurement system and supports for school and district improvement under an NCLB waiver, and thus provide a unique opportunity to examine and learn from the enactment of a system supported by accountability policy in this new era. This report examines the early implementation and effects of the CORE reform and seeks to inform the ongoing efforts within CORE as well as the development and implementation of future accountability policy in other states and districts.

Introduction

California is in the midst of a dramatic shift in how it educates children. In 2013, the Local Control Funding Formula (LCFF) transferred to local actors major decision-making authority over how to allocate resources to meet students' needs. California's related Local Control and Accountability Plan (LCAP) process encourages the use of multiple measures of school performance used locally to support continuous improvement. The passage of the federal Every Student Succeeds Act (ESSA) of 2015 reinforces this local control and similarly requires a more comprehensive approach to assessing school performance. ESSA includes both academic and non-academic measures and places the responsibility for intervening in low-performing schools primarily with districts. Underlying both of these major policy shifts is the idea that local leaders are in the best position to drive real educational improvement. However, their ability to do so hinges on their capacity to use data for improvement and to enact change at the district office and in schools system-wide. Given that the past decade under the No Child Left Behind (NCLB) act focused on compliance to a different federal approach to accountability, building this capacity is no small feat.

While much is known about the implementation and effects of NCLB (e.g., Dee & Jacob, 2011; Neal & Schanzenbach, 2010; Stecher et al., 2008), there is little known about the new set of accountability systems likely to emerge under ESSA (ESSA, 2015). And while some states have innovated slightly under the NCLB waiver policy, few made dramatic changes akin to those called for under ESSA. For example, the accountability systems in waiver states relied upon state-driven interventions for struggling schools and few incorporated measurement systems that include more than state test results in math and English language arts (Polikoff, McEachin, Wrabel, & Duque, 2014).

California's six CORE waiver districts—Fresno, Long Beach, Los Angeles, Oakland, Santa Ana, and San Francisco—provide a unique opportunity to understand and learn from the enactment of a system supported by accountability policy in this new era. Freed by the U.S. Department of Education from some of their federal obligations under NCLB in 2013, these six districts developed and are currently implementing CORE's accountability system (the School Quality Improvement System). Key features of this accountability system are (a) a measurement system (the School Quality Improvement Index, hereafter MS) that focuses on academic outcomes alongside non-academic measures of student success, including chronic absenteeism, suspension/expulsion, students' social-emotional skills, and school climate and culture; (b) peer-to-peer school improvement interventions; and (c) district-level capacity-building.

In this policy report, we seek to leverage the experiences of the CORE districts to promote a better understanding of what it means to implement a multiple-measure accountability system and a system of locally determined, collaborative improvement efforts.[i] Notably, we seek to identify the potential challenges and successes. We believe these findings from the early implementation of the CORE accountability system can inform the development and implementation of future accountability policy in states and districts nationwide. Specifically, the report addresses the following questions:

1. How did the CORE districts and schools respond to the new measurement system?
2. What was the districts' approach to school improvement and how was it working?
3. How has cross-district collaboration supported capacity for school and system improvement?

Using a multiple case study design (Yin, 2013), we gathered data on educators' experiences implementing CORE's new measurement system and peer-to-peer interventions at the school and district level. In 2015–16, researchers interviewed CORE leaders and central office administrators (n=45) and

principals (*n*=15) in the six CORE waiver districts, observed CORE meetings (42 hours), and gathered documents (e.g., presentations, peer review, and data reports). We analyzed each CORE district individually and then conducted cross-case analyses to examine how implementation varied by district and the factors associated with patterns acroFss sites (Miles, Huberman, & Saldaña, 2013).

In this report, we describe the implementation of the accountability and improvement efforts in the CORE districts. First we present background on the CORE districts and describe our research methods. Next, we present the results of our analysis of how districts and schools implemented and responded to the measurement system, the peer-to-peer interventions at the school level, and district-level capacity-building efforts. We then present cross-cutting factors broadly related to implementation and conclude with implications of our work for policy and practice.

Background on CORE

The CORE districts' collaboration began in 2010, when a group of districts came together to collaborate around the implementation of the challenging new Common Core State Standards. Informal networking opportunities led to increasingly formal collaborations, including an application for the U.S. Department of Education's Race to the Top program. Although the districts did not win, the process developed strong relationships and helped the districts identify overlap in their goals and values.[ii]

This experience laid the groundwork for the CORE districts' application in 2013 for an NCLB flexibility request (US DOE, 2012). At the point of application, the majority of states had received "waivers," and California's request had been denied (Johnson, 2012 December 26). Without the state waiver, leaders in the CORE districts, which represent 20% of the states' students and had a large proportion of the schools receiving sanctions under NCLB, felt the need to act. A primary motivation for the districts was to recoup federal dollars that were earmarked for specific interventions and use them in the way they saw fit. As written in the waiver: "A federal ESEA waiver will give Participating LEAs essential flexibility to target dollars directly towards the needs of students, which is vital to reorienting districts' individual and collective work around the moral imperative to prepare all students for college and careers and eliminate disparity and disproportionality" (CORE, 2013, p. 3).

CORE's pursuit of the waiver rose from the prior collaboration work in a natural way. The district leaders did not feel that the accountability system they were working under reflected their emerging vision (fueled by the Common Core) of how students succeed. As one leader said,

> Curriculum leaders in districts were struggling with the fact that school and district accountability [measures] were aligned to the old California standards, but yet the state had adopted and the districts were adopting the Common Core and beginning to design implementation plans. They were living with each foot in a different world: a world for accountability and a world for the future going forward and new standards. It always came up in conversation, "Wouldn't it be great if we could just go all in and not give the CSTs, have a waiver, and be held accountable with our new set of standards?"

The ultimate waiver plan, which was approved in August 2013, laid out CORE's theory of action: The CORE work would be a "system with a higher level of shared responsibility and accountability" (CORE, 2013, p. 1). Their collaboration under the waiver was undergirded by three overarching tenets:

1. *The importance of local control.* In both the development and implementation of the waiver, the districts emphasized that they wanted to learn from each other and hold each other accountable for outcomes, but that they wanted full autonomy to implement approaches locally, "not because of the desire to escape statewide expectations but because each community is truly unique" (CORE, 2013, pp. 17–18).
2. *A move from compliance to shared responsibility.* The districts fundamentally believed that NCLB's underlying theory of action was flawed: Schools/districts did not need sanctions and punishment; they needed flexibility to do what is best for their students and the support of one another in making big improvements. As stated in the waiver, "This is a paradigm shift away from a compliance-based accountability system to one driven by the collective and individual responsibility to adhere to this new set of principles, with shared responsibility and support building from educator to educator, from school to school, and from district to district" (ibid., p. 24). To support this vision, the CORE districts implemented a formal peer review process to ensure district compliance under the waiver (see section below for more details).
3. *Capacity-building through peer-to-peer collaboration with a focus on data.* Giving districts and schools the flexibility to improve hinges on the capacity of the district staff to identify problems and know how to fix them. For this reason, "it is CORE's hope to let data drive all actions and rely on peer-to-peer collaboration and support as much as possible" (ibid., p. 20). The districts believed that the capacity to improve existed within the districts, and that collaboration across districts would spread expertise.

The waiver was granted in August 2013, and the passage of ESSA in December 2015 changed the districts' implementation timeline substantially. The CORE districts had planned a three-year roll-out to full implementation, with the full measurement system built and used to identify (and redesignate) the lowest performing schools (focus and priority schools, see below for further description) for the 2016–17 school year. With the federal ESEA reauthorization, this school redesignation will now not occur. The waiver is only formally in effect through August of 2016, but the CORE districts remain committed to their collaboration. The districts (including Sacramento City and Garden Grove) are currently reorganizing as a Networked Improvement Community (Bryk, Gomez, Grunow, LeMahieu, 2015), using the data in the measurement system (MS) and the tools of Improvement Science to anchor their collaborative work together in cycles of inquiry.[iii] They will also welcome new members via the CORE "Data Collaborative," in which participating districts across the state will share data and benchmark student performance. Finally, the CORE districts have petitioned the state to allow them to continue their work as a research pilot under ESSA, in essence continuing to operate their accountability model.[iv]

CORE's work under the waiver was divided into three principles, with the work on each principle organized and executed separately: (a) the implementation of high-quality instruction under the new standards and assessment of student learning; (b) measurement of school quality and support of struggling schools; and (c) teacher, principal, and superintendent evaluation and support systems. These three principles together represented the multidimensional work envisioned to meet the districts' ambitious goals for student success. In this study, we focused primarily on the second area of work, which we detail below, during the 2015–16 school year, or the third year of implementation. It is important to note that not all aspects of the waiver system were fully operational at the time of our study, and that the system itself was in flux due to ESEA reauthorization in December 2015.

Measuring School Quality

CORE's MS focuses on academic outcomes alongside non-academic measures of student success, including chronic absenteeism, suspension/expulsion, students' social-emotional skills, and school climate and culture. CORE's systematic measurement of school and student performance on multiple dimensions is relatively unique, and has generated widespread national interest (Blad, 2015; Bornstein, 2015; Zernike, 2016).

CORE's measurement system aligns with new requirements for measurement under both LCFF and ESSA. LCFF requires districts to measure progress towards goals aligned with the state's eight priorities: student achievement, student engagement, parental involvement, school climate, basic services, implementing Common Core, course access, and other student outcomes. Similarly, ESSA expands measurement by requiring states to include multiple measures of student performance, including: academic performance as measured by proficiency on English Language Arts (ELA) and math tests; academic growth; graduation rate; development of English Learner (EL) proficiency; and at least one additional indicator of School Quality or Student Success (SQSS). The SQSS indicator can include measures of student engagement, educator engagement, student access to and completion of advanced coursework, post-secondary readiness, or school climate, safety, or other measures that meet the technical criteria, including SEL. CORE has a measurement system in each of these domains that they have aligned with research evidence (for details on the calculation methods and specific research behind the measures see CORE districts, 2015).[v] Table 1 defines these measures.

Table 1: Elements of CORE's Measurement System

CORE's Academic Domain	
Academic performance	Percentage of students testing proficient for ELA and math, based on Smarter Balanced Assessment Consortium (SBAC) test scores.
Academic growth	Growth percentile (rank from 0-100) comparing schools' contribution to student growth on ELA and math test scores, measuring the extent to which students in a given school have improved their performance on ELA and math tests from one year to the next relative to demographically similar students who started the school year with similar prior achievement.[vi]
Graduation	Percentage of students who graduate in a 4-, 5-, and 6-year cohort compared with the number of students enrolled in the school (accounting for students who transfer into and out of the school).
8th-grade students' high school readiness	Percentage of all eighth-grade students who meet the following criteria: 1) eighth-grade GPA of 2.5 or higher; 2) attendance of 96% or higher; 3) no grades of D or F in ELA or math in the final course grade; and 4) were not suspended in eighth grade. These performance thresholds were determined based on analysis of outcomes of students in the CORE districts.
EL redesignation	Percentage of students who are reclassified from English language learner status to "fluent English proficient" out of the number of all the English learners who are reclassified at a school site in the current year plus all those English learners who, after five years, were not reclassified at that school.[vii]
CORE's SEL/CC Domain	
Chronic absence	Percentage of students who have an attendance rate at or below 90% within a given school year.
School culture/climate (CC)	Percentage of positive responses in each school, similar to the indicator of social-emotional skills, produced from surveys of students (Grades 4-12), teachers, and parents that include questions about the climate of support for academic learning, knowledge and perceived fairness of discipline rules and norms, school safety, and sense of belonging and school connectedness.[viii]
Suspension/ expulsion	Percentage of students who are suspended and/or expelled at least once in a given school year.
Social-emotional skills/learning (SEL)	Percentage of positive responses in each school, produced from students' self-report surveys in Grades 4-12 that measure growth mindset, self-efficacy, self-management, and social awareness. Student responses on these surveys are translated into the percentage of positive responses in each school; for example, a school with a score of "80" would indicate that 80% of the survey questions were answered positively by students.[ix]

In the full MS (which would have been rolled out in 2016–17), the academic domain accounts for 60% of the final score, and measures in the ESSA SQSS domain account for 40%. Each domain in the MS is weighted, with those weights aggregated into a single number (which was a federal requirement for the identification of priority schools). For most metrics (except EL redesignation), MS points are divided between the all-students group and the four subgroup categories (lowest performing racial ethnic group,[x] English Learners, students with disabilities, and disadvantaged students). In the reporting of student subgroups, the CORE districts utilized an "n size" of 20.[xi] For each metric and each subgroup, schools are given an MS level score that compares them across other CORE schools. These metric cut points (1–10) are

established for each indicator based on an initial year of data and then maintained over several years. This was to avoid the outcome, as with California's previous Academic Performance Index (API), where 10% of schools are always identified in level 1. By setting the levels and keeping them for multiple years, schools can all show improvement on a metric.

The measurement reports, released for each school, are designed to display all of the MS measures, including three-year trends and comparisons with other schools and districts. These reports were made available publicly on the CORE website, but, according to CORE staff, were released "without much fanfare." It is important to note that in the 2015–16 school year (in which we conducted our research), social-emotional skills, culture climate, and academic growth did not yet count—they were measured but were not yet reported on the MS or included in the final score. However, surveys were administered in all schools as part of a field test, and the results were shared back with the districts and schools on a third-party website.[xii]

The guiding principle of the CORE districts in their use of multiple measures is that data should be used as "a flashlight, not a hammer." What indicators reveal about school performance should be used to help them improve rather than punish. This ethos permeated all of their decisions about what to measure, how to measure it, and how to report it. In evaluating specific measures for inclusion, it was determined that each construct should be "measurable, meaningful, and actionable." In other words, measures should have clear evidence of validity and reliability and should predict college and career outcomes, and schools must be able to influence and impact the outcome in question. This emphasized another important aspect of the MS, which is that all measures should be useful for continuous improvement, and not just for accountability. From its conception, the MS was envisioned as a subset of the measures that are used for continuous improvement within the districts. In this way, the MS would not replace individual districts' internal accountability systems (or district data dashboards) that reach beyond federal accountability expectations. This is another way in which the CORE waiver was focused on "ensuring that local values, as well as community cultures and needs, are included within district accountability systems" (CORE, 2013, pp. 38–39).

