

# Learning From Early Adopters in the New Accountability Era: Insights From California's CORE Waiver Districts

Educational Administration Quarterly  
2017, Vol. 53(3) 327–364

© The Author(s) 2017

Reprints and permissions:  
sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013161X16688064

journals.sagepub.com/home/eaq



Julie A. Marsh<sup>1</sup>, Susan Bush-Mecenas<sup>1</sup>, and Heather Hough<sup>2</sup>

## Abstract

**Purpose:** The Every Student Succeeds Act (ESSA) represents a notable shift in K-12 accountability, requiring a more comprehensive approach to assessing school performance and a less prescriptive approach to intervening in low-performing schools. In this article, we seek to leverage the experiences of California's CORE (California Office to Reform Education) waiver districts to better understand what it means to implement an ESSA-like system. Specifically, we examine educators' attitudes about CORE's accountability system, how it was implemented, and its intermediate outcomes. **Research Methods:** We use a multiple, embedded case study design, examining the implementation of CORE's accountability system across all six CORE Districts. We draw on interviews with CORE staff ( $n = 4$ ), district leaders ( $n = 6$ ) and administrators ( $n = 29$ ), and school principals ( $n = 15$ ); observations of CORE meetings (42 hours); and documentation. **Findings:** We find strong buy-in for CORE's accountability system and considerable adaptation of its key elements. District administrators also

---

<sup>1</sup>University of Southern California, Los Angeles, CA, USA

<sup>2</sup>Stanford University, Stanford, CA, USA

## Corresponding Author:

Julie A. Marsh, University of Southern California, 3470 Trousdale Parkway, WPH, 904C, Los Angeles, CA 90089-4039, USA.

Email: julie.marsh@usc.edu

reported challenges with achieving reciprocity in collaborative activities, and limited capacity, validity concerns, and policy misalignment constrained implementation. Reported effects on practice and learning indicate CORE efforts were a work in progress. **Implications for Research and Practice:** This research suggests lessons about what it means to be “data-driven” in a multiple-measures accountability era and raises questions about how to facilitate school improvement. While efforts to motivate change via test-based measures, sanctions, and prescribed interventions in prior accountability eras may not have yielded all the expected positive results, our study indicates that a shift to multiple measures, greater flexibility, and locally determined capacity-building efforts brings its own set of challenges.

### **Keywords**

accountability, data use, capacity building, organizational learning, policy implementation

The Every Student Succeeds Act (ESSA) represents a notable shift in K-12 accountability. Unlike the No Child Left Behind (NCLB) act, ESSA requires a more comprehensive approach to assessing school performance that includes both academic and nonacademic measures and a less prescriptive approach to intervening in low-performing schools. While much is known about the implementation and effects of NCLB (e.g., Dee & Jacob, 2011; Neal & Schanzenbach, 2010; Stecher et al., 2008), little is known about the new accountability systems likely to emerge under ESSA (2015). While some states 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 on state-driven interventions for struggling schools and few incorporated expansive measurement systems (e.g., most relied on test results in math and English language arts and few used non-state measures other than graduation rates; McNeil, 2014; Polikoff, McEachin, Wrabel, & Duque, 2014).

However, one state-like consortium of districts, the California Office to Reform Education (CORE), has designed and implemented a new accountability system well-aligned to ESSA's state requirements for holistic measurement systems, customized local support for school improvement, and public engagement with data (U.S. Department of Education [US DOE], 2016). The 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 an ESSA-like accountability system. Freed by

the U.S. Department of Education from some of their obligations under NCLB in 2013, these six districts developed and are implementing CORE's accountability system (the School Quality Improvement System) that provides comprehensive data on performance and emphasizes the importance of Fullan and Quinn's (2015) "right drivers" for school improvement. Key features of this system are (a) a measurement system (hereafter MS), formally referred to as the School Quality Improvement Index, that focuses on academic outcomes alongside nonacademic measures of student success, (b) peer-to-peer school improvement interventions, and (c) district-level capacity building.

In total, the CORE Districts represent more than one million students (20% of California students), including large percentages of students of color, students from low-income backgrounds, and English learners. Collectively, these enrollment figures exceed the total student population in more than two-thirds of states and reflect the diversity of students served nationally. As such, the implementation of the CORE accountability system across a diverse and geographically dispersed set of districts faces a set of potential challenges one might observe in states generally. In this article, 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 locally determined, collaborative improvement efforts. Specifically, we ask the following: *How did districts and schools implement and respond to the new accountability system, including the measurement system, school improvement efforts, and cross-district collaboration?* We examine educators' attitudes about the system, how it was implemented, what facilitated or constrained this process, and the intermediate outcomes of these efforts.

Findings from this study can inform the development and implementation of future accountability policy in states and districts nationwide. Notably, our research suggests lessons about what it means to be "data-driven" in a new multiple-measures era of accountability. The study also raises important questions about how to facilitate school improvement. While efforts to motivate change via sanctions and prescribed interventions in prior accountability eras may not have yielded all of the expected positive results, our research in the CORE Districts indicates that a shift to greater flexibility and locally determined capacity-building efforts brings its own set of challenges.

In the remainder of this article, we first provide background on the CORE Districts and their accountability model, followed by a review of the theoretical and empirical literature guiding our research. Next, we describe our research methods and present our findings related to the implementation of the measurement system and school and district capacity-building efforts. We conclude with a set of cross-cutting tensions and implications for policy and practice.

## Background on the CORE Districts

Building on a partnership started in 2010, the CORE Districts pursued an NCLB flexibility request in 2013 (CORE, 2013).<sup>1</sup> CORE's pursuit of this waiver arose naturally from many years of prior formal and informal collaboration around Common Core State Standards implementation.<sup>2</sup> Approved in August 2013, the waiver laid out CORE's theory of action: that 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, which remain true today:

1. *The importance of local control.* The districts committed to learning from each other and holding each other accountable for outcomes, but retaining 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). In this way, the CORE waiver intended to support, rather than replace, individual districts' internal accountability and measurement systems.
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, but instead flexibility to do what is best for their students and the support of one another in making big improvements. As stated in the flexibility request, "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" (CORE, 2013, p. 24).
3. *Capacity building through peer-to-peer collaboration with a focus on data.* Leaders believed that giving districts and schools flexibility to improve hinges on staff capacity to identify problems and know how to fix them. As such, "it is CORE's hope to let data drive all actions and rely on peer-to-peer collaboration and support as much as possible" (CORE, 2013, p. 20). Leaders asserted that performance data be used as "a flashlight, not a hammer," to help schools improve not punish.

In this study, we focused primarily on how CORE measured school quality and supported struggling schools during the 2015-2016 school year, or the third year of implementation.<sup>3</sup> It is important to note that not all aspects of the waiver system were fully operational yet at the time of our study, and that the system itself was in flux due to ESSA reauthorization in December 2015.

However, the waiver was in full effect while we were conducting our field work.

### *Measuring School Quality*

ESSA requires a more comprehensive approach to measurement than was required under NCLB, and CORE's MS exceeds these requirements, as illustrated in Table 1. ESSA requires states to include multiple measures of student academic achievement, 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, postsecondary readiness, or school climate and safety.

In the full MS, the academic domain accounts for 60% of the final score, and measures in the ESSA SQSS domain account for 40%. Each indicator in the MS is weighted, with those weights being aggregated into a single number (which was a federal requirement under the waiver for the identification of schools for intervention). For most metrics, points are divided between the "all students" group and the four subgroup categories (lowest-performing racial ethnic group,<sup>4</sup> English Learners, students with disabilities, and disadvantaged students) measured using a subgroup "n size" of 20.<sup>5</sup> For each metric and each subgroup, schools are given an index level score which 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, where 10% of schools are always identified as Level 1. By setting the levels and keeping them for multiple years, all schools can show improvement on a metric.

