The Properties of Non-Academic School Performance Measures

Rachel S. White
Old Dominion University

Morgan S. Polikoff
University of Southern California

Although there is a robust body of literature studying targets for academic indicators within school quality systems few studies explore target setting for non-academic indicators. Focusing on elementary schools within the CORE districts, we investigate how moving performance targets for non-academic indicators affects school quality ratings. We ask: (1) How does school performance on CORE's school quality improvement measures vary across schools and over time?; and (2) How does the setting of targets on CORE's non-academic indicators at various levels impact the number and types of schools that make progress toward or reach the target? We find that non-academic measures of school quality are less stable over time than static academic achievement measures, school demographics are not consistently associated with schools' ratings on non-academic measures, and schools' ratings on non-academic measures of school quality are sensitive to even very small changes in rating category thresholds.

VERSION: May 2019

The recent inclusion of non-academic measures in school performance measurement and accountability systems has been hailed as a way to hold schools accountable for the many functions they serve and provide the public with a more comprehensive understanding of the quality and outcomes of its public schools. These diverse outcomes of interest include schools' abilities to prepare students to excel academically and master K-12 curriculum standards, support students' social and emotional learning and growth, and foster a healthy, safe, and inclusive school culture and climate.

School performance measurement and accountability systems serve a number of functions. These include: identifying schools in need of additional resources and assistance; providing the general public with data about how the local schools that they fund through their tax dollars are performing; and celebrating and sharing the success of schools (Brooks, 2000). As such, most school performance measurement and accountability systems include performance targets or thresholds that make the data easily interpretable. Indeed, many statewide school accountability systems do not solely report averages of measures and scores, but instead give schools all-encompassing ratings based on pre-set thresholds for each of the rating categories. In some cases, these thresholds are criterion referenced – that is, thresholds are set based on performance relative to the goal. For example, if one goal is for schools to achieve a 100 percent attendance rate, the criterion referenced threshold in a system with five categories may set category thresholds every 20 percentage points (e.g., <19 percent, 20-39 percent, 40-59 percent, 60-79 percent, >80 percent). Schools would then be placed in their respective category regardless of how many other schools are in that category. However, this approach may result in widely skewed performance distributions among schools, making it difficult to differentiate among schools. Alternatively, in a system with norm-referenced thresholds, performance category floors and ceilings may be set based on performance relative to other schools. In particular, the setting of norm-reference thresholds may begin with the analysis of data to understand baseline measures of schools, followed by the setting of thresholds so that the number of schools in each rating category follows a certain distribution pattern (e.g., uniform, normal).

The setting of thresholds is an important component of any state or local school performance measurement or accountability system. In particular, though many schools categorized at "low performing" on school performance measures receive additional support from the district and/or state, research has shown that poor performance ratings affect the public's satisfaction with schools (Jacobsen, Snyder & Saultz, 2014), parental satisfaction with schools (Charbonneau & Van Ryzin, 2011); and accelerate teacher turnover in low-performing schools (Clotfelter, Ladd, Vigdor & Diaz, 2004; Feng et al., 2010). As such, the extent to which a school is just above or below particular performance category thresholds could have major implications for the public's perception of the quality of the school, as well as for the school's future performance. Put broadly, the information produced by non-academic indicators of school quality will undoubtedly be "recontextualized within a broader set of understandings about schools" (Bryk & Hermanson, 1993, p. 462).

While there is a robust body of literature related to target setting for academic indicators, such as standardized test scores, within state accountability systems (Ladd & Walsh, 2003; Richards & Sheu, 1992), few studies have explored target setting for non-academic indicators in school performance measurement or accountability systems. Therefore, we know very little about how and the extent to which ratings of schools' non-academic performance may shift from year-to-year and may be sensitive to a school's student demographics as well as the targets that are set by designers of the performance measurement system.

Focusing on schools within CORE Districts—Fresno Unified, Garden Grove Unified, Long Beach Unified, Los Angeles Unified, Oakland Unified, Sacramento City Unified, San Francisco Unified and Santa Ana Unified—we investigate the stability of non-academic performance indicators across time and schools, as well as how moving performance targets for non-academic indicators affects schools' performance ratings. The CORE Districts, among other innovative work, are engaged in efforts "to use more than just test scores to measure strengths and weaknesses in schools and to identify those in need of improvement" (CORE, n.d.). As such, the CORE districts have developed an extensive collection of data that informs a comprehensive school improvement. Using data that informs this system, we ask: (1) How does performance on CORE's School Quality Improvement measures vary across schools and over time? and, (2) How does the setting of targets on CORE's non-academic indicators at various levels impact the number and types of schools that make progress toward or reach the target? Our findings can help states and district consortia think about the ways to better design performance measurement systems to target the schools most in need of support and improvement.

Background

As detailed in this section, contemporary education reforms have resulted in more robust, transparent school performance measurement systems, and the types of data being collected and reported in these systems are expanding. Nonetheless, federal policies continue to provide substantial flexibility around some aspects of school performance measurement systems. Relevant to this report, federal law currently allows flexibility in how indicators of school performance and quality included in school performance and accountability systems are measured, and how school performance measurement systems differentiate schools based on their performance on said indicators.