Supporting Struggling Schools

Under NCLB/ESEA, schools performing in the bottom 5% of all schools in a state were required to undergo interventions for dramatic school turnaround, which included reconstitution, restructuring, closure, restart, and charter takeover.[xiii] As part of their waiver from the NCLB/ESEA requirements, the CORE districts designed a system of intervention they believed would better meet the needs of their low-performing schools. In developing their interventions, the CORE districts believed that barriers to school improvement and appropriate interventions are necessarily contextually driven and local in nature. That is, an approach that worked well in one school might not work well in a school across town. Furthermore, CORE leaders believed that practitioners already possess the knowledge and skills to improve, and schools can improve most quickly when allowed the autonomy and provided the support to meet, share successful practice, and engage in structured inquiry to guide improvement efforts. In the interventions that the CORE districts developed, schools were provided a framework for engaging in inquiry and knowledge sharing, but their ultimate decision-making and implementation were non-prescriptive in nature.

Specifically, 178 schools were identified across the CORE districts to engage in two tiers of improvement activities. Priority schools—those falling in the lowest 5% of all schools, receiving School Improvement Grants, or having a graduation rate below 60% in the CORE districts—were paired with reward schools, which were high performing or demonstrated high growth with similar student populations.

In these School Pairings, schools were expected to meet several times over the year, develop a plan with the assistance of a district facilitator, and were allocated funding to support the pair's activities. Focus schools, those with low performing subgroups or students, were grouped into Communities of Practice.[xiv] As with the pairing activities, within these Communities of Practice, schools were expected to meet several times per year, develop a plan with the help of a district facilitator, and were allocated funding to support these activities. Interestingly, some of the Communities of Practice and School Pairings crossed district lines.

As stated in the waiver, the interventions were expected to focus on local needs and to leverage the school-site staff as those best capable of improving the school:

> Specific interventions priority schools might pursue will be aligned to system reform drivers (fostering intrinsic motivation, continuous improvement, collective team work, and "allness")…This means focusing on interventions that situate the energy of educators and students as the central driving force of change.[xv] (CORE, 2013, p. 38)

In the case of pairings, it was assumed that the high-performing school leaders would serve as coaches to guide the improvement process in struggling schools: "CORE districts believe that the best experience-based reform experts come from within successful schools. Those experts are teachers and leaders who have demonstrated successful reform, achievement, and/or growth over time" (ibid., p. 64).
The School Pairing and Communities of Practice were launched in the middle of the 2013–14 school year. According to CORE staff, in that first year the support was light; schools were told "it's up to you, we trust you." This approach was not successful, and the CORE districts reworked the structures after a critical review from a U.S. Department of Education audit. For the 2014–15 school year, CORE centrally provided more tools, clear guidance, and training to support the work. Priority schools were encouraged to identify two to three problems of practice, develop an improvement plan based on ESEA's seven turnaround principles (listed in U.S. DOE, 2012), develop a structure for collaborative interaction, meet quarterly, and show evidence of "learning and progress." Schools in the Communities of Practice (two to four focus schools in each group) were required to identify problems of practice, come together around shared problems, meet quarterly, and run quarterly PDSA (Plan-Do-Study-Act) cycles. Together, collaborative matching, planning, and cycles of inquiry comprise the current Pairing and Communities of Practice interventions. After this initial central training, school support structures were run by the districts, with annual Peer Review from the other districts to monitor implementation.

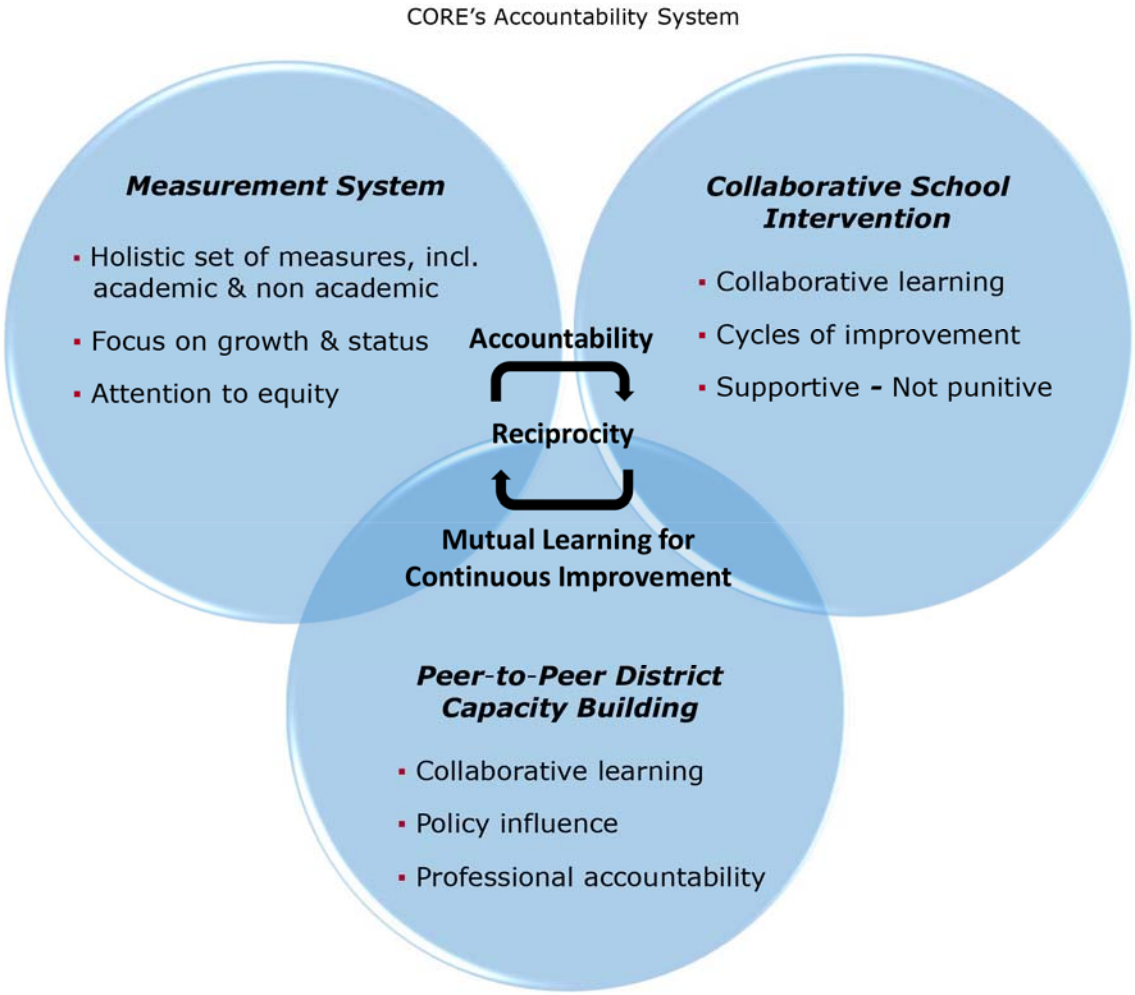
Cross-District Collaboration and Peer Review

The work of the CORE districts is supported by a nonprofit organization with approximately nine staff members—several of whom were employed by partner districts prior to joining CORE—who support and advance the work of the member districts. One important role staff play is to facilitate cross-district collaboration in order to build capacity and meaningfully engage with the MS. As such, CORE staff organized formal opportunities for this to occur, including board meetings, quarterly meetings for role-alikes—individuals from different districts fulfilling similar work roles—across districts, and regular phone meetings for key staff. Districts (in groups of three) were also required to complete a thorough self-evaluation and peer evaluation process measuring their progress against planned activities (e.g., collection of MS data, stakeholder engagement, fidelity of intervention implementation) using detailed rubrics.[xvi] After the self-evaluation, districts would review one another's work and make suggestions about how to improve implementation.

Conceptual Framework

We designed the study to examine the implementation of and interconnected relationships among the three main elements of the CORE accountability system: the MS, school interventions, and district peer-to-peer collaboration (illustrated in Figure 1). By design, these three elements rely upon accountability and mutual learning as key mechanisms to promote improvement, both predicated on the expectation of reciprocity among peers (district to district, school to school, administrator to administrator). We used this conceptual framework to guide our data collection and analysis.

Figure 1: Conceptual Framework



Research Methods

We used a multiple case study design (Yin, 2013) to gather data on educators' experiences implementing CORE's new MS and peer-to-peer interventions at school and district levels. In this section, we briefly describe our framework, research questions and methods of data collection and analysis.

Data Collection and Analysis

In each of the six CORE waiver districts, the research team conducted semi-structured interviews with central office administrators responsible for CORE-related work (*n*=41), including superintendents and cabinet-level administrators, along with district staff responsible for data, accountability, school support, curricula, and human resources (see Appendix A1). Notably, we targeted administrators with named roles in the CORE work ("data leads" and those coordinating the work related to Principles 1-3). In each district we also conducted interviews with a sample of school principals and facilitators engaged in the CORE-related intervention work. Specifically, in each district we targeted one principal each from a focus, priority, and reward school. During central office interviews, we asked administrators to identify a selection of challenged, typical, and exemplary schools in each category. We then selected randomly by school level from among the list of reported typical schools and recruited principals for interviews. In the end we interviewed 15 principals: six at the elementary level, nine at the secondary level (eight middle, one high school); and of those, four were from focus schools, six from priority schools, and five from reward schools. The majority of interviews were conducted in person, with a small subset conducted over the telephone (primarily for principal interviews). We also interviewed four leaders from within the CORE staff. We used semi-structured protocols in all interviews, which were audio recorded and transcribed.

We supplemented these interviews with observations of CORE meetings and training sessions (42 hours). Specifically, over the data collection period, we observed all regular cross-district meetings involving data and intervention leads, attended by three to five representatives from each waiver district. We also observed two local training sessions aimed at training intervention facilitators and working with paired schools. Finally, researchers gathered and analyzed documents pertinent to the overall CORE waiver and related activities (e.g., the initial NCLB waiver, meeting minutes, PowerPoint presentations given at meetings) as well as individual CORE districts (e.g., peer review reports, school-level data reports).

Through our case analysis, we sought to understand how districts implemented and responded to the CORE measurement system, school interventions, and capacity-building efforts. Guided by the conceptual framework, we first analyzed each CORE district individually, developing detailed case memos. These initial case study memos helped to specify the design and implementation of CORE activities locally and key contextual elements in each district. Next, we completed cross-case analysis, drawing on the case study memos and all transcripts, to examine how implementation varied by district and the factors associated with implementation (Miles et al., 2013). To further understand patterns across districts, we utilized matrix displays (with rows representing districts and columns representing constructs such as district characteristics and the local implementation of measures and interventions) (Miles et al., 2013). These matrix displays helped us to see patterns among multiple constructs, and our attention to alternative explanations also helped to ensure the robustness of findings (Yin, 2013). We also triangulated findings, wherever possible, among multiple respondents and data sources to strengthen the validity of our findings. In other words, overarching findings were considered valid only when corroborated by at least two other participant accounts, by direct observation, or by document review.

Several caveats are important to keep in mind. First, we examined the accountability system and the improvement efforts in their infancy, and in a time of transition. At the time of data collection, not all school-level MS data had been made public and many administrators acknowledged they were still building awareness and understanding of the MS. Second, as we discuss, given that many districts have integrated existing accountability systems and measures into CORE's accountability system or vice versa, we are unable to fully isolate the implementation and perceived effects of the work under the waiver from these prior

systems. Further, at this early stage of implementation, there were relatively low stakes attached to the accountability system, and there were questions within the districts about whether the system would be implemented in the following year. As such, we cannot fully infer how this system might play out in the face of high stakes. Finally, our sample includes only school administrators participating in the CORE school pairings or COPs and in some larger districts those interviewed represent a relatively small proportion of involved principals. The opinions they express (regarding buy-in, awareness, etc.) may not represent the views of all principals or others not involved in the intervention work.

Response to New Accountability Measurement System

Our study examined how the six CORE waiver districts understood, implemented, and responded to the new accountability system. In this section, we examine specifically the MS, asking:

- To what extent did educators value and support CORE's measurement system? How did educators understand this new MS and its component measures?
- How did schools respond to the MS and its component measures (in terms of attitudes, practices, etc.), and how do these responses vary across schools?
- What were the (intended and unintended) consequences of the use of the MS for accountability purposes?
- What were early indicators of conditions that facilitate and constrain implementation?

Findings

District and school administrators greatly appreciated a more holistic approach to measurement. Most administrators valued the MS and the use of a comprehensive set of academic and non-academic measures to assess school performance. As one superintendent explained, "The social emotional side…needs to play against the academic piece. If you have one without the other you're probably missing something." A principal in another district echoed this sentiment, "The CST isn't enough. This is a way better way, not just getting the academic but the SEL [social-emotional learning] too."