The measurement reports, released for each school, are designed to display all the MS measures, including 3-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." In the year of our study (2015-2016), social-emotional skills, culture climate, and academic growth did not yet "count"—they were measured, but not yet reported on the MS or included in the final score, as this was a field-test year for these measures. 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.<sup>6</sup> Also, at the time of our research,

**Table 1.** Elements of CORE's Measurement System, Organized by ESSA Domains.

ESSA Academic Domain	
Academic performance	Percentage of students testing proficient for ELA and mathematics, based on Smarter Balanced Assessment Consortium (SBAC) test scores.
Academic growth	Growth percentile (rank from 0 to 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 1 year to the next relative to demographically similar students who started the school year with similar prior achievement.
Graduation	Percentage of students who graduate in a 4-, 5-, or 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 8th grade students who meet the following criteria: (a) 8th grade GPA of 2.5 or higher, (b) attendance of 96% or higher, (c) no grades of D or F in ELA or math in the final course grade, and (d) were not suspended in 8th grade.
EL redesignation <sup>a</sup>	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 5 years, were not reclassified at that school. <sup>b</sup>
ESSA "School Quality or Student Success" 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 5-12), school staff, 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. <sup>c</sup>
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. <sup>d</sup>

<sup>a</sup>Although considered an academic indicator under ESSA, in CORE's measurement system, EL re-designation falls in the nonacademic, social-emotional/culture climate domain. <sup>b</sup>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). <sup>c</sup>More than 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 details on reliability and validity of the California Healthy Kids Survey or the California School Climate Survey, visit <http://cscs.wested.org/> and <http://chks.wested.org/>, respectively. <sup>d</sup>For further detail on the survey items measuring social-emotional learning, see <http://www.transformingeducation.org/measuringmesh/>.

schools were not yet reclassified based on these measures (i.e., no new schools were targeted for “intervention”).

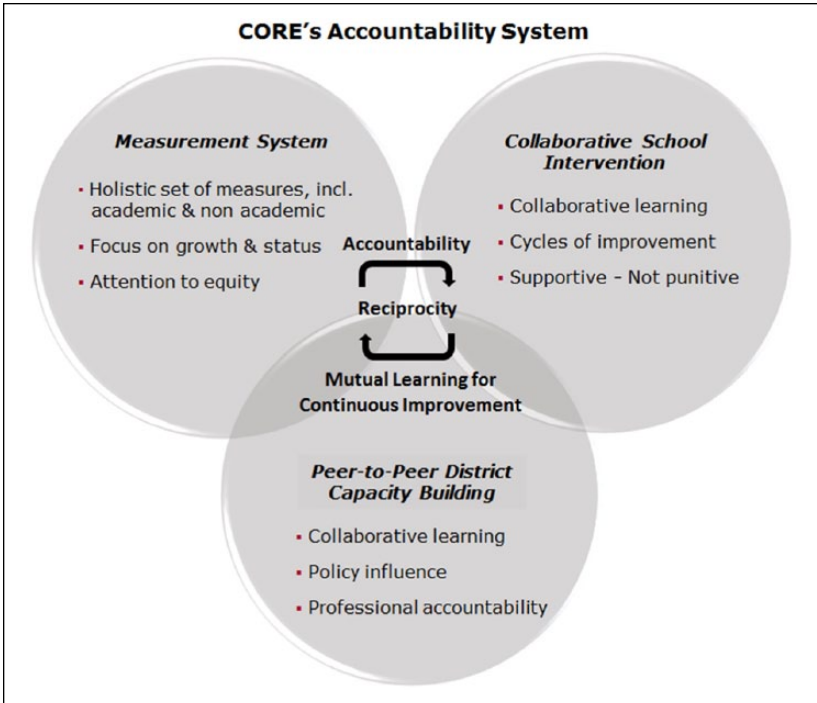
### *Supporting Struggling Schools*

Under NCLB (and just as ESSA requires), schools identified in the bottom 5% were required to undergo interventions for school turnaround. As part of their waiver from the NCLB requirements, the CORE Districts designed a system of intervention they believed would better meet the needs of their low-performing schools than the prescriptive NCLB interventions, which included reconstitution, restructuring, closure, and charter takeover. In line with their belief that schools can improve most quickly when allowed autonomy and encouraged to collaborate, in the CORE interventions, schools were provided a framework for engaging in inquiry and knowledge sharing, but their ultimate decision making and implementation were nonprescriptive in nature.

Specifically, 178 schools were identified across the CORE Districts to engage in two tiers of improvement activities. Schools falling in the lowest 5% of all schools in the CORE Districts were paired with high-performing schools or those demonstrating high growth with similar student populations. Within these Pairings, it was initially assumed that the high-performing school leaders would serve as coaches to guide the improvement process in struggling schools. Other schools with low-performing subgroups or students were grouped into Communities of Practice (COPs).<sup>7</sup> Paired schools were encouraged to identify 2 to 3 problems of practice, develop an improvement plan based on ESEA’s seven turnaround principles (listed in US DOE, 2012), create a structure for collaborative interaction, meet quarterly, and show evidence of “learning and progress.” COP schools (2-4 Focus schools in each group) were encouraged to identify problems of practice, come together around shared problems, meet quarterly, and run quarterly PDSA (Plan-Do-Study-Act) cycles. After an 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*

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 as a means to build capacity to meaningfully engage



**Figure 1.** Conceptual framework.

with the MS and to improve schools. As such, CORE staff organized formal opportunities for this to occur, including board meetings, quarterly meetings for “role-alike” staff 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.<sup>8</sup> After the self-evaluation, districts would review one another’s work and make suggestions about how to improve implementation.

## **Grounding the Inquiry: Literature on Accountability and Learning**

As illustrated in Figure 1, the design of CORE’s accountability system draws on two key mechanisms: accountability and organizational learning (OL),



both predicated on expectations of reciprocity. To guide our analysis of the implementation of this system we drew on theoretical and empirical literature in these two areas. Rather than formal theories prescribing our analysis, we used concepts from accountability and OL as sensitizing frameworks, giving “a general sense of reference and guidance in approaching empirical instances” and “suggesting directions along which to look” (Blumer, 1954, p.7; see also Charmaz, 2003).

### *Accountability*

We rely on several conceptions of accountability to inform our analysis of implementation of CORE’s accountability system. First, building on principal-agent theory (Holmstrom & Milgrom, 1991)<sup>9</sup> and derivative work (Hentschke & Wohlstetter, 2004; Loeb & McEwan, 2006), we frame accountability as a contractual relationship between an agent, who provides a service (in this case, educators who lead and teach students), and a provider, who sets the objectives for and often has the authority to reward agents (in this case, policy makers, parents, and other stakeholders). Given the decentralization of authority in education, the accountability arrangement helps ensure that educators provide the type of education desired by policy makers and other stakeholders. In this relationship, agents are held responsible for providing a particular service and/or reaching a specific goal. To incentivize and ensure alignment between principals’ objectives and agents actions, accountability relationships include an assessment of agents’ performance (measuring if they achieve specified goals) and consequences for their performance (ranging from more material rewards and sanctions, to psychosocial consequences such as the stigma of low performance). This framing thus defines the who/to whom, what, and how of accountability.