School Accountability Policy and Indicators of School Quality and Performance

Over the past two decades, federal, state and local authorities have established systems that aim to measure school quality and hold schools accountable for students' academic performance. For example, states like Massachusetts, North Carolina, Texas and Washington began to rate schools based on student performance and make such ratings available to the public in the early 1990s (MA 603 CMR 2.04(3)(c); NC S.L. 1995-716; 19 TexReg 8979, 1994; WA ESSB5418, 1993). As shown in Table 1, these state-initiated systems focused primarily on student performance on state standardized assessments and the identification of low-performing schools.

At the federal level, contemporary school accountability policies were formally established in the early 1990s and have undergone multiple iterations to present. First, under the Improving America's Schools Act (IASA)—the 1994 reauthorization of the Elementary and Secondary Education Act (ESEA)— states were required to establish yearly goals related to student academic progress and assess students at least once in elementary, middle and high school in at least mathematics and reading or language arts. Results from these assessments were to be "used as the primary means of determining the yearly performance of each local education agency and school" and "publicize and disseminate to teachers and other staff, parents, students and the community the results," including results disaggregated by student subgroups (Pub.L. 103-382 §1111(b)(3)(I); §1116(a)).

With the passage of the No Child Left Behind Act (NCLB) in 2001, IASA requirements were expanded to require testing of students in reading and mathematics in all of grades three through eight and in one high school grade, as well as science testing at least once in elementary, middle and high school. Similar to IASA, NCLB required states to establish adequate yearly progress (AYP) benchmarks and annual measurable objectives (AMOs) for student subgroups, determine what it meant for a student to be proficient on state assessments, and evaluate schools based on whether students—both in the aggregate and by subgroup—made AYP toward the ultimate goal of 100 percent proficiency in math and reading by 2014 (Pub.L. 107-110). Also similar to IASA, NCLB required states to disseminate an annual state report card that included information on student achievement on state assessments for all students, as well as for student subgroups.

While ESEA comes up for reauthorization every five years, U.S. Congress was unable to complete work on a reauthorization bill in 2007 and in subsequent years. As such, in 2011, contending that "NCLB requirements have unintentionally become barriers to State and local implementation and forward-looking reforms designed to raise academic achievement," the U.S. Department of Education began to allow states to apply for waivers related to NCLB requirements (USDOE, 2011, para. 2). The waivers provided state education agencies (SEAs) flexibility in meeting the 100 percent proficiency requirement of NCLB if they agreed to implement other federal school reform priorities. In particular, if SEAs agreed to include a measure of student growth in their calculation of school performance and established a system of identification and consequences for "reward," "focus," and "priority" schools, SEAs were then given flexibility in meeting their AMOs (USDOE, 2012; see Polikoff et al., 2014 for a discussion of waiver accountability systems).

Scholars, practitioners and policymakers criticized school performance measurement systems established under NCLB, the NCLB waiver system, and ESEA waiver system on a number of fronts, contending that the policies fostered perverse incentives to narrow curriculum and incentivized educators to focus on "bubble kids," a limited number of tested subjects and, in some cases, test preparation strategies (Dee, Jacob & Schwartz, 2013; Hamilton et al., 2007; Jennings & Bearak, 2014). Moreover, state school accountability systems put in place up to this point included school performance ratings systems based primarily on static measures of student performance on state standardized tests, which were found to be largely

correlated with student demographic characteristics and were not a robust measure of school effectiveness (Martinez-Garcia, LaPrairie & Slate, 2011).

Finally, in 2015, Congress enacted and President Barack Obama signed the latest reauthorization of ESEA—the Every Student Success Act (ESSA)—which is now current law. ESSA requires states to establish long-term goals, measure progress towards those goals, and publish data related to progress toward these goals in an annual school report. Building upon previous federal policy, each state is now required to measure progress towards goals they set in the following areas:

- Academic achievement, as measured by proficiency on annual assessments in mathematics and reading/language arts
- High school graduation rates
- Student growth, if determined appropriate by the State; or, another valid and reliable statewide academic indicator that allows for meaningful differentiation in school performance
- Progress in achieving English language proficiency; and
- At least one additional indicator of school quality or student success (SQSS) that allows for meaningful differentiation in school performance and is valid, reliable, comparable, and statewide (Pub.L. 115-141, §1111(c)(4)).

In response to ESSA's "additional indicator of SQSS" requirement, states began to incorporate various measures of school quality or student success into their school accountability systems, such as measures of student engagement (e.g., chronic absenteeism, suspension rates), access to and completion of advanced coursework, postsecondary readiness, and/or school climate and safety.

While many state ESSA plans make some mention of social-emotional learning, no state has included plans to fully utilize measures of social-emotional learning (SEL) in their state accountability system (ECS, 2017). Nonetheless, the CORE districts, a network of large urban districts in California, used flexibility afforded to them through the ESEA waiver process in 2013 to develop and implement an alternative school performance measurement system[1]. In particular, the CORE districts developed a performance measurement system that, in addition to traditional test-based measures (e.g., student achievement on state standardized assessment), includes survey-based measures of school culture-climate from the student, staff and parent perspectives; measures of SEL based on student surveys; and measures of chronic absenteeism, suspension rates, and college- and career-readiness. In the subsequent section, we provide a high-level overview of the more recent research related to these additional measures of school quality.