While many interviewees did not perceive the MS to be new in its entirety—many were using some to many similar MS measures prior to CORE—they generally acknowledged the value of having all of these measures accessible in one place. One district administrator called it the "umbrella index." A leader in another district similarly explained:

> When you can get all of those measures in one place and they're measures that make sense to people who use them, you get better at making decisions about what actions you need to take, how you use your resources, your dollars and your people to do that work and I think if anything, that is also linked to what we're trying to do with the Communities of Practice more at the school level.

Another leader noted that while they had been conducting climate surveys for years, "I guess maybe that's the difference. I think is that people now take a look at this and it becomes 40% of their accountability and you need to pay attention to it." By including a comprehensive set of measures, the goal was that schools would be held accountable, and recognized, for essential aspects of child development that, in prior years, were invisible in accountability systems. One district administrator reported that this measurement change in turn changed who in the system feels responsible for school improvement:

> Now that we've got an accountability index that has multiple domains and a more holistic set of measures with actionable targets, more departments that support schools are feeling that they have the agency to be able to take action to contribute to student success.

Although most interviewees endorsed the inclusion of nonacademic measures, a minority expressed concerns that such indicators could "distract" educators from the more essential work to support academic outcomes. One administrator said:

> I do think it's good that people are thinking about other factors besides academic, but I don't want them to do so at the expense of some of the academics because I think sometimes people think those things are easier to deal with, like attendance, "Oh we'll just hire a Pupil Services and Attendance counselor and we'll improve our attendance." The real work is the academics for kids and so I worry that sometimes those other pieces are distracters from what's really going to get kids to be able to be prepared and able to go to college.

By including a comprehensive set of measures, the goal was that schools would be held accountable, and recognized, for essential aspects of child development that, in prior years, were invisible in accountability systems. One district administrator reported that this measurement change in turn changed who in the system feels responsible for school improvement:

> Now that we've got an accountability index that has multiple domains and a more holistic set of measures with actionable targets, more departments that support schools are feeling that they have the agency to be able to take action to contribute to student success.

Administrators consistently endorsed the emphasis on growth over status. Administrators repeatedly praised the MS for including measures of growth in student achievement and the corresponding shifts in behavior they expected to see. One central office leader underscored the fairness of such a system:

> The growth measure … it's the only fair way really to measure because again you've got a school on this side of town, and this side of town you can only look to see how much they have grown, not compare one to the other….

A leader in another district appreciated that the MS allowed them to recognize schools that were rarely celebrated under the prior accountability system:

> How do we champion schools that are making tremendous progress? The great thing about the MS… is, you could pick any school and I could show you something they do well. We want to make sure they know that, build from that.

Similarly, several principals—particularly those in lower performing schools—appreciated the emphasis on growth. One explained, "When you're in a school that's historically failing, it's hard to see progress without that growth score." A principal in another district believed growth measures "level the playing field" for schools with underserved student populations. Others appreciated that such measures would incentivize educators to focus on more than just "bubble kids" or those scoring close to the proficiency cutoff, a response common in the NCLB era (Booher-Jennings, 2005; Jennings & Rentner, 2006; Hamilton et al., 2007; Marsh, Pane, & Hamilton, 2006).

Views about the decreased subgroup size used in MS calculations were mixed. When asked, administrators expressed general support for the smaller subgroup size criteria. In general, they believed decreasing the number of students required to constitute a "subgroup" from 100 to 20 increased the visibility of and accountability for these students. One superintendent believed the subgroup change advanced equity: "We drove the threshold in from 100 down to 20. We've exposed another 150,000 students across our collective system for somebody to be paying attention to their aggregated data. I think that's significant." Leaders in another district believed the smaller threshold allowed principals to better celebrate smaller subgroup victories: "If anything, folks are disappointed when they don't have 20 and they have the students that are 19 and below doing well and then not being represented." Two principals reported that this change allowed them to better identify and target problems. "The more you drill down and disaggregate, the better," said one principal.

Nevertheless, some administrators expressed more neutral views. One district leader noted that given the large size of schools, the change made little difference. An administrator in another district believed principals were simply expressing compliant acceptance to the change in subgroup measurement.

A few interviewees reported challenges resulting from this decrease in subgroup size, including: (a) the difficulty of obtaining 95% participation in testing for these now much smaller subgroups; (b) the challenges it imposes on magnet schools with particularly high performance (gaps between special education subgroups and the general population now appear to be quite large even though the performance of the subgroup was higher than in most schools districtwide); and (c) the instability it might create for intervention activities (with small subgroups, identification for intervention based on subgroup can shift quickly).

The school ranking system raised questions about the value of multiple versus single aggregate measures of performance. Some administrators endorsed the MS score school ranking system, which ranked schools across all CORE districts on individual measures using a decile system. One principal believed the rankings allowed her to identify schools she could contact for support in areas in which her school rated relatively weak. Others expressed more neutral positions of acceptance. One district leader explained, "I think people in public education are so used to that kind of ranking system that this is just another sort of 'you're red, you're orange.'" "It is what it is," said a principal in this district.

Yet others criticized the ranking system. One principal argued that these rankings wrongly promoted competition over collaboration. Moreover, leaders in another district intentionally de-emphasized the summative MS score ranking (a single number aggregating across all measures), noting its conflict with an accountability model intended to provide a holistic picture with multiple measures:

> The whole point of not making an index [summative MS] score or making it easy for anybody to rank schools was super intentional. This is what practitioners really didn't like about NCLB… that you line this all up, and that's not how schools work. There are nuances across that and what's really important is to look at the multiple measures and be able to make a strategic decision based on what those data points are telling you, not what the single score is telling you…. The whole point is multiple measures tell you a different story.

Developments in Los Angeles, in fact, demonstrate the difficulty of sustaining a focus on multiple measures and the deeper societal attraction to single numbers that are easy to "digest." While leaders in the Los Angeles Unified School District (LAUSD) had not publicly promoted the MS results and rankings, the online local media outlet *LA School Report* published a series of articles that could have oversimplified the

findings. The April 2016 series highlighted the lowest and highest performers based on the single ratio index with headlines such as "New data reveal best and worst of LAUSD schools" and "Stark differences for LAUSD elementary schools in the CORE accountability index" (Clough, 2016 April 11; 2016 April 26).

All districts adapted the MS to their local contexts. Rather than strictly adhering to the MS, all of the districts adapted it to fit with local values and needs, and incorporated it into existing frameworks. Notably, these adaptations included expanding but not removing indicators from the MS. Most districts embedded the MS into their own district's indicator systems and may not have even associated it with CORE internally. For example, some districts referred to their broader indicator system as the "School Quality Improvement Index" while other districts used their own label (such as "School Performance Framework"), which in some cases predated CORE. This was a conscious strategy on the part of central office leaders to build coherence and buy-in. As one district administrator explained, "Part of our strategy was to embed [the CORE MS] within the [local framework], so that we could essentially communicate that this is one in the same." As a result, not surprisingly, school principals and facilitators in several districts were largely unfamiliar with CORE's MS by name.

In these districts, CORE's MS was part of a broader data system that includes additional indicators based on a district's definition of student success and continuous improvement goals. In three districts, the system kept track of data beyond high school graduation, to include college access and readiness, with data on student eligibility and progress towards meeting college admissions criteria. In other districts with a history of reforms supporting social-emotional learning and climate improvement predating CORE, additional measures were added to the CORE's SEL and CC survey as they were believed to be "more impactful" and "stronger." In cases where additional measures were used, indicator systems often predated the development of the MS.

Most districts were still working on building awareness and understanding of the measures, particularly at the school level. Given the early stages of implementation, it was not surprising that most district leaders reported that they were still working on building staff awareness and understanding. Across all districts, central office administrators conveyed stronger familiarity with the MS than school-level administrators.

In districts with greater awareness and understanding at the school level, staff reported implementing a more deliberate roll-out led by the central office, including capacity-building support, such as professional development and technical training on using data systems. One district also expected their principals to report and explain the data to their School Site Councils. Gauging principals' understanding of the framework from a continuum of awareness to impact, a leader from this district with a roll-out plan shared:

> I would say we're looking at maybe 20% of our schools having impact [truly understanding and using the data to drive improvement]. We're looking at maybe 30% of our schools are emerging [starting to understand and use the data], and 50% of our schools at awareness [becoming familiar with the data].

The similarity of the MS to existing data management systems in some districts also complicated efforts to roll out the system and build awareness and understanding. Central office administrators believed that subtle differences in how metrics were calculated made it difficult for some schools to understand what was new about the MS and why they should not simply rely on existing measures.

Familiarity with the MS academic measures was high, but buy-in varied across and within districts. Most respondents understood which academic measures were included in the MS—SBAC results, high school readiness, and graduation rates—and could provide basic definitions. However, not everyone supported the inclusion of each of these measures in the MS.

First, the majority of districts and schools had a wait-and-see attitude with regard to SBAC test scores, as they were still making sense of the results and potential uses after the first year of test administration. In general, interviewees bought into the necessity of including SBAC scores in the accountability system and expressed appreciation for the focus on growth. However, there were mixed reactions to the utility of the SBAC for planning. Referring to the SBAC, one principal reported:

> It doesn't lie, right? So that is a fact. The fact is that 3% of our kids are proficient in math. It's like the "so what" factor. So what are we going to do about it?… For me, it's a reality check.

In one district, SBAC results were not viewed as relevant yet, especially as schools relied on other assessments that predated CORE. The district instead relied on an externally developed assessment of growth in math and reading. Several school leaders also indicated that they relied more on other academic assessments. When asked about SBAC, one principal replied that it "needs to be there" but that they needed to drill down with other benchmark assessments to monitor and support student learning such as reading skills. Another principal indicated that it was tough to think the SBAC results showed what her kids were capable of, because it was the first year of the test. The decision to consult other sources of data on achievement to guide instruction is neither surprising nor in conflict with the intent of CORE's MS. However, strong doubts about the value of state test results may weaken the potential for the MS results to guide school improvement strategies rather than simply identify schools for intervention.

Some principals also questioned the weight given to SBAC results. In high school, for example, students only take the SBAC in 11th grade. Some school leaders felt that a single test taken in a single grade level should not make up such a significant portion of the school's academic score. Similarly, central office administrators and middle school principals generally understood the high school readiness measure, but buy-in was mixed. One principal from a district that was generally enthusiastic about the MS explained its value:

> I also appreciate… the high school readiness [indicator] for us as a middle school. I like that measure because it takes into account a few different things and it gives us a sense when we start looking at this aggregating groups of who is more high school ready than others and that goes back into that work with the college readiness piece.

Within this district, the high school readiness indicator was seen as a substantive and important change for the middle schools in this district. Some, however, voiced concerns about the validity of these measures. One principal reported that there were different grading philosophies in the school: While some prefer a mastery-based grading approach, others grade based on a student's work completion or effort throughout the year. As we return to later, this same principal also feared that teachers might game the system and grade differently in response to the high school readiness measure.

While administrators at all levels appreciated the non-academic measures, some questioned their validity. In general, administrators conveyed strong awareness of and support for the non-academic measures included in the MS. Buy-in was particularly strong for the inclusion of absence and suspension rates, likely because they were already familiar indicators for most educators.[xvii]

District and school leaders also generally responded positively to the inclusion of SEL and CC measures, in large part because the districts felt these measures better reflected the realities of their schools (discussed below). One district with strong buy-in at both the district and school levels believed that the SEL and CC measures placed a spotlight on these issues in new ways. Although the district had an SEL- and CC-related survey in the past, these data were never accessible or used to the same extent. The new measures were now included in the district's dashboard and part of the accountability expectations. Among the central office leaders interviewed in this district, all had strong buy-in. One central office administrator believed there was more buy-in for SEL at the elementary school levels, where teachers were more likely to have a holistic orientation towards supporting student learning.

Nevertheless, several administrators voiced concerns about the validity of these measures: Both district and school leaders at one district questioned whether or not the SEL and CC surveys taken by the parents, students and staff truly measured the SEL/CC constructs at hand. As one district leader explained, there were concerns about whether the right questions were being asked: "The issue is a little bit more around the narrow set of questions and whether or not those are the right questions to get at the indicators." Leaders reported that they did not trust the results, but they were not sure of a better way to measure these constructs. As a result of these concerns, school leaders reported that they were not taking action based on these results. Another district believed that the CORE measures did not adequately address all SEL topics and included additional questions on the survey to compensate (e.g., using additional survey measures developed by Karen Mapp to examine parent support and school climate).

Other educators worried about the potential for distortive practices that could invalidate the MS indicators—a concern echoed by some scholars regarding the use of SEL measures in accountability systems (e.g., Duckworth & Yeager, 2015). We return to this topic later in the report.

Few administrators conveyed a detailed understanding of the specific SEL constructs, contributing, in part, to limited reported use of the measures in practice. Despite strong support for SEL and CC measures, few administrators articulated an understanding of specific SEL constructs and their measurement. Not surprisingly, SEL and CC leads within the districts were most knowledgeable about the various SEL competencies and how they are related to CC measures. Given the complexity and newness of these survey measures, it is not surprising that few administrators were highly knowledgeable about their calculation.

When asked about the SEL competencies in MS, some administrators responded by discussing climate indicators and student behavior. Describing how they use survey data in the district, one administrator explained:

> Each principal brought up that student SEL survey and what really had resonated with them in the context of the culture that they were trying to create at their schools and what they felt that they could do to create a better sense of belonging [a CC construct], or work on social awareness [an SEL construct] or any of those.

Others responded with generalities about the importance of social-emotional competencies. Still others acknowledged that they had a more superficial understanding of the concepts rather than the specific measures. One district administrator replied:

> We talk about growth mindsets a lot but we are not talking about growth mindset data. It is just everybody is saying, "Look I have a growth mindset and it is really ambiguous and vague," but we are not saying this is the percentage of kids that has a growth mindset or think this way.