Scholars have defined various types of accountability systems, ranging from bureaucratic, professional, community, market-based moral, legal, and political (Burke, 2005; Darling-Hammond & Ascher 1991; Firestone & Shippy, 2005; Goldberg & Morrison, 2003; O’Day, 2002). Several of these are particularly relevant to the design of CORE’s accountability system, including (a) external, bureaucratic accountability (relying on externally determined assessment and monitoring of progress toward measures of performance or compliance with rules and regulations), with elements of (b) internal, professional accountability (peers holding one another accountable) and (c) political accountability (democratically governed organizations being responsive to the public).<sup>10</sup> The central dimensions of this accountability system are listed in Table 2.

**Table 2.** Defining CORE Accountability.

	Bureaucratic			Political	Professional
To whom CORE Districts are accountable	The federal government and CORE's Board of Directors	Parents and the public at large	One another, as enforced by peer review		
What they are accountable for	For the academic and nonacademic outcomes of students as defined by the MS; For fulfilling the terms of the federal NCLB waiver	For the academic and nonacademic outcomes of students as defined by the MS	For undertaking actions to address the key goals of the waiver		
How are they held accountable—assessment	MS defines the measures of performance for CORE Districts and schools, based on surveys, tests, administrative data	MS results	The peer review system and rubrics measure activity-related performance		
How are they held accountable—consequences	High relative performance on the MS can lead to designations as higher-performing Paired schools (bringing financial rewards) and low relative performance can trigger designations as Paired and COP schools (requiring participation in interventions and supplying funding to facilitate improvement).	Public reporting of performance on the MS could also lead to both recognition and shame for schools and potential loss of gains in support and enrollment	The annual peer review process also has the potential to establish recognition and shame in the eyes of peer district educators		
	Should CORE Districts fail to uphold the terms of the waiver, the Board could recommend and/or the federal government could revoke the waiver and the associated Title I funds				

While CORE's accountability system blends multiple models, it nonetheless includes major elements of a bureaucratic accountability model rooted in U.S. education policy for decades and embraced most recently under NCLB (Mehta, 2013). As such, it is instructive to consider empirical evidence on how schools and educators responded to NCLB and antecedent reforms. Such accountability regimes have experienced serious challenges over the years, such as resistance to mandates, inadequate capacity among implementers, and failure to embrace the complexity of teaching (Mehta, 2013). Research indicates that administrators, teachers, and the public held mixed and complex attitudes about NCLB. While many praised NCLB for drawing attention to the achievement of student subgroups and promoting school responsibility and alignment to standards, others voiced concerns about the validity of state test data, the fairness of the system, and negative effects on teacher morale (Center on Education Policy [CEP], 2006; Murnane & Papay, 2010; Stecher et al., 2008; Stecher, Vernez, & Steinberg, 2010).

As for behavior, some literature found positive changes in administrator and teacher practice resulting from school-based accountability. Administrators report NCLB raised teachers' learning expectations, focused their attention on low-performing groups, and led to greater differentiation of instruction and increased use of achievement data for decision making (CEP, 2006; Hamilton et al., 2007). Studies of NCLB and similar state systems (e.g., Florida) also reveal reported increases in educators' emphasis on low-performing students, searches for more effective practices, and reorganization of learning environments and schedules (Murnane & Papay, 2010; Rouse, Hannaway, Goldhaber, & Figlio, 2013; Stecher et al., 2008). Jennings (2012) classified such responses as "productive," defined as "practices that improve student learning and do not invalidate the inferences about student- and school-level performance that policy makers, educators, and parents hope to make" (p. 4).

Yet other research uncovered responses to NCLB and similar accountability reforms that were "distortive," enhancing test scores and a school's chances of reaching proficiency targets rather than genuine improvement and learning (Booher-Jennings, 2005; Hamilton et al., 2007; Jennings, 2012; Mintrop, 2012; O'Day, 2002).<sup>11</sup> These behaviors include increasing time on tested topics (Dee, Jacob, & Schwartz, 2013; Jennings & Rentner, 2006; West, 2007) and "test prep" (Reback, Rockoff, & Schwartz, 2011), teaching to the test (Smith & Rottenberg, 1991), moving lower performing students around or out so their scores "don't count" (Figlio, 2006; Price, 2010), focusing on students scoring close to proficiency cut-offs (Hamilton et al., 2007; Jennings & Rentner, 2006) and cheating (Koretz et al., 1996). Some scholars

also observed that implementation of NCLB generally focused on compliance rather than substantive improvements (Manna, 2010).

Finally, studies indicate that responses to NCLB varied widely across states (who had considerable flexibility to interpret the policy; Davidson, Reback, Rockoff, & Schwartz, 2015; Murnane & Papay, 2010), as well as within districts and schools (Hamilton, Stecher, Russell, Marsh, & Miles, 2008). State flexibility at times led to misalignments between NCLB metrics and other performance measures (e.g., NAEP, state assessments), resulting in conflicting messages to schools and educators (Kim & Sunderman, 2005; Linn, 2005)

Collectively this literature suggests that educators respond to accountability, particularly externally determined systems, in varied ways and that well-intentioned systems may result in unintended consequences. To what extent will changes to the accountability system under ESSA—multiple measures of performance, less prescriptive interventions, a focus on capacity building—avoid some of the pitfalls observed under NCLB? Will the addition of elements of professional accountability mitigate these challenges? Early implementation findings from the CORE Districts can shed light on how these efforts may play out.

### *Organizational Learning*

We also draw on concepts from OL theory to examine the implementation of CORE's accountability efforts, particularly the elements calling on school and district collaboration. OL describes how organizations create, retain, transfer, and use knowledge, often to adapt to changing environmental conditions (Huber, 1991; Levitt & March, 1988). In particular, the processes through which organizations learn are thought to promote continuous improvement (Dixon, 1994), as organizations gather and respond to data, evidence, and experience to guide future actions. Adaptation and learning may be accomplished through straightforward error correction (i.e., identifying problems using data and implementing ready solutions), inquiry (i.e., responding to data by questioning underlying norms, values, or goals and seeking out promising new practices), or organizational metacognitive processes (i.e., changing organizational practices in an effort to learn how to learn; Argyris & Schon, 1996).

A major premise of district and school improvement efforts under CORE (and as required under ESSA) is that these organizations will use holistic measurement data to inform locally appropriate solutions. In this way, OL counterbalances and supports accountability: CORE's accountability system with its multiple measures and customized interventions relies on the ability

of districts and schools to learn by responding to data and developing appropriate improvement efforts. As such, OL concepts are particularly apt for understanding how CORE's school pairings and COPs, and district peer-to-peer collaboration, have played out over time.

OL theory has been increasingly used to help frame and understand the learning necessitated by education policy changes (e.g., Cohen & Sproull, 1996; Honig, 2008) and how schools and districts respond to state and federal accountability systems (Knapp, 2008). One strand of the empirical research finds a relationship between school factors, OL, and positive intermediate school outcomes. For example, Schechter (2008) found significant correlations between OL mechanisms (including analyzing, storing, and seeking information) and teachers' collective efficacy, commitment to their schools, and experience of school stability. Similarly, Marks and Louis (1999) found a positive, significant correlation between OL processes and teacher-reported empowerment. Another strand of empirical research links OL with school-level change process. For example, Spillane and colleagues (Sherer & Spillane, 2011; Spillane, Parise, & Sherer, 2011) analyze how schools adapt to changing accountability environments and found that schools use routines, repeated group behaviors that function to allow coordination, reduce conflict, and store organizational knowledge, to couple policy with school structure and practice (Spillane et al., 2011) and to influence norms, culture, and school change (Sherer & Spillane, 2011). In this article, we use OL concepts to better understand how CORE's efforts to promote inquiry, collaboration, and learning played out.