---

[1] This system supplements, and does not supplant, the state of California's school accountability system. CORE districts continue to receive a school accountability rating under the state system, while simultaneously receiving a rating under the CORE performance measurement system.

Research on the CORE Districts' Additional Indicators of SQSS

Social-emotional learning. Using data from the CORE districts, researchers have recently explored the development, interpretation, and implementation of SEL measures. In particular, West, Buckley, Krachman & Bookman (2018) explored how SEL competencies were selected for assessment, the process for curating & piloting student surveys, and preliminary evidence of the measures' reliability and validity as an indicator of school quality. Meyer, Wang & Rice (2018) studied the psychometric properties of the SEL measures and consistency of SEL measurement across grades & demographics groups. Subsequently, Meyer and colleagues used Item Response Theory (IRT) to develop scale scores for each SEL construct. Additionally, researchers examined how SEL develops in children over time and how these trends vary by student demographics (West et al., 2018); schools' contribution to students' SEL growth (Loeb et al., 2018); validity and reliability of CORE's student-report surveys on SEL (Gehlbach & Hough, 2018); the modeling, scaling and functionality of the SEL surveys (Meyer, Wang & Rice, 2018); and how districts and schools interpret SEL and how schools with strong self-reported data on SEL outcomes enact and support various conceptions of SEL and school-based practices intended to develop SEL (Marsh et al., 2018).

School culture and climate. A number of studies have focused on the relationship between school climate and student academic outcomes. For example, Berkowitz, Moore, Astor & Benbenishty (2017) conducted a review of research linking SES, school climate and academic achievement: positive school climate was found to mitigate the negative contributions of low SES on academic achievement. In particular, students and teachers in Chicago elementary schools that implemented the Comer School Development Program reported improving levels of school climate (Cook, Murphy & Hunt, 2000). Moreover, school climate, as reported on student, teacher/staff, and parent surveys, was positively related to mathematics and reading proficiency (Sherblom, Marshall & Sherblom, 2006).

One study by Gagnon & Schneider (2017) specifically focused on school culture as a part of a district's school performance measurement system. In particular, working in a mid-sized urban district in the state of Massachusetts, Gagnon & Schneider (2017) explored the relationship between the state's school accountability ratings and the district's comprehensive School Quality Framework that incorporated a larger set of school quality measures, including: teaching and the teaching environment, school culture, resources, student character, academic learning, and character and well-being. When it came to school culture, Gagnon & Schneider (2017) found that school culture metrics had a strong, positive relationship to the school's state accountability rating and a strong, negative relationship to school poverty.

English Learner Redesignation. Research has well documented the importance of monitoring EL redesignation. In particular, reclassifying an EL student too late has been shown to be associated with higher dropout rates, restricted access to honors and college preparatory coursework, decreased rates of college enrollment, and a greater likelihood of needing remedial coursework in college (Callahan, 2005; Cummins, 1980, 1981; Estrada, 2014; Flores & Drake, 2014; Harklau, 2002; Kanno & Kangas, 2014; Parrish et al., 2006; Silver et al., 2008).

Chronic Absenteeism and Suspension/Expulsion. While student chronic absenteeism or attendance rates have been reported in many states, districts, and schools for many years, under ESSA, many districts and schools are now held accountable for reducing student absenteeism and/or meeting specific thresholds for student attendance or absenteeism.

A review of research focused on absenteeism, as well as suspension and expulsion, underscores the importance of including such a measure in school accountability systems. Put simply, fewer days in school means fewer hours of instruction and results in less learning (Goodman, 2014). Indeed, presence in school is a robust predictor of student achievement (Chen & Stevenson, 1995; Gottfried, 2010; Gottfried & Kirksey, 2017) and high school graduation (Balfanz & Byrnes, 2012; Silver, Saunders & Zarate, 2008). Additionally, student absenteeism also has negative externalities: in schools with high chronic absence, the achievement of all students, not just those who are absent, is affected (Musser, 2011; Goodman, 2014; Gottfried, 2014). One potential explanation for this is that absent students often require remedial instruction, which adversely affects non-absent students because classroom instruction is slowed or large portions of instruction are dedicated to remediation (Monk & Ibrahim, 1994). For low-income children, reducing absenteeism is particularly important in the early years of schooling. For example, Ready (2010) found that children from low-income families with good attendance gained more literacy skills than peers from higher-income families during kindergarten and first grade.

Hutt (2018) outlines some of the challenges in producing uniform and reliable records on student attendance and suggests that we understand how incorporating attendance data into school accountability systems "might differentially affect individual schools and districts" (p. 8). Empirically, a number of studies have explored, at the student- or classroom-level, relationships between (1) absenteeism and student achievement on standardized tests or (2) absenteeism and students' perceptions of their personal educational experiences (e.g., boredom with content, irrelevant courses, bad relationships with teachers).