Some administrators believed the lack of familiarity with and capacity to interpret the new non-academic measures contributed to lower levels of use. Contrasting educators' familiarity with using academic data, one district administrator explained:

> Fifty percent of all high schoolers say they don't have a sense of self efficacy…. If you're a high school administrator, you say, "Oh God, what do I need to do?" I can imagine them feeling real pressure to respond and doing something about it…. We have a lot of data—we don't quite know how to interpret it, we don't quite know what it means, we don't know what the correlations are…. Because we haven't been practicing teaching self-efficacy in high school, we don't know where to start … Because there was no previous measurement on strategies that actually might work, we are clueless.

The depth of engagement with SEL and CC data varied across districts and related closely with a district's history of preexisting reform efforts focused on connecting student outcomes to SEL and CC. In a few cases, administrators reported using the results to guide professional development. In general, however, many educators questioned how "actionable" the MS's SEL measures were. One lead believed SEL data were "good initial measures" but "it's hard for teachers to make instructional decisions for students' SEL needs" based on them. When asked what makes the data actionable, another district respondent replied:

> It's indicative of groups and planning strategies around groups. I think at this point the social emotional data—just it being there, just understanding what it is—is such a big task and so deeply appreciated. As I said, just the fact that it's there, it's huge.

The SEL and CC lead at one district mentioned some of the measures were useful but would like some of the staff surveys to be more frequent so that they can be used to develop actionable plans throughout the year: "I would love… specifically for the staff piece, to have it done twice a year because it's very difficult to have the survey in February. Results come out in May, and then it's not that continuous cycle of improvement." Understanding how to analyze and use data, and the content and meaning of the actual measures, proved to be a challenge to district and school administrators alike.

Concerns about (mis)alignment with other measurement systems affected support for and use of the MS. Some interviewees expressed concerns about discrepancies between the MS and measures included in their local data and accountability systems. For example, a minority of interviewees believed that the MS did not "go far enough," failing to include indicators of college readiness (e.g., A–G course completion) and other indicators (e.g., chronic truancy) they believed to be important measures of success and equity.

Others identified misalignments between the MS measures and external accountability systems, notably the state's Local Control Accountability Plans. At two districts, interviewees reported that several MS indicators vary slightly from LCAP metrics, including suspension rates, making measures difficult to understand at the school level. One administrator noted:

> One is that there are different business rules for how some of those things are calculated than how they are calculated at the district level. The simplest version of that is chronic absenteeism. In the

> district, it's set at 91%. In CORE, it's set at 90%. If a principal were to look at their chronic absentee rate, it wouldn't necessarily read the same way. A more complicated version of that is the ELL redesignation rate, it's completely calculated differently. It's actually looking at a different slice of time as well. That's literally how the metrics are calculated.

A leader in another district shared similar questions about calculation differences between the MS and the state:

> Yeah. I mean, if you get down to the weeds, just the way we look at English learners under the school quality index [MS] and the way the state defines reclassification rates. I think that's one of the biggest questions we always get from schools. Why is my re-class rate different here than what the state says? And so we have to go into that whole explanation as to why. We were prepared to do that. We had all these scripts written when we knew it was going to be the accountability moving forward, but now, if it's not the accountability moving forward, then what's the justification to tell schools that, on this particular report, we're counting English learners this way and the state is counting them this way.

The misalignment was perceived to be particularly vexing for some. For example, in one of these districts, we were told about a school that had received recognition for its work around improving EL redesignation but then scored low on CORE's measure. Describing their principals as "data savvy," a district administrator reported that principals quickly identified discrepancies between MS data and district data. She shared:

> I think that our principals are just so data savvy that the minute you put something out, we knew it. The minute we put something out, I get a call from this tiny school, like, "My suspension rate says it's 0.23 and I didn't suspend anybody." Then we have to dig into it and it turned out that it was a kid that had moved from her school, but whoever generated the data reports had reported it at her location, so it's those kinds of things that people are very savvy to.

In light of the new Every Student Succeeds Act, some administrators also raised strong concerns about the potential incoherence of these multiple accountability systems in future years. These perceived inconsistencies in some cases decreased the support for the MS.

District level administrators most commonly used MS data to manage improvement, inform resource allocation, and communicate with stakeholders. Under the CORE vision, district and school educators were expected to regularly use the MS results to illuminate potential problems and generate collective inquiry and action for improvement. At the district and school levels, we heard of widespread engagement with the MS results, particularly to guide (a) improvement planning, (b) resource allocation, and (c) stakeholder communication. First, leaders used the data to inform improvement efforts. In one district, officials used the holistic data reports to reevaluate their view of school performance. An administrator shared an example of how the data led to reassessing leadership effectiveness in schools:

> I think we have one school that's a classic example… this leader, I think had been perceived for a long time as really effective… he's a really super nice guy and adults generally felt good, school generally looked good.… that teachers thought it was an awesome place and that was the general reputation it had.… But then all the other indicators were orange and red. It became clear. I think the narrative that jumps out in front of me is that "Oh this was a good place for the adults in the

> school, right, and not for the kids," and what does that mean about the leadership or what's needed there? Certainly we want places that are good for adults. I don't think you can get good for kids without good for adults, but it illuminated some stuff that wasn't, in a way, that our prior system didn't quite demonstrate.

Another district viewed the MS data as central to the way they managed improvement at the district and school level. This district planned to integrate the SEL survey domains into the student report cards in the elementary grades. These results will then be reported back to the schools at the grade- and school-level. Yet another district used the SEL data to plan professional development for school leadership teams. The SEL and CC lead and her team planned to hold meetings with school teams (e.g., administrator, five teachers, and classified staff) twice a year to look specifically at this data. She described one meeting:

> They went through the survey results. We walked them through it. We put some questions out there to help them process. They looked at celebrations, they looked at areas of growth. Then based off that they came up with action steps…. Then they take it back to their school sites and they figure out how. For example, if it's sense of belonging for students, like they scored low on that piece. Looking at student engagement activities that they could do at their school for kids to feel connected to the school.

Second, in all but two districts, administrators reported using the MS to identify resource needs, use, and effectiveness. One district tied all the CORE data to their district data dashboard and produced an "at-risk data report" for each school. Based on these data, the district assigned more staff to focus on improving results around high school readiness in the middle school. As a result of the increased flexibility in funding due to LCFF, schools were able to shift resources into the areas of need identified by their MS data.

Another district used the data for discussion and analysis with instructional leadership teams and the superintendent. In meetings, central office leaders would ask principals to reflect on MS results and how they were guiding school improvement plans, for example: "When you say you want $50,000 for something, which indicator are you using to make that argument and which indicator are you going to use?"

Finally, several districts reported using MS data to communicate with school leadership and parents. One district created a more succinct, one-page report with MS data because they found the CORE reports to be too difficult to use. The one-pager included all of the same data as the full MS, but also included the four SEL domains with school-level averages on the Likert scale (e.g., 4.2 score on growth mindset).

In addition to training their principals and school leadership teams, another district also engaged in parent education about their district's framework and the CORE MS. A parent–community specialist began trainings at the end of this year with key groups of parent leaders across the regions. The objectives of the sessions were to enable participants to walk away understanding the origins of the framework and the value of the indicators, what kinds of strategies can support improvement on these indicators, and how they can support the school to improve.

School principals most commonly reported using the MS for planning and goal-setting. As with other measures, principals across districts used academic data to create school improvement plans and set targets for students. For example, in one district, principals used data to lead cycles of inquiry and explore root cause analysis with their leadership teams as well as their School Site Councils (SSCs). Presenting these data to SSCs has caused principals to grapple with how to present a complex dashboard to

a parent audience. Principals described their aim to present data in a way that created buy-in, encouraged feedback, and promoted good questioning.

One principal in this district reported, "I have it [the MS report] hanging on my wall, in my office. I know the data points and what we're being held accountable for improving on." However, another principal shared that he is not using the district data as this principal believed that a separate, internally developed system was more useful.

At another district, school leaders embedded CORE's accountability system into their goals, which were discussed in quarterly meetings with school supervisors. School sites have designed new systems to address MS. One school site created committees to respond to particular types of data—attendance, high school readiness, SEL and CC survey—in which teachers developed programs and activities to guide improvement. For example, to boost attendance they developed a host of incentives (prizes for perfect attendance) and assigned mentors to chronic truants. For high school readiness, they planned to organize parent nights, inviting guest speakers, using educational posters all around campus. The same school also assigned staff to work on EL re-designation and suspension rates, using LCFF funds to pay for an EL coordinator.

Similarly, a principal in another district used the MS data with teachers and other support staff for instructional planning and intervention. This same principal also found the rankings in the MS reports helpful and used them to identify potential learning partners. Similarly, another principal in the same district shared the value of cross-site learning:

> Well, it's not just about how they did, but it opens up, I'm able to look at the MS. I can see the ranking, and I can say, "You know what I've been noticing? Your school is ranked here, what are some of the action items? What are you doing?" I can ask the principal, that generates conversation.

This principal found the high school readiness indicator as a good tool to communicate with students and parents. She uses it in her daily announcement to students and weekly letters to parents.

Several district and school administrators expressed concerns with the potential for distortive responses to the MS in higher stakes settings. While the potential for authentic learning was great, administrators across districts commonly cited the concern that some of the MS metrics could incentivize distortive behavior. Their concerns involved superficial approaches to improving performance on various indicators—approaches that would not necessarily alter the underlying behaviors likely to produce genuine improvement, particularly when high stakes set in. For example, one principal worried that teachers might produce different student grades in response to the high school readiness measure. Principals in three districts expressed similar concerns about the potential for "gaming" suspension measures or taking superficial approaches to reducing suspension numbers rather than underlying behaviors. One administrator shared his skepticism:

> I've been to a lot of schools where the culture has been horrible and the expectations for behavior are really low and then they have zero percent suspension rate. I think people either just send kids home and don't capture it as suspension or they are just ignoring behaviors that aren't acceptable. Because they know that is a place to score.

Others expressed concerns about distortive responses to the SEL measures (a concern echoed by some scholars, e.g., Duckworth & Yeager, 2015). As noted earlier, one central office administrator explained:

> Then how do you prevent gaming on the surveys, which I haven't really found places that have used surveys as an accountability metric…. The minute you attach an accountability label to it, people just want to know what are the questions you're going to be asking me, and how do I make sure we hit those, which just defeats the whole purpose of getting honest answers on surveys.

Another administrator in the same district echoed the concerns about survey-based measures used for accountability purposes and the responses they can generate:

> I think that once something becomes measured… it's a survey… I'll give an anecdote instead. The principal asks the question, "Will you tell us which questions contribute to this [rating] so that we can get the right answer?" Once you make an accountability, it's like you want it [survey results] to read well, not necessarily accurate. That's just an incentives thing…. The opposite can happen too where parents might be mad about something and they might be like, "Oh yeah, this is how I'm going to get the principal fired by…" You know what I mean? I don't know if that happens, it's just sort of like a macro question that I have about the surveys.

Other educators worried about the potential for distortive practices that could invalidate the MS indicators. One district leader explained:

> How do you weight surveys in an accountability system and what are the appropriate thresholds for response rates? Then how do you prevent gaming on the surveys?… The minute you attach an accountability label to it, people just want to know what are the questions you're going to asking me, and how do I make sure we hit those, which just defeats the whole purpose of getting honest answers on surveys.

While our findings on this point are speculative and we did not uncover evidence of such responses at this early stage of implementation, it is still worth noting how unsure some administrators were about the prospects of some of the MS results driving true learning and school improvement.

There were few to no formal consequences associated with MS results. At the time of our study, most interviewees reported that there were few or no formal stakes tied to the MS results yet (although many understood that in future years, schools would be re-identified for interventions based on these results). Thus far, most districts did not have explicit performance expectations or standardized requirements tied to data. Instead a common expectation was that data would be used to inform site planning and goal setting as well as justify resource allocation and interventions. For example, one district used the accountability data to determine the budget and autonomy of each school site based on rankings. Schools at the top of the ranking received the lowest amount of additional funding but the highest level of autonomy from the district; conversely schools at the bottom received the highest amount of additional funding but the lowest level of autonomy.

At another district, schools were expected to use the MS to set goals based on their current levels on each metric, both with all students and with relevant subgroups, and demonstrate at least one level of improvement each year (on a four-level scale). When asked about feeling pressured, several principals explained that they felt internal pressure driven by their sense of responsibility to their students and

communities rather than external pressure from the district. One principal felt that there was an expectation and "healthy competition" among school leaders to find ways to do better and improve their "score." While there appear to be few negative consequences for low performance, high-performance schools continue to be rewarded with additional funds or recognition.

Nevertheless, many educators believed active public dissemination of MS data—which was not occurring yet in all CORE districts[xviii]—would serve as a powerful incentive in this system. Some believed that once the public had access to and a better understanding of the MS, they would hold the district and schools accountable for student growth and performance. Some viewed this public pressure as a positive incentive to improve. Others highlighted the potential negative consequences of poor performance. One principal lamented that the school's low results on the MS made it difficult to attract prospective families. He noted that the MS brought more public attention to measures such as attendance and suspension, which were not publicly reported before. For schools that serve seriously disadvantaged students, the expanded measures do not "paint a pretty picture" of a school's performance, even if the school is doing very well with students who bring in serious challenges.

Implementation of School-Level Interventions

CORE's accountability system is built upon the assumption that both expanded access to data and information, and collaborative inquiry will result in school improvement by encouraging schools to rethink the status quo and try out new approaches. In this section we ask:

- To what extent did educators value and support the CORE school interventions (Communities of Practice and School Pairing)?
- How have districts implemented the CORE school interventions (Communities of Practice and School Pairing)?
- What factors facilitated or constrained implementation?
- How have schools responded to the new intervention system (in terms of attitudes, practices, etc.), and how did these responses vary across schools?
- What were the (intended and unintended) consequences of the use of new interventions?