## Methodology

We used a multiple case study design (Yin, 2013) to gather data on educators' experiences implementing CORE's accountability and intervention system at school and district levels in the 2015-2016 school year. In each of the six CORE Districts, the research team conducted semistructured interviews with central office administrators responsible for CORE-related work ( $n = 41$ ), including superintendents, cabinet-level administrators, and district staff responsible for data, accountability, school support, curriculum, and human resources (see Table 3).<sup>12</sup>

In each district, we also conducted interviews with a sample of school principals and facilitators engaged in the CORE-related school intervention work. Specifically, in each district we targeted one principal each from a COP, low-performing Paired school, and higher-performing Paired school. During central office interviews, we asked administrators to identify a selection of challenged, typical, and exemplary schools in each category. We

**Table 3.** Number of Interviews Conducted, by Organization and Role.

CORE Staff	Central Office			School Administrator			Total
	Superintendent <sup>a</sup>	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

<sup>a</sup>In one district, we interviewed a high-level district administrator who served on the CORE board in lieu of the superintendent.

selected from among the typical schools and recruited principals for interviews. In the end, we interviewed 15 principals: 6 at the elementary level, 9 at the secondary level (8 middle, 1 high), 4 from COP schools, 6 from low-performing Paired schools, and 5 from higher-performing Paired schools (Table 3). The majority of interviews were conducted in person, with a small subset conducted over the telephone (primarily principal interviews). We also interviewed four leaders from within the central CORE office. We used semi-structured protocols in all interviews, which were audio recorded and transcribed. Protocols addressed individual/district context, attitudes about CORE, awareness and use of the MS, school intervention implementation, and engagement with district collaborative activities.

We supplemented these interviews with observations of CORE meetings and trainings (42 hours). Finally, researchers gathered and analyzed documents pertinent to the overall CORE waiver and related activities (e.g., the NCLB waiver, meeting minutes, PowerPoint presentations) as well as documents produced in the 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 CORE measurement, 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 at 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, Huberman, & Saldaña, 2013).

To further understand patterns across districts, we used 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 paying attention to alternative explanations also helped to ensure the robustness of findings (Yin, 2013). Furthermore, we triangulated findings, wherever possible, among multiple respondents and data sources to strengthen the validity of our findings.

### *Limitations*

When reading our results two 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 further, 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. Furthermore, 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 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.

### **Findings**

We organize our findings around three aspects of implementation: educator attitudes and beliefs, implementation, and intermediate outcomes. In each section, we address the implementation of the measurement system, school-level improvement efforts, and district collaboration.

#### *Attitudes and Beliefs About the Accountability System*

District and school administrators reported overall strong support for CORE's accountability system. They endorsed the social-emotional skills/learning

(SEL) and academic growth measures included in the MS and supported the new focus on support over sanctions, communicated through the peer-to-peer interventions. A minority, however, raised concerns about the new measures and reporting system.

*District and school administrators greatly appreciated a more holistic approach to measurement and the focus on growth over status.* Most administrators valued the MS and the use of a comprehensive set of academic and nonacademic 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.” While many interviewees did not perceive the MS to be new in its entirety, as many were using some to many similar measures prior to CORE, they generally acknowledged the value of having all these measures accessible in one place and formally including them as expectations in the accountability system. One district administrator explained:

When you can get all of those measures in one place and they’re measures that make sense to people that 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.

Administrators also repeatedly praised the MS for including measures of growth in student achievement. 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 where they’re at.

*New measures and reporting mechanisms raised new questions.* Although most interviewees endorsed the inclusion of academic and nonacademic measures, a minority worried that nonacademic measures could “distract” educators from supporting academic outcomes. One administrator feared educators might focus on “easier” issues to tackle, like attendance, rather than “the real work” of academics. Several administrators also voiced concerns about the validity of these measures, such as the SEL and CC surveys taken by the parents, students, and staff. Finally, a minority of interviewees believed that the MS did not “go far enough,” failing to include indicators they believed to be important measures of success and equity, such as college readiness.

Similarly, there were conflicting opinions about how school performance was reported, specifically CORE’s index score system, which ranked schools



across all CORE Districts on individual measures using a 1 to 10 level ranking system. Some administrators believed the rankings allowed them to identify schools whose performance was relatively weak or to seek out the advice of schools performing well on particular indicators. Yet others criticized the ranking system. One principal argued that these rankings wrongly promoted competition over collaboration. Leaders in another district intentionally deemphasized the summative index ranking (a single number aggregating across all measures), noting its conflict with an accountability model intended to provide a multidimensional picture of schools:

The whole point of not making an index 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 LAUSD leaders had not publicly promoted the MS results and rankings, the online local media outlet *LA School Report* published a series of articles in April 2016, highlighting the lowest and highest performers based on the single summative index score 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, 2016a, 2016b). This example illustrates how multiple types of accountability embedded in the CORE accountability system interact and sometimes conflict: while CORE leaders may have wanted to hold one another accountable for a broad set of performance metrics, the public and political actors (e.g., the media) held a different set of expectations.

*Administrators embraced CORE's focus on support over sanctions and the vision of peer collaboration.* Most district and school staff believed that the CORE intervention model was better suited for school improvement than NCLB sanctions and appreciated that accountability was intended to be used as a “flashlight not hammer.” One central office leader described the system as “not about putting the red scarlet letter, it's about providing supports.”

Interviewees generally believed the purpose of CORE interventions was to encourage the sharing of ideas and successful practices at the school and district levels through *mutual learning*, defined as a shift in the overall tenor

of improvement efforts, often contrasting it with 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.

Others believed that a key purpose of the CORE interventions is to *build capacity* to engage in continuous improvement activities. 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 were perceived as helping districts and schools learn how to solve their problems (by engaging in OL), rather than accepting prescribed solutions. Specifically, this approach intended to promote trial and error and emphasized the importance of contextual fit over the sharing of "best practice."

Interviewees at about half the districts believed that *networking* through interaction among schools and districts was itself a main purpose of the CORE interventions and facilitated mutual learning and capacity building. 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, was thought to facilitate the sharing of "best practices" and innovative ideas, and was intended to allow for reciprocal, professional accountability. 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.

### *How the CORE Accountability System Was Implemented*

In terms of roll-out, districts adapted the MS and interventions substantially, leading to considerable variation and some concerns about alignment with other data systems. Although administrators greatly valued collaboration, particularly informal interactions, they often struggled to maintain reciprocity in those relationships. Limited capacity to engage with data and peer-to-peer interventions challenged implementation.

*All districts adapted the CORE accountability system to their local contexts.* Most districts embedded the MS into their own, more comprehensive indicator

systems. This was a conscious strategy on the part of central office leaders to build buy-in through coherence. 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.” 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 on college access and readiness, such as student eligibility and progress toward meeting college admissions criteria. In other districts with a history of pre-CORE reforms supporting improvement in SEL and climate, additional measures were added to the CORE 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.

Because of this adaptation, some interviewees expressed concerns about misalignment between the CORE MS and measures included in other data and accountability systems. For example, we were told about a school that had received state recognition for their work around improving EL redesignation but then scored low on the same indicator in CORE MS, since the two were calculated differently. 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

Similarly, districts adapted school pairing and COP work to fit local contexts, resulting in variation in the scope of their interventions. Most of the CORE Districts already used some form of school grouping, like Professional Learning Communities or principal supervision groups, with the intent of promoting cross-school learning. When the CORE interventions were rolled-out, however, districts varied in whether these interventions applied to only CORE-identified schools or all schools, and in whether they were integrated with existing cross-school collaborative structures versus treated separately. An administrator from one of two districts that integrated interventions in all schools said, “We have . . . [X] schools that were focus [COP] 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.’” Another district allowed non-CORE-identified schools to choose to participate in COPs based on preference. In the remaining three districts, COPs and Pairings were treated as distinct from existing district reform efforts. As a result, 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.