Our contribution. While the previously described research and literature offers useful evidence about the school performance measures that have recently been introduced in many state and local school accountability and/or school quality measurement systems, we know of few studies that examine target setting for these non-academic indicators. We also know a great deal about target setting on test-based measures of performance, but as states are considering the inclusion of a more diverse array of non-test measures within their school accountability systems, it is important to understand how these measures might thoughtfully be included into accountability systems as a whole. For example, a key feature in any performance measures system is setting annual targets and determining how high, low, realistic, and attainable the targets should be (Perie, Park & Klau, 2007). Additionally, with multiple goals in place for performance on both test- and non-test-based measures, local school and district leaders, and policymakers will want to ensure that the entire system is consistent and conveys a coherent theory of action (CCSSO, 2017). Our paper is a first effort to begin to

understand the properties of these measures and offer advice for how they might be better used in school performance management systems.

Data & Methods

Data

Data for this study come from over 800 elementary schools in the CORE districts of California. We focus on measures used in elementary schools, which make up about 60 percent of total schools in the CORE districts.[2] We use school-level data on 14 measures within the two domains of the CORE School Quality Improvement Index (SQII): four academic domain measures, which include academic achievement and growth in math and English language arts (ELA); and 10 social-emotional & culture-climate domain measures, which include suspension/expulsion rate, chronic absenteeism rate, English learner (EL) redesignation rate, four student social/emotional learning (SEL) measures, and three school culture and climate measures. Due to data limitations, we do not include the disproportionality in special education or college and career readiness measures that have been incorporated in to the CORE SQII.[3] Table 2 displays the number of elementary schools for which we have data for each SQII measure in school year.

---

[2] We focus on elementary schools because they represent a much larger number of schools than middle/high schools in our sample. As the patterns of association may differ across levels, it may be useful to replicate these analyses at other grade levels.

[3] At this time, the CORE districts are working with stakeholders to allay concerns with respect to include the disproportionality in special education measure in the School Quality Improvement Index (CORE Districts, 2017). Additionally, the CORE districts are currently working with Linked Learning certification and analytics to develop the college and career readiness measure.

Table 2. Number of distinct schools with data reported, by school year by grade level

	2014-15		2015-16		2016-17	
	Value Reported	No Value Reported	Value Reported	No Value Reported	Value Reported	No Value Reported
ELA Performance	800	59	801	117	801	43
Math Performance	800	59	801	117	801	43
ELA Growth Index	0	859	810	108	808	36
Math Growth Index	0	859	810	108	808	36
SEL Growth Mindset	679	180	733	185	155	689
SEL Social Awareness	678	181	734	184	155	689
SEL Self-Efficacy	678	181	692	226	155	689
SEL Self-Management	678	181	735	183	155	689
Student Culture-Climate	678	181	693	225	155	689
Staff Culture-Climate	593	266	670	248	151	693
Family Culture-Climate	0	859	691	277	157	687
Chronic Absenteeism	826	33	828	90	783	61
Suspension	773	86	828	90	830	14
EL Redesignation	604	255	717	201	654	190

Academic domain. The first academic domain measure in the CORE Districts' SQII is academic achievement—a static measure of a school's average absolute performance on standardized math and ELA assessments for all students in grade 3 through 8. The second academic domain measure is a measure of the average level of students' academic growth within a school. More specifically, the CORE academic growth model measures each student's growth relative to his/her academic peers while controlling for prior achievement and observable student demographics. Specific information about how the academic growth measure is calculated is available on the CORE Districts' webpage.[4]

Social-emotional and culture-climate domain. Within the social-emotional and culture-climate domain of the Core Districts' SQII, there are four SEL measures derived from surveys of students. The survey questions focus on views of students' own academic abilities (self-efficacy); ability to grow (growth mindset); ability to regulate emotions and behaviors (self-management); and understanding of others' perspectives and ability to empathize (social awareness).

Next, there are three school culture-climate measures derived from surveys of three populations: students (for elementary schools, surveys are administered to students in third

4 http://coredistricts.org/our-data-research/improvement-measures/

grade and above); teachers and staff; and parents, guardians and caregivers. Each population is surveyed on a range of school climate indicators that focus on four broad areas: teaching and learning, inter-personal relationships, safety, and school-community engagement. Survey questions in these four broad areas are combined to generate a single school-level culture-climate measure for each of the three populations.

The social-emotional and culture-climate domain also includes a measure of a school's suspension/expulsion rate, measured as the percent of individual students in a school that were suspended/expelled during the year; chronic absenteeism rate, measured as the proportion of students that have an attendance rate of less than or equal to 90 percent; and an EL progress measure, measured by dividing the number of ELs who were redesignated as English language proficient during the school year by the sum of the number of ELs who were redesignated plus all non-redesignated ELs with five or more years of instruction in US schools. The EL progress measure excludes ELs identified for special education who cannot take the California English Language Development Test with reasonable accommodation.

The central dataset used for our analysis are derived from the CORE Districts' SQII index levels, which are measured on a scale from 0-100, as well as the SQII status categories, which place schools into five California Accountability Dashboard Level (CADL) status levels based on their index score. Table 3 displays how the 14 SQII measures are translated in to the 0-100 SQII index level. Information on how the index levels map on to the CADL categories are available on the CORE District website.[5] For the CADL categories, the five levels, in order of lowest- to highest-performing are: red, orange, yellow, green and blue.