Findings

Districts and school leaders supported the CORE intervention model—particularly the focus on support over sanctions—but expressed concerns about implementation. Most district and school staff believed that the intervention model was better suited for school improvement than NCLB sanctions. Echoing the message from CORE leaders and documents, district and school leaders consistently noted and appreciated that accountability was intended to be used "as a flashlight not a hammer." One central office leader described the system as "not about putting the red scarlet letter, it's about providing supports". She went on to explain:

> The accountability efforts are less like a hammer and more of a way to improve on our efforts in a more constructive way. And say the word hammer kind of referring to maybe the perception of what NCLB held us accountable at. Looking at standards with a cut-off line and you're either yay or nay. So either one way or another. So this really allows us to look more discretely and closely at groups of students, how they're doing and what efforts are being put in place to help to support them.

A leader in another district concurred:

> It wasn't viewed as a punitive thing. It was more you've been identified, but here are resources to help you. We're going to pair you with other schools. We're going to create communities of practice, so I don't know how it actually played out in terms of the obstacles and in terms of the implementation, but I think just the philosophy behind it was much more positive than a negative thing.

At the time of our interviews, buy-in appeared somewhat positive across all districts. As one district administrator stated, "I think it is absolutely the right thing to do. I think it's absolutely positive for schools." In practice, however, district and school staff acknowledged that the interventions might not see consistently high results across schools. In one district, an administrator noted that "it depends on to what extent we're able to keep it robust, and up, and going."

Overall, while district administrators expressed their support for the Communities of Practice and School Pairing work in theory, they believed that the success of these interventions depended on school-level buy-in and capacity. Interestingly, when we examined the buy-in across districts, it emerged that those districts that had schools involved primarily in one intervention (either Communities of Practice or School Pairing) were more positive than districts that had schools involved in both interventions. Districts with very few interdistrict pairings (discussed below) also reported more positive beliefs and buy-in to the CORE interventions. It may be that engaging in just one intervention and/or utilizing only intradistrict pairings simplified the CORE intervention work and thereby influenced understanding and buy-in.

According to administrators, CORE interventions are designed to promote mutual learning, capacity-building, and network development. Overall, district and school leaders reported three intertwined main purposes of the Community of Practice and School Pairing interventions: mutual learning, building capacity to engage in continuous improvement, and networking.

In four districts, the theme of **mutual learning** was most salient. Interviewees in these districts believed the purpose of CORE interventions was to encourage the sharing of ideas and successful practices among schools in both the Communities of Practice and School Pairs. Moreover, most interviewees interpreted mutual learning as a shift in the overall tenor of improvement efforts: They often contrasted this mutual learning approach to prior prescriptive, top-down reforms. As one superintendent stated:

> In the NCLB days you had to be determined… you're bad, you're in trouble, we're going to send somebody to fix you, kind of thing, versus a CORE approach of matching schools that have similarity with demographics but [are] dissimilar in their outcomes. How can we help each other? What can I learn from you? What can you learn from me? [It] is a much more powerful model.

While this purpose came up the most frequently in interviews, respondents also acknowledged challenges in implementing this vision. For example, one district administrator shared that although the district sought to emphasize the reciprocal nature of learning (by creating "a kind of respectful, mutual learning space") this vision was not always realized.

Three districts believed that a key purpose of the CORE interventions is to **build capacity to engage in continuous improvement** activities, such as cycles of improvement. In the words of one superintendent, "I don't want you to help them, I want you to help them get better." That is, these interventions could help districts and schools learn how to solve their problems, rather than learning

solutions. By pairing improvement protocols and a culture of inquiry with access to new ideas and perspectives, schools may have been better prepared to engage in deeper mutual learning.

Building capacity for continuous improvement also involved changing the school culture to a culture of safety and innovation, in which failure is an acceptable and necessary part of progress. As an administrator in one district stated, the culture of intervention schools should "allow [teachers] to safely practice and give them effective feedback that's going to change their practice," or, in the words of an administrator at another district, "[give school staff] license to really try to uncover and understand what are the issues that are impacting the school." Through the Plan-Do-Study-Act cycle prescribed in CORE's Community of Practice intervention,[xix] schools were encouraged to identify challenges, search for innovative solutions, try out these new ideas, and assess their effectiveness within the particular school context. Although similar to the purpose of mutual learning, the capacity-building purpose promotes trial and error and emphasizes the importance of contextual fit over the sharing of "best practice."

Interviewees at three districts believed that the relationships or **networks**, developed by encouraging interaction and learning among schools in Communities of Practice and School Pairs, were in themselves a main purpose of the CORE interventions. Indeed, these relationships played an important role in facilitating mutual learning and continuous improvement activities. As one district administrator shared, "I think CORE's mission really is to develop a truly collaborative networked improvement community that is pushing each other's ideas, getting each other's feedback, creating a space where districts can learn."

Relationship building and networking, in these districts, were thought to facilitate the sharing of innovative ideas and were also intended to allow for reciprocal accountability among schools. These lofty goals, however, were not uniformly realized across all schools and districts. Next, we describe the variation in implementation of CORE interventions across districts and the challenges experienced.

Districts adapted the scope and content of interventions. Districts varied substantially in the scope of their interventions (that is, whether they utilized CORE interventions in all schools or just CORE-identified focus, priority, and reward schools) and the features of their local intervention system (e.g., the extent to which activities and strategies were adapted at the district level).

Most of the CORE districts already utilized some form of school grouping, like Professional Learning Communities or principal supervision groups, with the intent of promoting cross-school learning. The extent to which these structures had encouraged learning prior to the beginning of CORE interventions is unclear. When the CORE interventions were rolled out, however, districts varied in whether the CORE interventions were restricted to only CORE-identified schools (versus all schools) and whether the CORE interventions were integrated with existing cross-school collaborative and/or improvement structures (versus held separately).

For example, two districts rolled out the Communities of Practice intervention to all district schools, grouping schools together and utilizing the Plan-Do-Study-Act cycle (or a derivative). In the words of one district administrator, "We have… [X] schools that were focus schools. By the time we looked at it, it was like, 'Let's not hold that work separately. We're going to implement the strategy across all of our schools.'" These districts chose to utilize Communities of Practice across all schools in part because the CORE intervention resonated with their prior practice and in part due to strong district-level buy-in for collaborative improvement efforts. Interestingly, both districts intentionally used alternative terms (e.g., network, cohort) to describe the groupings and intervention. As such, schools in these districts had mixed understandings of whether or not they were involved in CORE interventions versus local interventions. An

additional district allowed non-CORE-identified schools to choose to participate in Communities of Practice based on preference, essentially facilitating the ground-up creation of improvement communities.

In the remaining three districts, Communities of Practice and School Pairing were treated as distinct from existing district reform efforts. For example, though one district grouped schools regionally and facilitated regular principal meetings as well as cross-school observation, the Communities of Practice was considered a separate intervention (and non-CORE schools did not utilize the Plan-Do-Study-Act cycle). A few administrators expressed concerns that the CORE interventions had become increasingly isolated and focused on compliance, primarily to Title I spending restrictions, rather than a central part of the school improvement work. As such, the scope and integration—and possibly adaptation—of CORE interventions at the district level may have influenced implementation efforts.

In addition to varying the scope and integration of CORE interventions, districts also chose to adapt the intervention and select prescribed elements locally. First, CORE developed the Communities of Practice and School Pairing interventions with a few key differences. Both interventions required that schools were identified for participation based on past performance, met regularly with collaborating schools, worked with a district facilitator to identify a problem of practice and develop school improvement plans, and utilized a subset of Title I funds to implement these plans.

CORE created the Communities of Practice, however, with a slightly more prescriptive approach: These schools were required to engage in a set of three Plan-Do-Study-Act cycles each year, while Paired schools were afforded more autonomy to determine their improvement activities over the year. Notably, one district chose to implement the more robust Plan-Do-Study-Act cycle in Paired schools as well. In contrast, the central CORE staff was more centrally involved in matching Paired schools together (and less so in grouping schools in Communities of Practice), sometimes across district borders. Due to challenges experienced in implementing cross-district School Pairings (described below), a few districts chose to match their priority schools with alternative reward schools within the district.

Interestingly, most schools across all CORE districts utilized cross-school walk-throughs (i.e., structured observations of classroom instruction by personnel from both the host school and the partnered school(s)) as a central part of their improvement plans. Already a common practice in all CORE districts, and referred to in local terms such as instructional rounds and continuous improvement visits, these walk-throughs were intended to promote reflection, sharing, and mutual learning. Districts identified varying levels of success in aligning existing walk-through practices to the Communities of Practice work versus serving as an add-on practice.

District and school administrators identified four main challenges affecting implementation: validity of identification, fit, reciprocity, and capacity constraints. Despite relatively positive perceptions of the reform design, interviewees identified several challenges that they experienced in implementing CORE interventions, including multiple concerns about the identification and matching of schools as well as time, capacity, and resource constraints.

Interviewees expressed concerns regarding the **validity of identification** of schools for CORE interventions. As noted above, schools were initially identified as focus, priority, and reward schools in the 2013–14 school year. In subsequent years, although a few new schools were dropped from focus or priority status based on student performance, few focus and priority schools were identified and reward school classifications remained fixed (new pairings were expected to occur with data from the full MS in 2016, which did not happen due to changes in federal legislation). As a result, district and school staff shared

concerns about whether the schools selected were truly in need of improvement and if reward schools were best positioned to provide support, due to concerns about both the timeliness and validity of selection criteria.

In the case of priority and focus schools, some districts also questioned the use of School Improvement Grant (SIG) status as a criterion for CORE inclusion: "Most of our priority schools never would have been identified to begin with as they were identified solely because they were SIG schools." Moreover, some district administrators felt that they had additional schools in need of improvement that were not identified by CORE because of the relatively large proportion of low-performing schools in other districts. In their view, several schools that could have benefited from School Pairing and Communities of Practice were excluded from participation.

Another major concern was **fit** in the matching of schools into School Pairing and Community of Practice relationships. First, priority and reward schools were primarily matched based on having similar demographics. As a result, interviewees expressed concerns that matched schools had much different contexts, despite seemingly similar demographics. While these concerns were particularly salient at the school level, district level interviewees believed that school contextual differences were of less concern. One superintendent shared these concerns, saying:

> Every school tells me, "Well, we've got special kids." Why are these kids special? Well, these kids are EL or these kids come from this side of town or that side of town. They are in poverty, they are in affluence. Every school has its unique demographics, so there is this sense that nobody else could inform my work and part of that frankly is fear, too. We have to create ways to get beyond that and say, "I'm not saying you are bad, let's just learn from each other and go visit that school. Tell me what you learn… and how that affects our practice back here."

Interviewees also reported wanting additional autonomy to determine pairings. In the words of one district administrator, "There were times when the forced pairing, where you have to stick with a school over time, didn't actually resonate. I think contexts matter, a lot." On the whole, principals in both Communities of Practice and School Pairing believed that cross-school collaboration functioned best when schools shared not only demographic characteristics (along with other contextual factors, capacity, and personality match) but also were well matched on school challenges and successes. In two districts, principals believed that they would be better served by selecting their own collaborative partner schools by targeting a school's specific areas of need.

Overwhelmingly, interviewees stated that interdistrict School Pairings were costly, time consuming, and ineffective. In one case, a reward principal described the interdistrict pairing as "a joke," stating that the pairing lacked accountability since the schools were supervised by different districts. In another district, a priority principal emphasized that differing geographic and district policy contexts impeded collaborative work in their interdistrict pairing. District administrators stressed that pairing across districts was also costly and time consuming due to travel. As this administrator shared, "It was very costly and cumbersome, because we were paired with schools that were pretty far away… it was a little too prescriptive, time consuming. Over time, it felt like more work than enhancement and facilitative work." As a result, three districts modified their pairings to provide intradistrict pairings for all priority schools.

Reciprocity also represented a challenge to implementation. The intent of the School Pairing intervention was to allow for two-way, reciprocal learning between the priority and reward schools. In some cases, reward principals reported learning a great deal from their paired priority school. For example, one

reward principal shared, "We're looking at their [the priority school's] practices and how those practices can be brought back to improve our [reward] school in the areas where we feel they're doing well." In other cases, however, the expectation of reciprocal learning was called into question. To begin with, as this reward principal shared, not all schools were clear on the specific roles of priority and reward schools in School Pairings: "It was never really clear or articulated what our role was as a reward school."

Moreover, some interviewees noted resistance to learn from schools perceived as poorly performing. As one principal shared:

> They didn't really look to us as like, "Oh, you guys have found some success." They liked what they saw, but they never implemented a single thing we suggested. They were, sort of, looking at us like we were equals and we certainly didn't feel that way. We went to their school and thought it was a horrible mess.

This tension was so widespread across districts that one school district changed the terminology from reward school to partner school. The belief that paired schools were not equipped to reciprocate feedback and ideas may have interfered with cross-school mutual learning for improvement.

Interviewees also identified a set of typical **capacity constraints** to implementation, including limited time to engage in collaboration and inadequate funding to support the actual work of school improvement (since intervention funds were reserved for "trying out" new ideas, rather than implementing school-wide programs). Besides time and resources, districts also struggled to provide trained facilitators to push forward Community of Practice and School Pairing work. Notably, the quality of facilitation arose as a concern in all districts.