*Achieving reciprocity of school and district peer-to-peer interactions proved to be challenging.* Reciprocity represented a significant challenge to collaboration at the school and district levels. In school collaboration, although the initial intent of the Pairing intervention was coaching of a low-performing school by a high-performing partner, over time the goal changed to promoting two-way learning across both schools. In some cases, higher-performing Paired principals reported learning a great deal from their lower-performing Paired school. One of these principals shared, “We’re looking at [the lower-performing school’s] practices and how those practices can be brought back to our school to improve our school in the areas that we feel they’re doing well.” Leaders in other schools, however, raised questions about the potential for mutual learning. To begin with, as a higher-performing Paired principal shared, not all schools understood the specific roles of schools in Pairings: “It was never really clear or articulated what our role was as a reward [higher-performing Paired] school.” Moreover, some interviewees noted resistance to learn from schools perceived as poorly performing. As one principal said:

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.

The matching process and concerns about “fit” contributed to this challenge of ensuring mutual learning among schools in Pairings and COPs. Paired schools were primarily matched based on having similar demographics. As a result, interviewees expressed concerns that matched schools had different contexts, despite seemingly similar demographics. On the whole, principals in COPs and Pairings believed that cross-school collaboration functioned best when schools shared not only similar students, but also similar challenges and successes. In two districts, principals believed that they would be better served by selecting a partner based on a school’s specific areas of need.

Similar to school intervention work, several district leaders questioned how much they could learn from CORE district collaboration given the differences in local contexts. 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 CORE accountability system. For example, adequate parent engagement for a small district might

mean 150 parents, while a large district might wish to see many more parents engaged. As a result, comparing rubric ratings across districts complicated the review's accountability aim.

Moreover, challenges to reciprocity and the belief that both paired organizations were not equipped to reciprocally provide one another feedback and ideas may have interfered with cross-school mutual learning for improvement (discussed below). Administrators in several districts also 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 specific strategies or different implementation pieces, best practices that have worked for them that we can then use on our own.

While such relationships may be beneficial in the short term, they raise questions about the potential for longer-term engagement and continuous improvement, as we discuss later.

*Formal activities facilitated informal collaboration.* Throughout the year of study, we observed many meetings during which CORE staff facilitated discussions of specific data metrics and of implementation successes and challenges. In addition, our review of documents indicates educators spent considerable time on annual peer reviews. Although district administrators regularly participated in these formal activities, they tended to prefer, and value more, informal activities—such as contacting other CORE District administrators between meetings as issues arose. Nearly all the Superintendents reported routinely calling and texting each other to consult on emerging issues. Similarly, district administrators leading CORE work reported reaching out to the CORE community when working through implementation challenges. One district administrator explained:

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.

However, district administrators stated that formal, quarterly CORE meetings themselves were less helpful. An administrator noted the technical nature of discussions at times, saying, "I think some of those activities . . . 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 re-designation?" 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 the accountability system, which may have necessitated more focused discussion rather than learning opportunities. Notably, many district administrators acknowledged that the formal events were necessary to build the relationships enabling informal collaboration.

*Limited capacity created obstacles to 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 nonacademic data, the use of SEL and CC measures was very new. As expected in any situation when new measures are introduced, educators were still actively learning how to interpret and respond to them. In particular, few administrators articulated a clear understanding of specific SEL constructs and their measurement. Some administrators believed the lack of familiarity with and capacity to interpret the new nonacademic measures contributed to lower levels of use. Contrasting educators' familiarity with using academic data, one district administrator explained:

50% 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, because there was no previous measurement on strategies that actually might work, we are clueless.

Capacity constraints also affected school intervention work. In addition to the ubiquitous lack of time and inadequate funding, administrators spoke about inconsistency in the motivation, skills, and availability of facilitators. The role of the facilitator also 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 existing duties. In some cases, this meant facilitators were

not fully committed to the intervention work, rarely attended meetings, and did not properly review school plans. Overall, facilitation was not as substantive as many hoped and may not have optimally promoted learning among schools.

### *Intermediate Outcomes*

CORE's theory of action suggests that if implemented, new accountability measures, school supports, and cross-district collaboration will result in learning and changes in practice, which will ultimately lead to improved student outcomes. Here, we examine the intermediate outcomes of early implementation of these efforts. Across all elements of the CORE work, our evidence suggests that while progress has been made, more work is needed to fully achieve the vision of data-driven practice and deep learning resulting from peer-to-peer collaboration.

*District staff report beginning to use the MS to inform decisions.* 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. In all but two districts, administrators reported using the MS to identify resource needs, use, and effectiveness. One district used the CORE data to produce an "at-risk data report" for each school. Based on these data, the district assigned more staff to focus on improving results for these schools. In another district, central office leaders regularly asked principals to reflect on the 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?"

Other administrators used the MS data for improvement planning. Leaders in one district used the holistic data reports to reevaluate their view of school performance. An administrator shared an example of how the data led to re-assessing leadership effectiveness:

I think we have one school that's a classic example that . . . this leader, I think had been perceived for a long time as really effective . . . But then all the other indicators were orange and red. It became clear . . . "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?

Similarly, in two districts, school leaders used the MS to guide work with leadership teams, including using data to lead cycles of inquiry and

embedding the MS into their school goals. In another district, the SEL and CC lead and her team held meetings with school teams twice a year to look at these data and plan PD. 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, looking at student engagement activities that they could do at their school for kids to feel connected to the school.

Another district planned to integrate the SEL survey domains into the student report cards in the elementary grades, reporting results at the grade- and school-level.

Yet these examples of deep engagement with nonacademic data were more the exception than the rule. While educators reported ongoing use of academic data to guide their practice, not surprisingly, there was much greater variation in the depth of reported engagement with newer nonacademic measures. In fact, many educators questioned how “actionable” the SEL measures were.

*Despite evidence of productive learning, several district and school administrators noted the potential for distortive responses in a higher stakes setting.* While the potential for authentic learning was great, administrators across districts commonly cited concern that some of the MS metrics could incentivize distortive behavior that would preclude such learning and improvement, particularly when high stakes set in. For example, one principal worried that teachers might start grading students differently 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 0% 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). One central office administrator explained:



How do you prevent gaming on the surveys? . . . the minute you attach and 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 newer MS indicators driving true learning and improvement.

*Respondents reported fewer examples of deep learning resulting from school and district collaborative interventions.* At the school level, some administrators reported powerful learning, while others gleaned little from these collaborative interventions. In particular, some individuals questioned the appropriateness for such relatively “light touch” interventions to solve chronic performance problems in schools. As one district administrator stated, “[In] 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.” Nonetheless, administrators shared several examples of learning achieved through the CORE school-level interventions. These involved schools picking up “best practices” from other schools to facilitate their implementation of existing curricula and programs. For example, one higher-performing Paired principal shared that they had learned several logistical processes from their low-performing partner to “make our special education program more compliant.” This kind of learning involved error correction: recognizing that Individualized Educational Plans were not being submitted in a timely manner, this school learned how the lower-performing school managed their flow of paperwork.

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

I think the intervention work is helping us home 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 happening? And

how do we help them identify those areas of challenge and help them figure out why those are still areas of challenge?

District-level collaborative activities also appeared to result in useful technical problem solving, but fewer reports of deep learning. Much of the reported learning that took place, as captured through interview and observation, concerned solving pragmatic problems and developing messaging for rolling-out the MS. For example, district administrators discussed challenges in using technology for SBAC testing, metrics for measuring and strategies for improving SEL, managing relationships with data platform vendors, and designs for useful data reports. Administrators largely felt that they had learned best practices from others to tackle these and other problems and benefited from the common language the MS provided in their discussions with parents and faculty.