Methods

To explore how schools' performance on CORE's School Quality Improvement measures vary across schools and over time, we present descriptive statistics of the SQII measures for all three years for which we have data (2014-15, 2015-16 and 2016-17). Subsequently, we display transition tables of the SQII measures for the 2015-16 to 2016-17 school years. We also report Kendall's rank correlation coefficients, a measure of the degree of similarity between a school's 2015-16 and 2016-17 color ratings.

Next, we examine how the setting of targets on CORE's non-academic indicators at various levels impact the number and types of schools that make progress toward or reach the target. We use the 2015-16 data and shift the CADL threshold categories one index level unit to the left (i.e., narrowing the lowest-performing (red) category) and one index level unit to the right (i.e., widening the lowest-performing category[6]). Figure 1 displays one example of what this shift looks like for the elementary school-level SEL growth mindset measure. We focus on the 2015-16 school year data for two reasons. First, while we have data from the 2014-15 school year, this was a pilot year for the CORE districts. In particular, the CORE districts piloted

---

[5] Ibid.

[6] Due to the negative skew in the suspension rate measure, we impose a 0.25 index level unit shift in lieu of a one index level unit shift.

the culture-climate and SEL surveys in 2014 and conducted a CORE-wide field test of a refined set of items on both surveys in Spring 2015. The results of the field tests informed the baseline setting for inclusion in the CORE SQII, starting with the measurement of these factors in Spring 2016. The SEL and culture-climate surveys were formally incorporated in to the CORE SQII in the fall of 2016. Second, we do not have complete school-level social-emotional and culture-climate domain data for the 2016-17 school year, as it is still in the process of being cleaned.

Using the aforementioned data alongside school-level student demographic data for the 2015-16 school year, we employ descriptive statistics as well as ordinary least squares (OLS) regression models with district fixed effects to examine how a one-unit shift in the CADL threshold impacts the number and types of schools that are in each color category. In particular, we model

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + v_l + \epsilon_{il}$$

for i=1,…, n elementary schools in the CORE districts, where y_i is the CADL category for school i; β_0 is a constant; the X_i are the values of the following school-level characteristics: log enrollment, proportion of students that are white, proportion of students that are African American, proportion of students that are Asian, proportion of students identified as English learners, proportion of students identified as socio-economically disadvantaged, and proportion of students with disabilities; v_l is a vector of district fixed effects; and ϵ_{il} is a standard zero-mean error term. Significant coefficients in this model indicate that schools with more or less of the different characteristics are more or less likely to be sensitive to modest target setting shifts.

Figure 1. Example of right- and left-shifts in CADL category thresholds for analyses

	Red		Orange		Yellow		Green		Blue	
	Low	High	Low	High	Low	High	Low	High	Low	High
Current	0	54	55	58	59	64	65	70	71	100
Right shift	0	55	56	59	60	65	66	71	72	100
Left shift	0	53	54	57	58	63	64	69	70	100

Table 3. CORE District School Quality Improvement Index Measures

Measure	Measure Details	SQII Index Level example: In an elementary school of 1,000 students…
ELA Performance Math Performance	School-level average of the percent of students who met level 3 or above on the state standardized assessment.	If the average percent of students achieving level 3 or higher was 77, the school's performance level would be a 77 on the 0-100 index level scale.
ELA Growth Index Math Growth Index	School-level average of all growth percentiles (SGPs) generated for each student with similar performance histories and observable demographics. An SGP can range from 1 to 99, with 50 being average growth.	If the average student growth percentile for ELA was 77, the school's growth rating would be a 77 on the 0-100 index level scale
SEL Growth Mindset	The percent of favorable responses to the four growth mindset survey questions (denominator is the number of answers received, not number of students). Favorable responses are those in the top two categories among various 5-point Likert scale responses.	If all 1,000 students answered all four growth mindset survey questions for a total of 4,000 responses, and 3,000 of the responses were favorable, the school's SEL growth mindset rating would be 75 on the 0-100 index level scale.
SEL Social Awareness	The percent of favorable responses to the eight social awareness survey questions (denominator is the number of answers received, not number of students). Favorable responses are those in the top two categories among various 5-point Likert scale responses.	If all 1,000 students answered all eight social awareness survey questions for a total of 4,000 responses, and 3,000 of the responses were favorable, the school's SEL growth mindset rating would be 75 on the 0-100 index level scale.
SEL Self-Efficacy	The percent of favorable responses to the four self-efficacy survey questions (denominator is the number of answers received, not number of students). Favorable responses are those in the top two categories among various 5-point Likert scale responses.	If all 1,000 students answered all four self-efficacy survey questions for a total of 4,000 responses, and 3,000 of the responses were favorable the school's SEL growth mindset rating would be 75 on the 0-100 index level scale.
SEL Self-Management	The percent of favorable responses to the eight self-management survey questions (denominator is the number of answers received, not number of students). Favorable responses are those in the top two categories among various 5-point Likert scale responses.	If all 1,000 students answered all eight self-management survey questions for a total of 4,000 responses, and 3,000 of the responses were favorable, the school's SEL growth mindset rating would be 75 on the 0-100 index level scale.
Student Culture-Climate	Percent of students at the school who respond favorably to each of the 24 elementary-level culture-climate survey questions. For 18 of the questions, a favorable response is one in the top two categories among a 4-point Likert scale of: No, never; Yes some of the time; Yes, most of the time; and Yes, all of the time. For eight questions that are focused on safety/bullying, zero bullying incidents is considered favorable. As such, a favorable response for these eight questions is the lowest category among the 4-point Likert scale: No, never.	If all 1,000 students answered all 24 culture-climate survey questions for a total of 24,000 responses, and 20,000 of the responses were favorable, the school's student culture-climate rating would be 83 on the 0-100 index level scale.