First, the role of the facilitator appeared to be unclear and inconsistent across districts. Centralized training was not provided to facilitators in all districts and, in most districts, existing principal supervisors took on facilitation in addition to their current duties. In some cases, this meant that facilitators were not fully committed to the CORE intervention work. For example, one district administrator shared that facilitators were minimally involved in School Pairing and Community of Practice meetings and did not properly review school plans. As this administrator stated, "If the [facilitator is] not interested in the school plan, I think that's problematic. If it's not a good school plan, I think the [facilitator] ought to be having conversations with the school about that." In another district, an administrator shared that the facilitator's role was primarily logistical—ensuring that schools met regularly, completed the school plan and related forms, and that funding requests were appropriate. A lack of training, along with the many existing responsibilities and priorities of principal supervisors, may have contributed to a compliance orientation towards the CORE interventions mentioned by several interviewees.

While interviewees' perceptions varied widely in terms of facilitation quality, most districts emphasized the need for facilitators to take a more hands-on approach: utilizing Socratic coaching methods, reviewing school plan content critically, and following up with schools to ensure that planned collaboration and activities were realized. In sum, facilitation was not as substantive as hoped and may not have optimally facilitated learning among CORE intervention schools.

Respondents reported fewer examples of deep learning resulting from interventions. Interviewees expressed a range of perspectives on the effects of CORE interventions at the school and district levels. Some districts and schools reported powerful learning, while others gleaned little from these collaborative interventions. In particular, some districts questioned the appropriateness for such relatively

"light touch" collaborative interventions to solve chronic performance problems in schools. As one district administrator stated, "The pairing work, we gave them guidelines that we expected them to meet a minimum of three times. I don't know if that's enough for anything to matter in the long run…." Indeed, another district administrator expressed concern over the apparent intractability of performance problems at some troubled schools, saying, "[In a very low performing school in the district], we've had a really hard time finding a… leader who really wants to embrace it because of all the challenges there."

Administrators shared several examples of learning achieved through the CORE interventions, such as picking up "best practices" from other schools to facilitate their implementation of existing curricula and programs. For example, one reward principal shared that they had picked up several logistical processes from their paired priority school to "make our special education program more compliant." In essence, this kind of learning involved error correction; recognizing that Individualized Educational Plans were not being submitted in a timely manner, the reward school learned how the priority school managed their flow of paperwork.

While these superficial learnings and changes were common and, at times, quite useful, interviewees provided fewer examples of deeper inquiry directed towards continuous improvement. One district administrator shared that school principals involved in CORE interventions were learning a basic step in reflective inquiry: "learning how to ask one another questions and they're taking hard lessons away." District administrators echoed support for a gradual shift towards inquiry-oriented organizational learning, which might lead to eventual improvement. One district administrator shared:

> I think the intervention work is helping us hone in on our skill set of using a cycle of continuous improvement to look at both the implementation and the impact…. How do we help them identify where those successes are and be really super mindful and explicit about why they think those are happening…how do we help them identify those areas of challenge and help them figure out why those are still areas of challenge… what are the practices that may have had longevity at their schools?

While the ultimate outcomes of the CORE interventions remain to be seen, these perceived effects point to a shifting focus from quick, prescriptive school turnaround towards gradual, locally led, inquiry-based organizational change.

Cross-District Collaboration

Similar to its school-level corollary, district peer-to-peer collaboration is a central element of the CORE partnership, intended to build district and school capacity to engage in CORE's accountability system and improvement efforts and to promote professional accountability, mutual learning, and collaborative inquiry among the member districts. In this section, we ask:

- To what extent did district administrators value and support the CORE-led peer-to-peer collaboration?
- What were factors that facilitated or constrained collaboration?
- How did administrators respond to peer-to-peer collaboration?

Findings

District administrators supported peer-to-peer collaboration as a means to promote mutual learning and capacity-building. Through the CORE partnership, role-alikes were brought together to help design and implement CORE's accountability system. As the accountability system moved into implementation, the collaboration pivoted to focusing on professional accountability (through peer review), learning among districts, and working together on implementation challenges. Many district leaders supported peer-to-peer collaboration as the primary purpose of the CORE partnership and a crucial step in building capacity for continuous improvement. As one district administrator shared:

> I think CORE's mission really is to develop a truly collaborative network improvement community that is pushing each other's ideas, getting each other's feedback, creating a space where districts can learn about the work that is happening in the different districts around these principles that are of importance for moving student learning.

Similar to the school-level interventions, interviewees believed that the purposes of district-level collaboration were to promote mutual learning and develop networks. District administrators with unique roles and those in smaller districts especially appreciated the opportunity to engage with role-alikes, as one interviewee explained:

> You are like the one and only person who does that work in your district and you don't get to engage and exchange around ideas with other people leading similar work in districts. That was the big like, "There are other people, like me, doing the same stuff and we can talk to each other about what some of these challenges are, and bounce ideas off each other."

Additionally, district administrators supported their enhanced role in designing CORE's accountability system and leveraging the CORE partnership to influence state policy. For many district administrators typically involved in implementing state policy, CORE provided a valuable opportunity to contribute to policy design.

Formal activities facilitated informal collaboration, which was perceived as more powerful. Over the course of our year of study, the districts met together regularly. This included meetings during which CORE staff facilitated discussions of specific data metrics and of implementation successes and challenges and an in-depth peer review process. Although district leaders and administrators regularly participated in formal administrative activities (such as board meetings, quarterly meetings on each CORE principle, the peer review process, and regular phone meetings for data leads), they tended to prefer, and value more, the informal activities (such as contacting other CORE district administrators in person between meetings, or by phone or email as issues arose). Nearly all of the Superintendents reported that they routinely called and texted each other to consult on emerging issues. Similarly, district administrators leading CORE work reported reaching out to the CORE community when working through local implementation challenges. As one district administrator shared:

> What's happened with CORE is that now we're routinely… shooting out messages [to district role-alikes]: "Hey we're wrestling with this issue. How are you guys dealing with that?" There's a cross sharing that's really been… I'm going to call it a widening and a bigger circle of collaboration than we ever had before.

Formal collaboration activities necessarily facilitated the creation of this network, while providing the time, space, and climate to promote relationship building among role-alikes. Nonetheless, interviewees in all districts reported that they valued informal over formal collaboration. As one district administrator stated, "If I'm going to reach out and call someone I'm most likely to call them not because I'm required to…. [A] relationship is something that happens over time and we continue to talk about a whole variety of things." In contrast, district administrators stated that quarterly CORE meetings were less helpful, with too much "fluff stuff" and compliance activities (particularly concerning the peer review process). Another administrator noted the technical nature of discussions at times, saying:

> I think some of those activities… that we've engaged in when we're together have predominately been specific to getting something done. What are the questions that need to be included in the survey? How are we going to count EL redesignation?

Notably, district administrators stated that CORE had responded to their concerns and provided more time in recent meetings to allow districts to share their experience with implementation and to learn from one another. As a district administrator stated:

> I thought this last meeting was one of the better ones because it was really the first opportunity I had to just really sit and listen to how people were operating COPs and pairing and to learn from them, and to ask questions and for it to just be more of an informal dialogue to really begin to learn from one another.

In part, interviewees may have valued informal collaboration over formal activities because of the content of formal meetings: In the early years of CORE, district administrators were involved with designing and rolling out CORE's accountability system, which may have necessitated more focused discussion rather than learning opportunities. Now that districts are primarily involved with ongoing implementation, future formal collaboration may continue to include more time for cross-district reflection and learning.

Districts were still working to develop authentic professional accountability. Part of the CORE theory of action rests on the assumptions that member districts will hold one another accountable for implementing CORE's accountability system and helping one another improve. CORE's peer review process is designed to structure professional accountability, by mandating that districts (in groups of three) complete a thorough self-evaluation and peer evaluation. In most districts, interviewees remarked that the peer review process played out as a compliance activity, described by one interviewee as "check, check, check." Another interviewee described the peer review process as "basically grading someone's paper" with "very, very minimal" conversation.

The structure of the peer review process appeared to hinder mutual learning and professional accountability. A district administrator explained:

> It was frustrating, it was cumbersome, it took a long time. I didn't walk away feeling like, "Now I can think about my work differently." I think that was a function of the way that it was designed, the structure that we were working with, more than it was what anybody wanted.

Further, because districts were grouped in triads, they were not involved with reviewing all districts in the partnership. One district administrator noted the possibility of inconsistent peer ratings, stating that once she viewed other districts' peer review reports:

> When I looked at their evidence, I was like, "That's not a rating of a four or that's like a rating two." I would turn to my colleague and be like, "What do you think about that evidence?" They'd be like, "Well, I don't think that meets a four level but now it's too late."

This tendency to focus on compliance is not surprising, but was dissatisfying to some district leaders, who then pushed to go deeper with the collaboration to "dig into to some nitty-gritty problems of practice at a district level." To this end, the CORE districts are now strengthening their commitment to one another by focusing on a specific, measurable goal for cross-district school improvement and harnessing collective resources to do so as a Networked Improvement Community. In the spirit of continuous improvement, the districts learned from their successes and challenges in waiver implementation and are organizing with more rigor in this next phase.

Achieving reciprocity was challenging. Reciprocity among administrators from different districts is an important precondition for mutual learning and professional accountability. That is, district administrators might be more likely to learn from and listen to one another when they believe that their peer districts have something to teach and a legitimate role in holding them accountable. In practice, achieving reciprocity was a challenge.

First, as with school intervention work, district administrators expressed concerns about whether learnings from other districts were applicable to their context. A district administrator shared, "It always felt like maybe this is our own struggle and there's no way any other district understands our struggle, which may not have been the case but often felt like that when we went to those [CORE] meetings."

The peer review process faced similar challenges, as administrators remarked that the wide variation in district size influenced their implementation of the accountability system. As a result, a small district may have done a good job with parent engagement by bringing 60 parents to a meeting, while a large district might expect to see many more parents engaged. As a result, comparing rubric ratings became complex and may have undermined the review's accountability aim.
Moreover, challenges to reciprocity threatened mutual learning across districts. Administrators in several districts stated that they felt that they were "further along" in implementing CORE-related activities. As a result, in the words of one district administrator:

> We have felt more like the teachers of this…. We are informing other districts of how this work looks in [our district] but have had very little reciprocity. We don't really hear from other districts about specific strategies or different implementation pieces, best practices that have worked for them that we can then use on our own…. We have a similar population of students across some of these districts but it just feels like either our ship has sailed long ago or some of these districts are starting on some of this work… that there has been little opportunity to really be informed by what other districts are doing. I think we have definitely provided information, instead, to the collaborative.

While such relationships may be beneficial in the short term, longer term engagement and continuous improvement depends upon reciprocal, mutual learning among districts.

Peer-to-peer collaboration appeared to result in useful pragmatic problem-solving. Despite the challenge experienced in achieving reciprocity, interviewees reported learning through district peer-to-peer collaboration. Much of the learning that took place, as captured through interview reports and

observation, concerned solving technical problems and developing messaging for rolling out CORE's accountability system.

Each district produced a variety of examples of learning that they had gained through peer-to-peer collaboration. For example, district administrators discussed challenges in using technology during the roll-out of SBAC testing, , managing relationships with data platform vendors, designs for useful data reports, and metrics for measuring and strategies for improving social-emotional learning for students. District administrators largely felt that they had gained valuable strategies in learning best practices from others in order to tackle these and other problems. District administrators also reported that their districts benefited from discussing messaging and common language around CORE's accountability system. Examples included discussions about engaging with and explaining the accountability system to parents and providing assistance to school personnel to read and use the MS data reports.

Peer-to-peer collaboration also engendered deeper reflection and learning in a few instances. One district, for example, reported looking into how other districts structure their central office departments, in an effort to improve their own organizational structure. An administrator in another district indicated that CORE district administrators, as a group, reflected on the progress of the accountability system and meaning of the CORE partnership. As this administrator shared, "All the data leads met, and I think we had a very good discussion about what does it mean to still be in a collaborative together?… You know, what are the benefits and how do we proceed?"

Conclusions and Implications

In summary, district and school administrators reported overall strong buy-in for CORE's accountability system. In particular, interviewees endorsed the measures included in the MS for the enhanced focus on SEL and academic growth as well as the focus on addressing smaller subgroups to promote equity. Similarly, interviewees supported the new accountability focus on support over sanctions, which were communicated through the peer-to-peer collaborative interventions.

Despite strong buy-in, interviewees expressed concerns regarding the implementation of several aspects of the accountability system. Regarding the MS, some interviewees questioned the validity of non-academic measures and conveyed low levels of understanding of the SEL measures. District administrators also questioned the validity of school identification for inclusion in interventions and noted challenges in matching schools based on demographics.

In terms of roll-out, districts adapted the accountability system substantially. For example, some districts implemented the interventions across all schools in the district, while others utilized them only with CORE-identified schools. Some districts combined the MS and interventions with existing accountability systems, while others used the accountability system as an add-on.

Across all elements of the CORE work, the perceived effects indicated that implementation was still a work in progress. That is, while administrators reported using the MS to engage in planning, goal-setting, and managing resource allocation and improvement, they reported that interventions and district collaboration resulted in technical problem-solving more than deeper learning.

Four key factors appeared to affect the accountability system's implementation and explain the variation observed across and within CORE districts. First, each district came to the CORE work with a different history of prior accountability and intervention efforts. Some districts had administered student

surveys for many years, and a few had detailed accountability systems already drawing on SEL and CC data. In contrast, a couple of districts had historically primarily used only the state accountability measures and related interventions. It appears that those districts with a strong history implementing accountability and intervention systems similar to CORE's accountability system were more likely to express strong understanding of the MS and use of the interventions. These findings point to the steep learning curve in understanding new accountability systems and measures and determining appropriate interventions.