Furthermore, districts struggled to develop authentic professional accountability, as most interviewees described the peer review process as a compliance activity. One leader noted, “[it was] basically grading someone’s paper” with “very, very minimal” conversation. Another administrator reported that the peer review was “frustrating” and “cumbersome” and did not promote reflection.

This overall tendency to focus on compliance and technical problem solving is not surprising, but was dissatisfying to some district leaders, who expressed a desire to go deeper with the collaboration to “dig into to some nitty gritty problems of practice at a district level.”<sup>13</sup>

## **Conclusion and Discussion**

In summary, district and school administrators reported overall strong buy-in for the CORE accountability system. Interviewees endorsed the measures included in the MS for the enhanced focus on nonacademics and academic growth, yet some questioned the validity of certain measures and conveyed only emerging levels of understanding of the new SEL and CC indicators. They also appreciated the focus on support over sanctions, executed through collaborative interventions. In terms of roll-out, districts adapted the elements of CORE’s accountability system, in some cases combining the MS and interventions with existing systems. District administrators also noted challenges with school collaborative matching, fit and reciprocity. Across all elements of the CORE work, the perceived consequences indicated that implementation was still a work in progress. That is, while administrators reported using some MS measures to engage in planning and managing

resource allocation and improvement, they struggled to interpret and address SEL and CC data and worried about distortive responses. Furthermore, they reported that school and district collaboration resulted in technical problem solving more than deeper learning.

These findings echo extant literature on the challenges of implementing *external bureaucratic* accountability systems, particularly concerns about validity, fairness, and capacity (CEP, 2006; Stecher et al., 2008; Stecher et al., 2010), misalignments (Linn, 2005), and the potential for distortive behaviors (e.g., Booher-Jennings, 2005; Hamilton et al., 2007; Mintrop, 2012; Jennings & Rentner, 2006). Similarly, while literature suggests that *internal professional* accountability (peer review) can counterbalance bureaucratic accountability pressures and promote learning (O'Day, 2002), CORE faced considerable obstacles to achieving this balanced model. Furthermore, at this early stage of implementation, the public had not yet gained enough awareness and understanding of CORE's MS to promote *political* accountability (yet media attempts to facilitate this form of accountability suggest a potential narrowing of expectations that could conflict with CORE's aims).

While CORE's efforts were not universally successful in the early years, it takes time for systems to change and these efforts hold great potential. Although some readers may be inclined to write off this early CORE story as another example of inertial forces resisting change in schools (Hess, 2011; Tyack & Cuban, 1997), we see promise in educators' strong support for the underlying principles of CORE's accountability system and their commitment to learn from and adjust it over time. As such, CORE's experiences may be useful in informing the design and implementation of ESSA-aligned accountability systems under development around the country. Next, we consider a set of cross-cutting tensions and corresponding implications to inform future accountability policy.

### *Cross-Cutting Tensions and Implications*

Our analysis surfaced three key tensions in the implementation of CORE's accountability system across the participating districts, which we predict may emerge as states and districts implement less-standardized, multiple-measure accountability systems under ESSA. These tensions capture the challenge of introducing a new accountability system in settings with varied local values, balancing the desire to use performance measures for both accountability and continuous improvement, and ensuring the inputs necessary to achieve outcomes. Below we examine these tensions and implications for policy makers and practitioners seeking to mitigate them in the future.

*Customization versus standardization.* Many district leaders highly valued that CORE's accountability system was not a "one size fits all" model and allowed for local adaptation. As noted, many districts embedded the MS into their existing frameworks. Some were quite intentional in their efforts to customize, viewing it as a way to build support among educators who are generally suspicious of outside agencies' imposing accountability on them. "It's about ownership," said one district leader, "It's about the ability to remove a bogeyman: 'This outside group coming in measuring.' No, we measure. I think that's really important. . . . We know what a quality school is." Likewise, districts appreciated the ability to adapt the school interventions to their local context.

Yet this local adaptation also created challenges. First, at times it led to misalignments and confusion within districts, which led to inconsistency of awareness and implementation. More important, however, are the potential effects of local adaptation on the use of comparative data across districts and interventions. If districts decide that a broader set of measures, beyond those in the MS or slightly different from the MS, best captures local definitions of "success," then this raises serious questions about the meaning and value of the MS and its accompanying school rankings. Furthermore, if rankings based on MS data trigger consequences, identified schools may not in fact be the lowest performers most in need of support or the highest performers deserving accolades according to local values (a concern voiced by several interviewees). Similar to the contradictory pressures of standardization and local control identified in past studies of accountability (e.g., Graue, Wilinski, & Nocera, 2016), this tension observed in CORE districts suggests that those seeking to implement similar approaches consider both the importance of local buy-in and the potential challenges resulting from adaptation and variation. Notably, buy-in for performance measures may take more than customization, and too much variation may threaten the legitimacy of the system.

*Accountability versus continuous improvement.* By design, CORE leaders intended the accountability system to provide data to guide and hold schools accountable and interventions to help schools improve, and these dual purposes were viewed as mutually reinforcing. In theory, accountability, and the accompanying consequences, provided incentives for schools to engage in continuous improvement, while efforts to reflect on data and continuously learn were seen as the means to ensuring schools achieve accountability expectations.

While administrators agreed in principle with the intent of the accountability system, it was not always clear that its implementation accomplished these dual purposes. First, the variation in implementation and local adaptation may weaken the ability of MS results to advance accountability *and*

improvement. If what districts value locally strays from what gets measured in the formal accountability system, educators may start to see the CORE MS as compliance measures for external accountability, but not ones facilitating reflection and improvement. Similarly, compliant responses to interventions may undermine the opportunity for continuous improvement—an outcome observed in the implementation of past accountability systems (Manna, 2010).

Second, the concerns about strategic “gaming” behaviors reported above suggest that the MS could become more of an accountability tool than one facilitating continuous improvement. While gaming can lead to better results in the accountability MS (e.g., inflating grades to boost high school readiness indicators, not suspending students to keep rates low), it precludes true learning and improvement. This tension is well documented in studies of past accountability policies, as noted earlier. By design, the multiple measures embedded in the MS could reduce incentives for gaming (superficially boosting results on one measure has a proportionately smaller effect on the aggregate rating the more you increase the number of measures included and adjust weights). But if leaders continue to favor a “dashboard” approach to measurement that considers each indicator separately rather than the aggregate, single measure of performance in the index ratio, then the measures giving rise to potential distortive responses deserve more attention.

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 numbers but not their practices. To ensure productive responses to new measures (and the ultimate validity of results), administrators should carefully monitor schools and ensure consistent messages about the purposes and proper responses. To respond in meaningful ways to data and interventions, educators also 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 teacher and administrator preparation programs, supervisory supports, and in-service programs.

Finally, more experimentation and research is needed to evolve district peer-review into opportunities for true professional accountability and learning. Future design-based research might examine different models of peer review and other activities to inform improvements in this area.

*Inputs versus outcomes.* Several educators raised concerns about a classic accountability dilemma. While holding schools and districts accountable for

outcomes is desirable, and clearly preferable to a system that only measures inputs, some argued it was only fair to do so if there was simultaneous accountability for ensuring schools have the inputs needed to achieve those outcome goals. For example, one principal voiced strong concerns about being held accountable for outcomes without consideration for her lack of control over personnel decisions and the fact that her school was the district's "dumping ground" for ineffective teachers. With the shift in policy under ESSA and new flexibility around teacher evaluation, CORE's early results suggest the need for continued attention to the human capital inputs contributing to accountability and improvement.