Measure	Measure Details	SQII Index Level example: In an elementary school of 1,000 students…
Staff Culture-Climate	Percent of staff at the school who respond favorably to each of the 26 culture-climate survey questions. For 21 of the questions, a favorable response is one in the top two categories among various Likert scales. For five questions that are focused on safety/bullying, fewer incidents is considered favorable. As such, a favorable response for these five questions is the lowest category among the various Likert scales.	If there are a total of 250 teachers and staff and all 250 answered all 26 culture-climate survey questions for a total of 6,500 responses, and 5,000 of the responses were favorable, the school's staff culture-climate rating would be 77 on the 0-100 index level scale.
Family Culture-Climate	Percent of family members of students at the school who respond favorably to each of the 13 culture-climate survey questions. For all 13 of the questions, a favorable response is one in the top two categories among various Likert scales.	If there are a total of 1,500 family members of students in the school answered all 13 culture-climate survey questions for a total of 19,500 responses, and 15,000 of the responses were favorable, the school's family culture-climate rating would be 77 on the 0-100 index level scale.
Chronic Absenteeism Rate	Percent of students at the school in the current year who have an attendance rate of less than or equal to 90 percent. A student's attendance rate is calculated by dividing the number of days attended during the school year by the number of days in which the student was enrolled. Chronic absence is only determined for students enrolled in the district at least 45 days, which is one-quarter of a 180-day school year. If a student was enrolled in more than one school during the school year, the student's attendance and enrollment data are aggregated across all schools attended that year. The aggregated figures are then used to determine if the student was chronically absent, and the result for that student is associated with the student's last known school of attendance.	If 50 out of 1,000 students in the school have an attendance rate of less than 90 percent, the school's chronic absenteeism index level would be a 5 on the 0-100 index level scale.
Suspension/Expulsion Rate	The percent of individual students suspended/expelled at a school during the school year (i.e., unduplicated counts of students' suspensions divided by student enrollment).	If 100 different students, out of a total of 1,000 students in the school, were suspended or expelled during the school year, the school's suspension/expulsion index level would be a 10 on the 0-100 index level scale.
EL Redesignation Rate	Percent of ELs who redesignate at the school site in the current year, as measured in a percentage (0 to 100) of the number of ELs who redesignated in the school (no matter how long they've been ELs) divided by the number of ELs who redesignate at the school site in the current year + all 5-year plus non-redesignated ELs at the school.	If a school of 1,000 students had a total of 650 EL students, 150 of which redesignate but have been in EL status for less than five years and 500 of which have been in EL status for over five years, and 100 of the five-year plus students were redesignated, the school's EL redesignation rate would be a 20 on the 0-100 index level scale.

Results: How Performance on the CORE Districts' SQII Measures Vary Over Time

General Descriptives of CORE Districts' SQII Measures

SQII Index Levels. We begin with descriptive statistics related to the 0-100 index levels for the SQII measures for elementary schools within the CORE districts across all three school years for which we have data. As shown in Table 4, on average, schools consistently made gains in ELA and math achievement levels between 2014-15, 2015-16 and 2016-17. Additionally, compared to other measures of school quality, schools, on average, made the most positive growth in academic achievement measures. While schools, on average, made gains in ELA growth, math growth was approximately zero between 2015-16 and 2016-17.

The SEL and culture-climate measures appear to be less stable than the academic achievement and growth measures. Only one SEL measure—growth mindset—consistently improved across all three school years. For the other SEL measures, they either increased then decreased (social awareness, social efficiency, student and staff culture-climate scores) or decreased then increased (self-management) from year to year. With regard to EL redesignation, average EL redesignation rates declined between 2014-15 and 2015-16 and then increased between 2015-16 and 2016-17. Changes in schools' average chronic absenteeism, suspension rates, and family culture-climate were relatively minimal. In short, whereas the test-based measures have a general upward trend (which is typically the case on test-based measures, both in California and elsewhere), the other measures have different patterns of year-to-year change. This may complicate the narrative about school performance over time on these measures.