Second, district capacity also influenced the implementation of CORE activities. Districts varied widely in their ability to manage, interpret, and use the MS data to engage in improvement activities. Even for those districts whose administrators had a great deal of facility with academic and traditional non-academic data, the use of SEL and CC measures was very new. Similarly, these data do not, at this point, evoke a clear set of relevant interventions. For example, poor SBAC math performance among seventh graders might lead administrators to adopt a new mathematics curriculum. In contrast, there appears to be little widespread awareness of ready solutions for low-growth mindset survey results among seventh graders.[xx] This difficulty might, in fact, be a strength, as this disruption may promote the use of CORE's key intervention strategies: cycles of improvement, mutual learning, and peer collaboration. That is, absent a ready-made solution for improving growth mindset, educators may be more likely to engage in inquiry, examine the root causes of low-growth mindset, explore various possible interventions, and evaluate the effectiveness of these solutions. In practice, however, capacity limitations impeded improvement work around SEL and CC at this early stage of implementation. Capacity constraints also affected school intervention work. In addition to the ubiquitous lack of time, administrators also spoke about inconsistency in the skills, motivation, and availability of facilitators.

Third, a perceived misalignment between the MS and existing accountability systems complicated the implementation of both the MS and school interventions. In addition to difficulties aligning CORE's accountability system with district policies including sometimes overlapping, but different, metrics, districts also struggled to maintain a sense of alignment in communicating the accountability system to school-level administrators, teachers, and parents. Further, administrators also raised questions about alignment with the district LCAP. One central office administrator shared:

> I'm really hoping that the state's new accountability program is part of that LCAP process and not separate, which I still think is going to be a little tricky because LCAP is the state and you still have federal requirements around those federal dollars. I think there is a way to do it, but sometimes they don't think that way, I'm hoping that they do… I know that the state has been thinking about it, but now that we have these LCAP plans, that's creating a bit of dissonance between the [federal] LEA plan and an LCAP plan. How do we use the LCAP to meet some of those federal requirements?

In essence, districts were responding to three accountability systems concurrently: their local system, the LCAP, and CORE. The perceived lack of alignment contributed to feelings that at times administrators were complying with CORE activities without necessarily engaging in the intended deeper learning and continuous improvement.

Finally, ESSA and other policies introduced a level of uncertainty that colored the implementation of CORE waiver activities. Ambiguity about the future state accountability system and state of the waiver came up frequently as a barrier to implementation at a couple of the districts. One district acknowledged that without the NCLB waiver and with a new, ESSA-oriented accountability system coming from the state, there may be less of an incentive to participate in CORE. As one central office staff reported:

> Where the intention is there to keep up with CORE and be a part of CORE, the intention is very much there, I would say with anyone you speak with but then it becomes an issue just to be honest with you of there's so much on everyone's plate. … unless it's mandated by someone high up, how much involvement will we have?

The ambiguity around the purpose of CORE, apart from the waiver, further complicated implementation. Nevertheless, some central office staff also viewed the unknown accountability system as an opportunity for CORE to evolve to meet the needs of its members and serve as a model for the state's accountability system. ESSA reauthorization presented the opportunity for district administrators to engage in strategic planning with CORE, to reimagine what the CORE partnership might look like and how to structure local interventions. All this is possible because of the interdistrict collaborative built by CORE which in some ways, appears to have increased a sense of agency among district administrators.

Implications

The experiences of the CORE districts have the potential to inform the design and implementation of accountability systems currently under development around the country. As such, our findings are relevant to state and district leaders creating and responding to new accountability systems.

New measurement systems. If the experience of the CORE districts is generalizable, educators will be receptive to the inclusion of a more diverse set of indicators as required under ESSA. Educators in our study indicated that tending to the "whole child" has always been a central part of their work, and that the extended measurement served the purpose of better aligning the accountability system with their existing views and approaches. The inclusion of multiple measures in this way serves to identify the strengths and weaknesses of schools on many dimensions, which the CORE districts appreciated.

In order for new data systems to truly be useful for improvement, several conditions appear to be essential:[xxi]

- **Buy-in**: If educators and leaders are going to take the time to review and reflect on data, they need to perceive them to be valid. They also want to feel that the indicators reflect what they value. District efforts to adapt measurement systems to reflect local values and needs may be important for building local buy-in, but these adaptations also come with challenges. If multiple data systems are perceived to be misaligned, it can undermine this support. In fact, achieving wide-scale buy-in may be difficult. Even among CORE districts, where leaders created the system and selected the indicators, there was considerable variation in buy-in within and across districts. Involving educators in the development of the systems is important, but likely not sufficient. Education efforts that demonstrate the validity and value of the measures may also be needed.

- **Capacity**: To use the new data to drive practice and improvement, educators and leaders also need to understand it and know how to respond. As scholars and practitioners have long observed, "data [alone] don't drive" (Dowd, 2005) and do not immediately lead to action without the capacity to interpret and know how to act (Marsh, 2012; Marsh, Bertrand, & Huguet, 2015). While there is generally a developed capacity to interpret and respond to academic data (e.g., state, benchmark, interim, and diagnostic assessment results), this capacity needs to be built around the newer academic and non-academic measures. Educators in the CORE districts in particular conveyed the need for more support around SEL data. While there are resources available on research-based strategies to

promote SEL competencies (e.g., growth mindset instructional strategies), there may not be widespread awareness and access to them. Much of the guidance on student achievement data use suggests that it is valuable to link data to clear next steps for instruction (e.g., through sample lesson plans that address specific weaknesses identified in the data) (e.g., Hamilton et al., 2009). A similar linking process may be needed to promote informed responses to non-academic data. There are also added challenges of acting within the context of a multiple-metric system. With the shift away from solely academic measures to multiple academic and non-academic measures, educators and leaders need to learn how to interpret and respond to data that may show positive results on some types of indicators and negative results on others.[xxii] Knowing that 60% of students are proficient in mathematics is very different from knowing students are high on five indicators and low on five other indicators of math performance. Schools once rated "good" or "bad" may now be rated "good" in some areas and not on others. For example, an academically high-achieving school might find that its school culture or students' social-emotional outcomes are relatively low. Rather than just considering SEL and CC as the means to an end, the holistic dashboard approach charges districts and schools with addressing many kinds of academic and non-academic outcomes.

While some may be inclined to create readily available solutions and a menu of "what works," capacity to respond may require more. Educators need to understand the data and what might be the underlying reasons for lower-than-expected results. They may also need support accessing possible ways to address these underlying causes. Such efforts might require networking with experts, support from peers and coaches, and access to research. The complexity of new data systems also requires greater communication strategies and ways of helping all stakeholders understand what it means for a school to be rated high on some measures and not on others, and then what to do about it.

- **Culture**: To respond in meaningful ways to data, educators need the support of colleagues and a culture that supports reflection over compliance. While we cannot expect a shared commitment to new data systems to appear overnight, there may be opportunities to build shared understandings about the new measures over time via preparation programs, supervisory supports, and in-service programs.

In addition, our interviews with CORE district educators indicate the need to maintain a careful eye on the possibility of distortive responses to these new accountability systems. Under NCLB, research demonstrated ways in which educators narrowed instruction, focused on "bubble kids," and emphasized test-taking strategies to assist schools in meeting adequate yearly progress (AYP) (Booher-Jennings, 2005; Jennings & Bearak, 2014; Hamilton et al., 2007; Marsh, Pane, & Hamilton, 2006). The new accountability systems introduce a host of new measures with the potential for similar "gaming" strategies, as CORE district educators readily reported. Even without the "high stakes" of sanctions possible under NCLB, pressures to "look good" for the public and to attract and retain students in contexts of declining enrollment and school choice could create incentives for educators to improve the numbers but not the practices in schools. To ensure productive responses to these new measures (and the ultimate validity of the results), administrators should carefully monitor schools and ensure consistent messages about the purposes and proper responses. Further investment in capacity-building, particularly around changing the culture of data use and accountability in schools, may facilitate this shift.

School interventions. Under ESSA, districts will now be responsible for school improvement. Given the backlash against NCLB's approach to improvement, it is highly likely that many districts will pursue capacity-building activities in the same spirit as the CORE school intervention work. Once again, the

experiences of educators within the CORE districts suggest several important lessons. If districts elect to implement collaborative approaches similar to the peer-to-peer interventions used in CORE, they should consider the importance of several key factors:

- **Fit and reciprocity**: To maximize the potential for authentic, mutual learning, schools need to be matched in ways that guarantee all parties have something to learn from one another. Pairing schools based on demographics and performance alone may not always lead to meaningful exchanges, as some administrators in CORE district schools shared with us (despite their support for the idea of using cross-school collaboration to address performance problems). Instead, matching should also consider the local context, as well as geographic proximity, and ensure there is adequate time for meetings to occur. The content of improvement efforts is also an important consideration: A school that wishes to develop new practices to address low English learner performance may be best paired with a similar school that has had success in improving their EL program. Leaders may also consider allowing schools greater agency in selecting their collaborative partners (which may also promote buy-in), while brokering connections with similar schools that may best assist with specific improvement issues.

  Reciprocity also appears to depend on clear expectations and roles. If in fact there is a commitment to pair "higher performing" to "lower performing" schools, administrators should consider the incentives for higher performing schools to engage in these efforts and the social-political dynamics that may complicate interactions. Is mutual learning a reasonable expectation in such arrangements? How might meaningful collaboration among schools be facilitated? Moreover, considering schools to be expert in specific content (e.g., ELL programming), may open up opportunities for collaboration for all kinds of schools based on specific needs (e.g., the EL expert school may need to find a partner to assist with growth mindset).

- **High-quality facilitation**: Successful collaboration also rests on the availability of well-trained, committed facilitators. As witnessed in the CORE districts, schools taking a compliance orientation toward the intervention work often lacked a facilitator who had: a shared commitment to collaborative interventions; a solid understanding of the schools and their plans; and the capability to meet regularly, push educators' thinking, and monitor progress. Districts should consider allocating resources to support staff time and training for these positions, as well as investing in tools that can promote facilitators' work with schools (e.g., protocols for leading meetings around processes such as Plan-Do-Study-Act, materials to guide coaching).

Not surprisingly, capacity-building activities of any type rely on sufficient buy-in and capacity at the school sites. Without basic building blocks—such as stable teachers and leaders, sufficient funding, safety—improvement efforts may be misdirected or short-lived (e.g., investing in in-depth staff training regarding growth mindset in a setting with extremely high staff turnover). The experiences of the CORE districts indicate that the people inside schools matter greatly, as does the environment in which they work. This realization leads to two important implications:

First, regardless of which approach to capacity-building is taken, states and districts should consider other policy levers to ensure that all schools, particularly those performing in the bottom 5%, are staffed effectively. Teachers, staff, and leaders should be committed to take on improvement efforts and promote a culture that supports educators as they reflect on data, try out new strategies, and monitor progress. While

not explicitly included in ESSA, leaders should not forget the importance of personnel policies and practices that promote better recruitment, preparation, development, and retention of educators.

Second, given the considerable challenges often facing low-performing schools—such as a high student mobility and staff turnover, safety concerns, low morale, the lack of trust or professional culture—one must ask if a peer-to-peer intervention model goes far enough to address these difficult conditions and to promote deep learning and improvement. Providing educators the opportunity to discuss and access new knowledge may build reflective capacity, but does it build sufficient capacity to do the actual work? Is it possible that some schools' needs go beyond collaboration? Perhaps under certain conditions, other models may be needed—models that rely on more in-depth needs assessments, training, technical assistance, structural changes, new leadership, etc. While ESSA clearly seeks to move away from NCLB-era interventions that were perceived by many to be "draconian," we should not rule out the possibility that, in some schools, true improvement will require organizational changes and more intensive capacity-building.

CORE leaders are currently in the process of revising their intervention strategies and considering many of the questions we pose here, as should others seeking to learn from the early experiences of CORE districts.

District-level capacity-building. Under ESSA, districts will be responsible for facilitating data use and improvement in schools. Cross-district collaboration is one approach for building district capacity to engage in these processes. Once again, we can learn a lot from CORE about district capacity-building generally and networked approaches more specifically.

First, it takes a long time to build relationships, but once they are established, districts can communicate directly about problem solving. Building these relationships requires dedicated time and, ideally, shared projects to build trust and shared language. Our research in CORE districts suggests that formal collaboration opportunities were necessary but not sufficient: Administrators highly valued frequent opportunities for informal interactions with individuals serving in similar positions.

Second, moving the work requires having the right people at the table. As CORE district members shared with us, if attendees do not have the authority to make needed changes in their respective districts in response to the ideas generated in meetings, progress will be slow if not stalled. Moreover, frequent turnover and rotating attendees were seen as an obstacle to building momentum (as well as trust among participants). Administrators seeking to build cross-district collaboration should consider ways to promote continuity among those participating in these efforts (e.g., expectations, incentives) and to ensure that individuals with agency attend.

Finally, much like school-level interventions, issues of fit and reciprocity are paramount. Participating districts must feel that they have something to learn from one another. This raises questions about whether districts differing in size and local context are best suited to collaborative work and indicates the need for purposeful matching. This matching should also consider the alignment of policies and goals between districts. That is, do the goals of collaborative work fit with districts' local priorities? Are there common problems that can be identified that truly are central to multiple districts and not viewed as add-on work that draws administrators away from what they care about most? If matching is expanded, policymakers should consider both the context (e.g., district size, demographics, policies, and location) as well as the content (e.g., addressing teacher retention issues, managing data) of district improvement needs to properly facilitate relationship- and network-building across districts.