A few CORE Districts were in fact systematically tracking information on inputs, but were doing so on their own volition and not as part of the CORE initiative. One district had conducted a systematic regional analysis of school choice patterns, enrollment patterns, facilities capacity, teacher turnover, and environment stress factors across the district. "That gives us a sense of equity issues that aren't visible when you just look at a list of schools and kind of rank order where they are," explained a district official. The district then used school level "scores" on an index of conditions in conjunction with outcome measures from the MS to drive resource allocations: schools with less favorable environmental conditions were given additional resources. Moreover, California's new finance system attempts to address the input side by providing greater flexibility around resource allocation and increasing funds for higher-needs students. However, some administrators implied that more was needed to guarantee, in the words of Richard Elmore, "reciprocity of accountability for capacity" (2004, p. 7). In other words, they wanted assurance that while they were responsible for improving, system leaders were equally responsible for ensuring their capacity to do so.

This tension highlights the critical need for attention to inputs, such as stable teachers and leaders, sufficient funding, and safety, as well as capacity building. Regardless of what approach to capacity building is taken, states and districts should consider other policy levers to ensure that all schools are staffed with effective and committed teachers and leaders who can take on improvement efforts and promote a culture that supports educators to reflect on data, to try out new strategies, and monitor progress. While not explicitly included in ESSA, leaders should attend to personnel policies that promote better recruitment, preparation, development, and retention of educators.

Second, to use new data to drive improvement, educators and leaders need to understand them 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 act (Marsh,

2012; Marsh, Bertrand, & Huguet, 2015). This capacity needs to be built around the newer academic and nonacademic measures via preparation and PD. The complexity of new data systems also requires greater communication strategies and support to help 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.

Finally, given the considerable challenges facing low-performing schools—such as a high student mobility and staff turnover, safety concerns, low morale, 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. Under certain conditions, other models may be needed. 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 intensive capacity building.

In the midst of planning for the future under ESSA, it behooves state and district policy makers to reflect on these issues, tensions, and questions. In fact, the CORE Districts have been doing just that, meeting regularly to reflect on and adjust their work, and are already moving into a new phase that builds on their collective learning. Their ongoing work will focus on using the principles of improvement science to learn from prior efforts and engage in continuous improvement.<sup>14</sup> Officials in other districts and states should likewise carefully consider not only the technical specifications of new accountability systems and interventions, but also the factors likely to facilitate implementation, to ensure that investments truly support ESSA’s improvement goals.

### **Acknowledgments**

We greatly appreciate the cooperation of the educators that participated in our research, as well as contributions from other members of our research team, including Vicki Park (University of Utah), Taylor Allbright (University of Southern California), Michelle Hall (University of Southern California), and Holly Glover (Stanford University).

### **Declaration of Conflicting Interests**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We are grateful to the generous sponsor of this research, the S. D. Bechtel, Jr. Foundation.

**Notes**

1. See Knudson and Garibaldi (2015) for a more detailed account of CORE's early years.
2. Given this history and prior efforts that built up district-level capacity to engage in reform, one might argue these districts are "atypical" of all California districts.
3. Our study did not focus on two other components of the CORE work under the waiver: (a) the implementation of high-quality instruction under the new standards and assessment of student learning and (b) teacher, principal, and superintendent evaluation and support systems.
4. 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 Index calculation. This means that a school could have a different "lowest performing racial ethnic group" across different index elements.
5. For a further discussion of student subgroup sizes and the impact on accountability measurement, see Hough and Witte (2016).
6. In 2014-2015, all of the CORE districts surveyed Grades 5 to 12, and some schools/districts surveyed Grades 3 and 4 additionally. In 2015-2016 and beyond, students in Grades 4 to 12 were surveyed in all districts. 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).
7. 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; Seashore Louis & Kruse, 1995; Stoll, Bolam, McMahon, Wallace, & Thomas, 2006). In this report, we use the capitalized term, Communities of Practice, to denote the specific intervention used as part of the accountability system.
8. 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."
9. The principal-agent problem has been studied by economists, political scientists, and other scholars, and focuses on the challenges of workers having goals and preferences that may not be consistent with those of their employers. When these goals are misaligned, agents are expected to act in ways to help improve their own outcomes, often at the expense of principals' goals.
10. Some might also categorize this as "community accountability" (Goldberg & Morrison, 2003).
11. Jennings (2012) defines "distortive" as the "use of test score data to make instructional and organizational decisions produc[ing] score gains that do not generalize to other measures of learning . . . and thus leads [teachers] to make invalid inferences about which schools, teachers, and programs are effective" (p. 4).



12. Given the newness of CORE's accountability system and the lack of prior knowledge, we believed it was essential to visit all six CORE districts to understand how the new system was being implemented. Furthermore, it is the cross-district understanding that promised to yield broader, nationally relevant insights for states and districts about to implement new state accountability systems under ESSA. For this reason we chose a multiple case study design rather than in-depth examination of one or two districts.
13. To this end, in the year following our data collection, the CORE districts have begun organizing as a Networked Improvement Community and focusing on a specific, measurable goal for cross-district school improvement, using MS data and the tools of Improvement Science (Bryk, Gomez, Grunow, LeMahieu, 2015). See the next Note for further discussion.
14. 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).

## References

- Argyris, C., & Schon, D. A. (1996). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison-Wesley.
- Blumer, H. (1954). What is wrong with social theory? *American Sociological Review*, 18, 3-10.
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42, 231-268.
- Burke, J. C. (2005). *Achieving accountability in higher education: Balancing public, academic, and market demands*. San Francisco, CA: Jossey-Bass.
- 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*. Cambridge, MA: Harvard Education Publishing.
- Carranza, R. (2015, September 20). *How best to measure success with English learners: Redesigning accountability*. Retrieved from <https://edsources.org/2015/holding-school-districts-accountable-for-success-with-english-learners/87121>
- Center on Education Policy. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Author.
- Charmaz, K. (2003). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies for qualitative inquiry* (2nd ed., pp. 249-291). Thousand Oaks, CA: Sage.

- Clough, C. (2016a, 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. (2016b, 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/>
- Cohen, M. D., & Sproull, L. S. (1996). *Organisational learning*. Thousand Oaks, CA: Sage.
- CORE. (2013). *ESEA flexibility request*. Retrieved from <http://coredistricts.org/wp-content/uploads/2013/02/CORE-ESEA-Flexibility-Request.pdf>
- Darling-Hammond, L., & Ascher, C. (1991). *Creating accountability in big city school systems* (Urban Diversity Series No. 102). New York, NY: National Center for Restructuring Education, Schools, and Teaching at Teachers College.
- Davidson, E., Reback, R., Rockoff, J., & Schwartz, H. L. (2015). Fifty ways to leave a child behind: Idiosyncrasies and discrepancies in states' implementation of NCLB. *Educational Researcher*, 44, 347-358.
- Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30, 418-446.
- Dee, T. S., Jacob, B., & Schwartz, N. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, 35, 252-279.
- Dixon, N. M. (1994). *The organizational learning cycle: How we can learn collectively*. New York, NY: McGraw-Hill.
- 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, 237-251.
- Elmore, R. F. (2004). *School reform from the inside out: Policy, practice, and performance*. Cambridge, MA: Harvard Educational Publishing Group.
- Figlio, D. N. (2006). Testing, crime and punishment. *Journal of Public Economics*, 90, 837-851.
- Firestone, W. A., & Shippy, D. (2005). How do leaders interpret conflicting accountabilities to improve student learning. In W. A. Firestone & C. Reihl (Eds.), *A new agenda for research in educational leadership* (pp. 81-100). New York, NY: Teachers College Press.
- Fullan, M., & Quinn, J. (2015). *Coherence: The right drivers in action for schools, districts, and systems*. Thousand Oaks, CA: Corwin Press.
- Goldberg, B., & Morrison, D. M. (2003). Co-nect: Purpose, accountability, and school leadership. In J. Murphy & A. Datnow (Eds.), *Leadership lessons from comprehensive school reforms* (pp. 57-82). Thousand Oaks, CA: Corwin Press.
- Graue, M. E., Wilinski, B., & Nocera, A. (2016). Local control in the era of accountability: A case study of Wisconsin PreK. *Education Policy Analysis Archives*, 24(60). doi:10.14507/epaa.24.2366