In addition to measuring performance on static, single-year measures of school quality, elementary schools in the CORE districts also receive a performance measure related to year-to-year improvements in SQII measures. Table 5 displays school-level descriptive statistics of the changes in index levels for each of the SQIIs. We find that, on average, schools' academic achievement, SEL growth mindset, and student culture-climate made positive year-to-year changes. Year-to-year changes for all other variables were less stable, with a positive change from 2014-15 to 2015-16 and a negative change from 2015-16 to 2016-17, or vice versa. Readers should keep in mind for both the static and year-to-year improvement SQII measures that the 2014-15 school year was a pilot year for the SEL and culture-climate surveys, and the averages and standard deviations shown in the columns showing changes from 2015-16 to 2016-17 in Table 5 is only based on a limited number of schools that have SEL and culture-climate survey data for the 2016-17 school year. As such, inferences across years should not be made and year-to-year trends cannot be inferred.

Table 4. Summary statistics for elementary school SQII Index Levels, all students, by year

	2014-15			2015-16				2016-17				
	n	mean	st.dev	n	mean	st.dev	1yr Δ in μ	n	mean	st.dev	1yr Δ in μ	2yr Δ in μ
ELA Performance	800	31.9	18.0	801	36.9	18.5	5.0	801	38.0	18.7	1.1	6.1
Math Performance	800	27.4	17.2	801	31.8	18.0	4.4	801	33.6	18.1	1.8	6.2
ELA Growth Index	0	-	-	810	52.2	28.1	NA	808	55.2	28.0	3.0	NA
Math Growth Index	0	-	-	810	53.5	28.9	NA	808	53.3	28.7	-0.2	NA
SEL Growth Mindset	679	62.0	8.2	733	62.6	8.5	0.6	155	65.3	8.6	2.7	3.3
SEL Social Awareness	678	71.2	5.7	734	73.2	5.9	2.0	155	71.2	5.5	-2.0	0.0
SEL Self-Efficacy	678	61.8	7.1	692	61.9	7.8	0.1	155	61.5	8.8	-0.4	-0.3
SEL Self-Management	678	71.5	71.2	735	71.3	7.2	-0.2	155	71.5	6.7	0.2	0.0
Student Culture-Climate	678	77.6	4.6	693	79.3	5.2	1.7	155	78.7	4.5	-0.6	1.1
Staff Culture-Climate	593	82.4	10.0	670	85.2	8.3	2.8	151	84.2	7.5	-1.0	1.8
Family Culture-Climate	0	-	-	691	93.4	3.1	NA	157	93.3	3.5	-0.1	NA
Chronic Absenteeism	826	11.5	8.9	828	11.5	8.5	0.0	783	11.9	7.7	0.4	0.4
Suspension	773	0.9	1.6	828	0.9	1.7	0.0	830	0.9	1.8	0.0	0.0

Table 5. Summary statistics for elementary school SQII index level changes, all students, by year-to-year change

	2014-15 to 2015-16			2015-16 to 2016-17			
	n	mean	st.dev	n	mean	st.dev	1yr Δ in μ change
ELA Performance	799	5.0	5.5	799	1.1	5.4	-3.9
Math Performance	799	4.4	5.7	799	1.9	5.5	-2.5
SEL Growth Mindset	678	1.0	6.7	155	4.7	7.1	3.7
SEL Social Awareness	677	1.7	4.2	155	-0.8	4.3	-2.5
SEL Self-Efficacy	677	0.0	5.9	155	-0.3	5.9	-0.3
SEL Self-Management	677	-0.2	4.4	155	-0.3	4.1	-0.1
Student Culture-Climate	678	1.8	3.6	155	0.1	3.6	-1.8
Staff Culture-Climate	576	2.9	6.7	143	-1.2	4.3	-4.1
Family Culture-Climate	0	-	-	155	0.1	2.8	NA
Chronic Absenteeism	824	0.0	3.5	782	0.2	3.6	0.2
Suspension	770	0.0	1.1	827	0.0	1.2	0.0
EL Redesignation	568	-5.8	14.2	624	3.9	14.1	9.7

California Accountability Dashboard Levels. Table 6 displays school-level descriptive statistics of the five-category CADL status measures for each of the SQII measures. Additionally, Figure 2 displays the proportion of schools that fall into each of the five CADL color categories for each SQII over time. Similar to the previously discussed index level changes, some SQII measures saw gains in each year, while others experienced a seesaw trend. Overall, the CORE elementary schools received higher rating on most measures by 2016-17 than in 2014-15, with the exception of minor declines for EL redesignation, self-efficacy, and chronic absenteeism.

Elementary schools are given a CADL for both static and year-to-year change measures for 12 of the 14 elementary-specific SQIIs (math and ELA growth are not included in the change accountability measures). Table 7 displays school-level descriptive statistics for the CADL change categories for each of the SQII measures. For most measures, schools were above a 3.0 on the change measure for 2014-15 to 2015-16. For 2015-16 to 2016-17, schools scored below a 3.0 on most measures. This trend implies some slowing of positive change, as implied by other analyses.

Figure 3 presents transition tables for the CADL categories for the elementary school quality measures between 2015-16 and 2016-17. Below each transition table, we also present Kendall's rank correlation coefficient, τ_b—which is an estimate of the degree of similarity between a school's 2015-16 and 2016-17 color ranking, allowing for ties in rank across years. For both ELA and math performance, the majority of schools are categorized as orange; and, for the static measures of performance, there is minimal movement in CADL categories from one year to the next. In contrast to the static measures of academic performance, the CADL categories associated with the academic growth measures are more evenly distributed among the color categories and we see more movement among the CADL categories, among both schools that improve and worsen their color category.