Once again, CORE leaders and member districts are in the midst of reflecting on a similar set of issues and questions as they plan for their future together. It behooves leaders elsewhere to similarly assess their plans and monitor implementation over time to ensure that investments truly build district capacity to support the improvement goals of new accountability systems in this next phase of policy implementation.

References

Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading data use in schools: Organizational conditions and practices at the school and district levels. *Leadership and Policy in Schools, 9*(3), 292-327. Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231-268.

Blad, E. (2015, December 10). CORE districts to Share Social-Emotional Measures to Inform Accountability Plans. Education Week. Retrieved from http://blogs.edweek.org/edweek/rulesforengagement/2015/12/core_districts__social-emotional_measures_to_inform_accountability_essa.html

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American educational research journal*, *42*(2), 231-268.

Bornstein, D. (2015). Teaching social skills to improve grades and lives. Retrieved from http://opinionator.blogs.nytimes.com/2015/07/24/building-social-skills-to-do-well-in-math/?_r=0

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*.

Carranza, R. (2015) How best to measure success with English learners. Redesigning Accountability. Retrieved from https://edsource.org/2015/holding-school-districts-accountable-for-success-with-english-learners/87121.

Clough, C. (2016, April 11). "New data reveal best and worst of LAUSD schools." *LA School Report*. Retrieved from http://laschoolreport.com/new-data-reveal-best-and-worst-of-lausd-schools/

Clough, C. (2016, April 26). "Stark differences for LAUSD elementary schools in the CORE accountability index." *LA School Report*. Retrieved from http://laschoolreport.com/stark-differences-for-lausd-elementary-schools-in-the-core-accountability-index/

CORE (2013). *ESEA Flexibility Request*. Retrieved from http://coredistricts.org/wp-content/uploads/2013/02/CORE-ESEA-Flexibility-Request.pdf

CORE Districts (2015). A technical guide to the CORE School Quality Improvement Index: SY 2014-15. http://coredistricts.org/wp-content/uploads/2016/02/CORE-Index-Technical-Guide-SY-2014-15-updated-2.1.16.pdf

Cosner, S. (2012). Leading the on-going development of collaborative data practices: Advancing a schema for diagnosis and intervention. *Leadership and Policy in Schools, 11*(1), 26-65.

Datnow, A., Park, V., & Kennedy-Lewis, B. (2012). High school teachers' use of data to inform instruction. *Journal of Education for Students Placed at Risk, 17*(4), 247-265.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and management*, *30*(3), 418-446.

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record, 106*(6), 1145-1176.

Dowd, A. C. (2005). *Data don't drive: Building a practitioner-driven culture of inquiry to assess community college performance*. Boston, MA: University of Massachusetts.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes. Educational Researcher , 44(4), 237–251.

Every Student Succeeds Act, Public Law 114-95, 114th Cong., 1st sess. (December 10, 2015), available at https://www.congress.gov/bill/114th-congress/senate-bill/1177?q={%22search%22%3A[%22every+student+succeeds+act%22]}&resultIndex=1.

Farrell, C. & Marsh, J. (2016). Metrics matter: How properties of data shape teachers' instructional responses. *Educational Administration Quarterly, 52*(3), 423-462.

Fullan, M. (2011). *Choosing the wrong drivers for whole system reform*. East Melbourne: Centre for Strategic Education.

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). Using student achievement data to support instructional decision making (NCEE 2009-4067). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Hamilton, L., Stecher, B., Marsh, J., McCombs, J. S., Robyn, A., Russell, J. L., Barney, H. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: RAND Corporation.

Hough, H. J., & Witte, J. (2016). *Making Students Visible: Comparing Different Student Subgroup Sizes for Accountability*. Retrieved from http://www.edpolicyinca.org/publications/making-students-visible-comparing-different-student-subgroup-sizes-accountability

Hough, H., Penner, E., & Witte, J. (2016). Identity crisis: Multiple measures and the identification of schools under ESSA. Stanford, CA: PACE. http://edpolicyinca.org/sites/default/files/PACE_PolicyMemo_1603.pdf

Jennings, J., & Bearak, J. (2014). "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher, 43*(8), 381-389.

Jennings, J., & Rentner, D. S. (2006). Ten big effects of the No Child Left Behind Act on public schools. *Phi Delta Kappan, 88*(2), 110.

Johnson, M. A. (2012, December 26). "US turns down No Child Left Behind waiver for California." *NBC News*. Retrieved from http://usnews.nbcnews.com/_news/2012/12/26/16171331-us-turns-down-no-child-left-behind-waiver-for-california?lite

Kirst, M. (October 9, 2016). Strong, locally generated data key to school improvement strategies. *Edsource*. https://edsource.org/2016/strong-locally-generated-data-key-to-californias-school-improvement-strategies/570340

Knudson, J., & Garibaldi, M. (2015). *None of Us Are as Good As All of Us: Early Lessons from the CORE Districts*. AIR. Retrieved from http://coredistricts.org/wp-content/uploads/2015/08/CORE-Cross-District-Collaboration-Report-August-2015.pdf

Krachman, S. B., Arnold, R., & LaRocca, R. (2016). Expanding Our Definition of Student Success: A Case Study of the CORE Districts. Boston, MA: Transforming Education. https://static1.squarespace.com/static/55bb6b62e4b00dce923f1666/t/57ea8a3cbe6594387dad0b11/1474988682108/Transforming+Education+Case+Study+FINAL+(1).pdf

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*: Cambridge University Press.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, *114*(11), 1-48.

Marsh, J., Bertrand, M. & Huguet, A. (2015). Using data to alter instructional practice: The mediating role of coaches and professional learning communities. *Teachers College Record, 117*(4), 1-40.

Marsh, J. A., Pane, J. F., & Hamilton, L. (2006). *Making sense of data-driven decision making in education*. Washington, D. C.: RAND Corporation.

McLaughlin, M. W., & Talbert, J. E. (1993). *Contexts that matter for teaching and learning*. Stanford, CA: Center for Research on the Context of Secondary School Teaching, Stanford University.

McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago: University of Chicago Press.

Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011). Teachers' ability to use data to inform instruction: Challenges and supports: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). Qualitative data analysis: A methods sourcebook. Thousand Oaks: Sage.

Moen, R. D., & Norman, C. L. (2010). Circling back. *Quality Progress, 43*(11), 22.

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics, 92*(2), 263-283.

Polikoff, M. S., McEachin, A. J., Wrabel, S. L., & Duque, M. (2014). The waive of the future? School accountability in the waiver era. *Educational Researcher, 43*(1), 45-54.

Seashore Louis, K., & Kruse, S. D. (1995). *Professionalism and community: perspectives on reforming urban schools*. Thousand Oaks, CA: Corwin Press.

Stecher, B. M., Epstein, S., Hamilton, L. S., Marsh, J. A., Robyn, A., McCombs, J. S., ... & Naftel, S. (2008). Pain and gain: Implementing No Child Left Behind in California, Georgia, and Pennsylvania, 2004 to 2006. *The RAND Corporation*.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change, 7*(4), 221-258.

US DOE (2012) *ESEA Flexibility*. Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.

Yin, R. K. (2013). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

Zernike, K. (2016, February 29, 2016). Testing for joy and grit: Schools nationwide push to measure students' emotional skills. New York Times. Retrieved from http://www.nytimes.com/2016/03/01/us/testing-for-joy-and-grit-schools-nationwide-push-to-measure-students-emotional-skills.html?action=click&contentCollection=Opinion&module=RelatedCoverage&region=EndOfArticle&pgtype=article

Appendix A: Additional Details about Our Research Methods

Table A1 provides further details about the individuals we interviewed across the CORE districts.

Table 1: Number of Interviews Conducted, by organization and role

	CORE staff	central office			school administrator			TOTAL
		Superintendent*	Administrator	Facilitator	Priority	Reward	Focus (COP)	
CORE	4							4
SFUSD		1	4	2	1	1	1	*10*
SAUSD		1	7	0	1	1	0	*10*
FUSD		1	4	2	1	0	1	*9*
OUSD		1	5	2	3	0	2	*13*
LAUSD		1	4	0	0	2	0	*7*
LBUSD		1	5	0	0	1	0	*7*
TOTAL	*4*	*6*	*29*	*6*	*6*	*5*	*4*	***60***

*In one district, we interviewed a high level district adminsitrator who served on the CORE board in lieu of the superintendent.

[i] The broader CORE-PACE Research Partnership focuses on producing research that informs continuous improvement in the CORE districts and policy in California and beyond. Other research reports can be found at http://www.edpolicyinca.org/projects/core-pace-research-partnership.

[ii] See Knudson and Garibaldi (2015) for a more detailed account of CORE's early years.

[iii] Improvement science aims to "develop the necessary *know-how* for a reform idea ultimately to spread faster and more effectively" among local school actors (Bryk et al., 2015, p. 8). Drawing on empirical evidence, educators are expected to constantly and iteratively ask three improvement questions: "What is the specific problem I am now trying to solve? What change might I introduce and why? And, how will I know whether the change is actually an improvement?" (p. 9). Improvement science interventions focus attention on these key questions through the use of a set of standard tools (e.g., driver diagrams). In addition, Networked Improvement Communities describe intentionally created social networks focused on using these tools to solve problems (Bryk et al., 2015).

[iv] California's theory of action in this new phase of policy implementation is that local districts will innovate on measurement and accountability and that these lessons will be spread across the state to facilitate continuous improvement. In order for the state to learn from the CORE district's data for use in accountability systems, the CORE districts propose a research pilot so that CORE's measures are actually used in the context of an accountability system. Specifically, this means that CORE's measures would be used in place of state measures for identifying schools for improvement and monitoring progress toward goals (Kirst, 2016).

[v] The measurement of SEL is the most controversial of CORE's measures, as some researchers have expressed concern that such measures are not suitable for use in accountability systems (e.g., Duckworth & Yeager, 2015). However, early evidence from CORE's field test shows that the SEL surveys demonstrate high levels of validity and reliability (West, 2016; Measuring MESH)

[vi] The growth indicator is currently under development; the version included in these analyses is preliminary and is only available for elementary and middle schools at this time.

[vii] CORE's measure of EL proficiency is slightly different than what is specified in ESSA. Rather than using only test score results to determine progress on English proficiency, the CORE districts chose to report reclassification rates, which are a combination of language proficiency scores and academic performance (Carranza, 2015).
[viii] Over 85% of CORE's student items are from the California Healthy Kids Survey or the California School Climate Survey, both of which have been used extensively across California. For further detail on reliability and validity of the California Healthy Kids Survey or the California School Climate Survey, visit http://chks.wested.org/ and http://cscs.wested.org/, respectively.
[ix] For further detail on the survey items measuring social-emotional learning, see http://www.transformingeducation.org/measuringmesh/.
[x] While performance is reported for each racial/ethnic subgroup, the racial/ethnic subgroup with the worst metric performance on each metric is identified for inclusion in the MS calculation. This means that a school could have a different "lowest performing racial ethnic group" across different MS elements.
[xi] Subgroup size is the minimum number of students in a subgroup (e.g., English learners, Special Education) in a school needed for that group's performance data to be reported and counted. For a further discussion of student subgroup sizes and the impact on measurement, see Hough and Witte (2016).
[xii] For a more detailed treatment of the development of SEL and CC measures, integration into the MS, and training for districts, see Krachman, Arnold, and LaRocca (2016).
[xiii] The vast majority of schools that were identified for improvement used a minimally disruptive intervention rather than the more dramatic approaches often emphasized in the media. (Jennings & Rentner, 2006).
[xiv] Communities of Practice encompasses a broad literature concerning contextually embedded social learning (Lave & Wenger, 1991; Wenger, 1999), often taking place within school communities (e.g., McLaughlin & Talbert, 1993, 2001; Seashore Louis & Kruse, 1995; Stoll et al., 2006). In this report, we use the capitalized term *Communities of Practice* to denote the specific intervention utilized as part of CORE's accountability system.
[xv] These ideas were based on the work of Fullan (2011).
[xvi] For example, possible selections for the rubric item on stakeholder engagement ranged from "LEA conducted **regular,** meaningful stakeholder engagement on the School Quality Improvement Index and School Support Interventions to administrators, teachers, other staff, parents, and the local community in multiple easy-to-access fashions" to "LEA **did not attempt** to consult on Principle 2 to relevant stakeholders."
[xvii] We learned less about the Special Education identification rate indicator, in large part because it was not yet formally a part of the MS. For this exact reason, central office staff in several districts acknowledged that they either have not been paying "a whole lot of attention to the SPED piece" or were "not even touching" it.
[xviii] Although MS results were released on the CORE website, not all districts actively publicized them.
[xix] Deming's Plan-Do-Study-Act cycle is a series of steps designed to promote continuous improvement, during which organizational actors iteratively plan (set a goal and theory of action, define metrics of success, and make an action plan), do (implement the plan), study (monitor plan outcomes), and act (learn from the process by adjusting the goal, changing methods, or rethinking theory of action) over and over until the problem is solved (Moen & Norman, 2010).
[xx] While there is a growing collection of research-based strategies to address SEL and CC constructs, district- and school-level administrators may not be readily aware of them.
[xxi] Prior research on the use of data in schools confirms the importance of these elements (see, for example, Anderson, Leithwood, & Strauss, 2010; Cosner, 2012; Datnow, Park & Kennedy-Lewis, 2012; Diamond & Spillane, 2004; Farrell & Marsh, 2016; Means, Chen, DeBarger, & Padilla, 2011).
[xxii] Recent analyses indicate that in fact the majority of schools is high on some measures and low on others, indicating that multiple measures show different dimensions of school performance (Hough, Penner, & Witte, 2016).

http://edpolicyinca.org