- 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.
- Hamilton, L., Stecher, B., Russell, J., Marsh, J., & Miles, J. (2008). Accountability and teaching practices: School-level actions and teacher responses. Strong states, weak schools: The benefits and dilemmas of centralized accountability. *Research in Sociology of Education, 16*, 31-66.
- Hentschke, G. C., & Wohlstetter, P. (2004, Spring/Summer). Cracking the code of accountability. *Urban Education, 17*-19.
- Hess, F. M. (2011). *Spinning wheels: The politics of urban school reform*. Washington, DC: Brookings Institution Press.
- Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization, 7*, 24-52.
- Honig, M. I. (2008). District central offices as learning organizations: How sociocultural and organizational learning theories elaborate district central office administrators' participation in teaching and learning improvement efforts. *American Journal of Education, 114*, 627-664.
- Hough, H. J., & Witte, J. (2016). *Making students visible: Comparing different student subgroup sizes for accountability*. Stanford: Policy Analysis for California Education. Retrieved from <http://www.edpolicyinca.org/publications/making-students-visiblecomparing-different-student-subgroup-sizes-accountability>
- Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science, 2*, 88-115.
- Jennings, J. L. (2012). The effects of accountability system design on teachers' use of test score data. *Teachers College Record, 114*(11), 1-23.
- Jennings, J., & Rentner, D. S. (2006). Ten big effects of the No Child Left Behind Act on public schools. *Phi Delta Kappan, 88*(2), 110-113.
- Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher, 34*(8), 3-13.
- Knapp, M. S. (2008). How can organizational and sociocultural learning theories shed light on district instructional reform? *American Journal of Education, 114*, 521-539.
- Knudson, J., & Garibaldi, M. (2015). *None of us are as good as all of us: Early lessons from the CORE districts*. San Mateo, CA: American Institutes for Research.
- Koretz, D. M. (1996). *Perceived effects of the Kentucky Instructional Results Information System (KIRIS)*. Santa Monica, CA: RAND.
- 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. Retrieved from [https://static1.squarespace.com/static/55bb6b62e4b00dce923f1666/t/57ea8a3cbe6594387dad0b11/1474988682108/Transforming+Education+Case+Study+FINAL+\(1\).pdf](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*. New York, NY: Cambridge University Press.
- Levitt, B., & March, J. G. (1988). Organizational learning. *American Review of Sociology, 14*, 319-340.
- Linn, R. L. (2005). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives, 13*, 33.
- Loeb, S., & McEwan, P. J. (2006). An economic approach to education policy implementation. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 169-186). Albany: State University of New York Press.
- Manna, P. (2010). *Collision course: Federal education policy meets state and local realities*. Thousand Oaks, CA: CQ Press.
- Marks, H. M., & Louis, K. S. (1999). Teacher empowerment and the capacity for organizational learning. *Educational Administration Quarterly, 35*, 707-750.
- Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record, 114*(11), 1-48.
- Marsh, J. A., Bertrand, M., & Huguot, A. (2015). Using data to alter instructional practice: The mediating role of coaches and professional learning communities. *Teachers College Record, 117*(4), 1-40.
- 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.
- McNeil, J. D. (2014). *Contemporary curriculum: In thought and action*. Hoboken, NJ: Wiley.
- Mehta, J. (2013). *The allure of order: High hopes, dashed expectations, and the troubled quest to remake American schooling*. New York, NY: Oxford University Press.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks: Sage.
- Mintrop, H. (2012). Bridging accountability obligations, professional values and (perceived) student needs with integrity. *Journal of Educational Administration, 50*, 695-726.
- Murnane, R. J., & Papay, J. P. (2010). Teachers' views on No Child Left Behind: Support for the principles, concerns about the practices. *Journal of Economic Perspectives, 24*, 151-166.
- Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics, 92*, 263-283.
- O'Day, J. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review, 72*, 293-329.
- 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.

- Price, H. E. (2010). Does No Child Left Behind really capture school quality? Evidence from an urban school district. *Educational Policy, 24*, 799-814.
- Reback, R., Rockoff, J., & Schwartz, H. L. (2011). *Under press: Job security, resource allocation, and productivity in schools under NCLB* (NBER Working Paper No. 16745). Cambridge, MA: National Bureau of Economic Research.
- Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida Heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy, 5*, 251-281.
- Schechter, C. (2008). Organizational learning mechanisms: The meaning, measure, and implications for school improvement. *Educational Administration Quarterly, 44*, 155-186.
- Seashore Louis, K., & Kruse, S. D. (1995). *Professionalism and community: Perspectives on reforming urban schools*. Thousand Oaks, CA: Corwin Press.
- Sherer, J. Z., & Spillane, J. (2011). Constancy and change in work practice in schools: The role of organizational routines. *Teachers College Record, 113*, 611-657.
- Smith, M. L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice, 10*(4), 7-11.
- Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational routines as coupling mechanisms policy, school administration, and the technical core. *American Educational Research Journal, 48*, 586-619.
- 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*. Santa Monica, CA: RAND Corporation.
- Stecher, B. M., Vernez, G., & Steinberg, P. (2010). *Reauthorizing no child left behind: Facts and recommendations* (Vol. 977). Santa Monica, CA: RAND.
- Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change, 7*, 221-258.
- Tyack, D. B., & Cuban, L. (1997). *Tinkering toward utopia*. Cambridge, MA: Harvard University Press.
- U.S. Department of Education. (2012) *ESEA flexibility*. Retrieved from <http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>
- U.S. Department of Education. (2016). *Summary: Proposed regulations on accountability, state plans, and data reporting under ESSA*. Retrieved from <http://www2.ed.gov/policy/elsec/leg/essa/essaaccountabilitynprmsummary52016.pdf>
- Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Boston, MA: Harvard Business School Press.
- West, M. (2007). Testing, learning, and teaching: The effects of test-based accountability on student achievement and instructional time in core academic subjects. In C. E. Finn Jr. & D. Ravitch (Eds.), *Beyond the basics: Achieving a liberal education for all children* (pp. 45-61). Washington, DC: Thomas B. Fordham Institute.

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

### **Author Biographies**

**Julie A. Marsh**, PhD, is an associate professor of education policy at the University of Southern California's Rossier School of Education and specializes in research on K-12 policy. Her research blends perspectives in education, sociology, and political science. Her research focuses on the implementation and effects of accountability and instructional reform policies, including the roles of central office administrators, intermediary organizations, and community members in educational reform and the use of data to guide decision making.

**Susan Bush-Mecenas** is a PhD candidate and Provost's fellow at the University of Southern California's Rossier School of Education. Her research interests include organizational learning, district reform, district and school capacity building, and accountability policy.

**Heather Hough**, PhD, is executive director of the CORE-PACE Research Partnership at Policy Analysis for California Education. Her area of expertise is in district- and state-level policymaking and implementation, with a particular focus on teacher compensation, support, and accountability; policy coherence; and system improvement.