Compared to the academic achievement measures, there is a more even distribution of schools in the various CADL categories for the SEL measures, as is evidenced by the smaller Kendall's rank correlation coefficient of τ_b>0.75 for academic achievement measures and, for SEL measures, 0.40< τ_b<0.55. This is also the case with the culture-climate measures, although there is more of a positive skew and we see more positive movement, especially among those who received an orange and yellow rating in the 2015-16 school year in the culture-climate measures. This, again, is evidenced by a Kendall's tau correlation coefficient for the culture-climate measures (0.36< τ_b<0.45) that is substantially smaller than that of the academic achievement measures but only slightly smaller than that of the SEL measures.

The chronic absenteeism and suspension measures are relatively stable across time, as we see little movement among CADL color categories among schools between 2015-16 and 2016-17, which is supported by the relatively small Kendall's tau correlation coefficients: τ_b of 0.38 and 0.19 for chronic absenteeism and suspension, respectively. The same is true for the EL redesignation measure, which has quite a bit of stability (supported by a Kendall's tau correlation coefficient, τ_b of 0.19). For the suspension measure, we see little change in the large proportion of schools in the highest CADL category (blue). Similarly, for the EL GROWTH

16-17redesignation measure, we see little movement among the large proportion of schools in the lowest CADL category (red). For the chronic absenteeism measure, we see little movement among both the highest and lowest CADL color categories. Again, the finding of lower stability on the non-test measures implies that the ratings will jump around more from year to year. Explaining these year-to-year changes to both educators and the general public may be a new challenge for systems incorporating these less stable measures.

Table 6. Summary statistics for elementary school SQII CADL status categories, all students, by year

	2014-15			2015-16				2016-17				
	n	Mean	St.Dev.	n	Mean	St.Dev.	*1yr Δ in μ*	n	Mean	St.Dev.	*1yr Δ in μ*	*2yr Δ in μ*
ELA Performance	800	2.0	1.0	801	2.30	1.0	*0.3*	801	2.3	1.0	*0.0*	*0.3*
Math Performance	800	2.2	0.9	801	2.4	0.9	*0.2*	801	2.4	0.9	*0.0*	*0.2*
ELA Growth Index	0	-	-	810	3.1	1.4	*NA*	808	3.3	1.4	*0.2*	*NA*
Math Growth Index	0	-	-	810	3.2	1.4	*NA*	808	3.1	1.4	*-0.1*	*NA*
SEL Growth Mindset	679	3.0	1.3	733	3.0	1.4	*0.0*	155	3.4	1.3	*0.4*	*0.4*
SEL Social Awareness	678	2.7	1.3	734	3.1	1.3	*0.4*	155	2.7	1.4	*-0.4*	*0.0*
SEL Self-Efficacy	678	3.1	1.3	692	3.1	1.4	*0.0*	155	3.0	1.6	*-0.1*	*-0.1*
SEL Self-Management	678	2.9	1.4	735	2.9	1.4	*0.0*	155	2.9	1.4	*0.0*	*0.0*
Student Culture-Climate	678	2.8	1.3	693	2.3	1.3	*-0.5*	155	3.0	1.3	*0.7*	*0.2*
Staff Culture-Climate	593	2.9	1.3	670	3.3	1.3	*0.4*	151	3.1	1.3	*-0.2*	*0.2*
Family Culture-Climate	0	-	-	691	3.5	1.3	*NA*	157	3.4	1.4	*-0.1*	*NA*
Chronic Absenteeism	826	3.1	1.4	828	3.1	1.4	*0.0*	783	3.0	1.4	*-0.1*	*-0.1*
Suspension	773	4.3	1.1	828	4.3	1.1	*0.0*	830	4.3	1.1	*0.0*	*0.0*
EL Redesignation	604	2.6	1.5	717	2.2	1.3	*-0.4*	654	2.5	1.5	*0.3*	*-0.1*

Table 7. Summary statistics for elementary school SQII CADL change categories, all students, by year-to-year change

	2014-15 to 2015-16			2015-16 to 2016-17		
	n	mean	st.dev	n	mean	st.dev
ELA Performance	799	3.8	0.9	799	3.0	0.0
Math Performance	799	3.7	1.1	799	3.2	1.1
SEL Growth Mindset	678	3.2	1.3	155	3.9	1.2
SEL Social Awareness	677	3.4	1.2	155	2.7	1.3
SEL Self-Efficacy	677	3.0	1.2	155	2.9	1.3
SEL Self-Management	677	2.9	1.2	155	2.8	1.2
Student Culture-Climate	678	3.8	1.3	155	3.0	1.4
Staff Culture-Climate	576	3.6	1.3	143	2.6	1.1
Family Culture-Climate	-	-	-	155	2.9	1.4
Chronic Absenteeism	824	3.0	1.2	951	3.4	0.7
Suspension	770	3.2	0.8	678	3.2	0.9
EL Redesignation	568	2.3	1.4	624	3.4	1.5

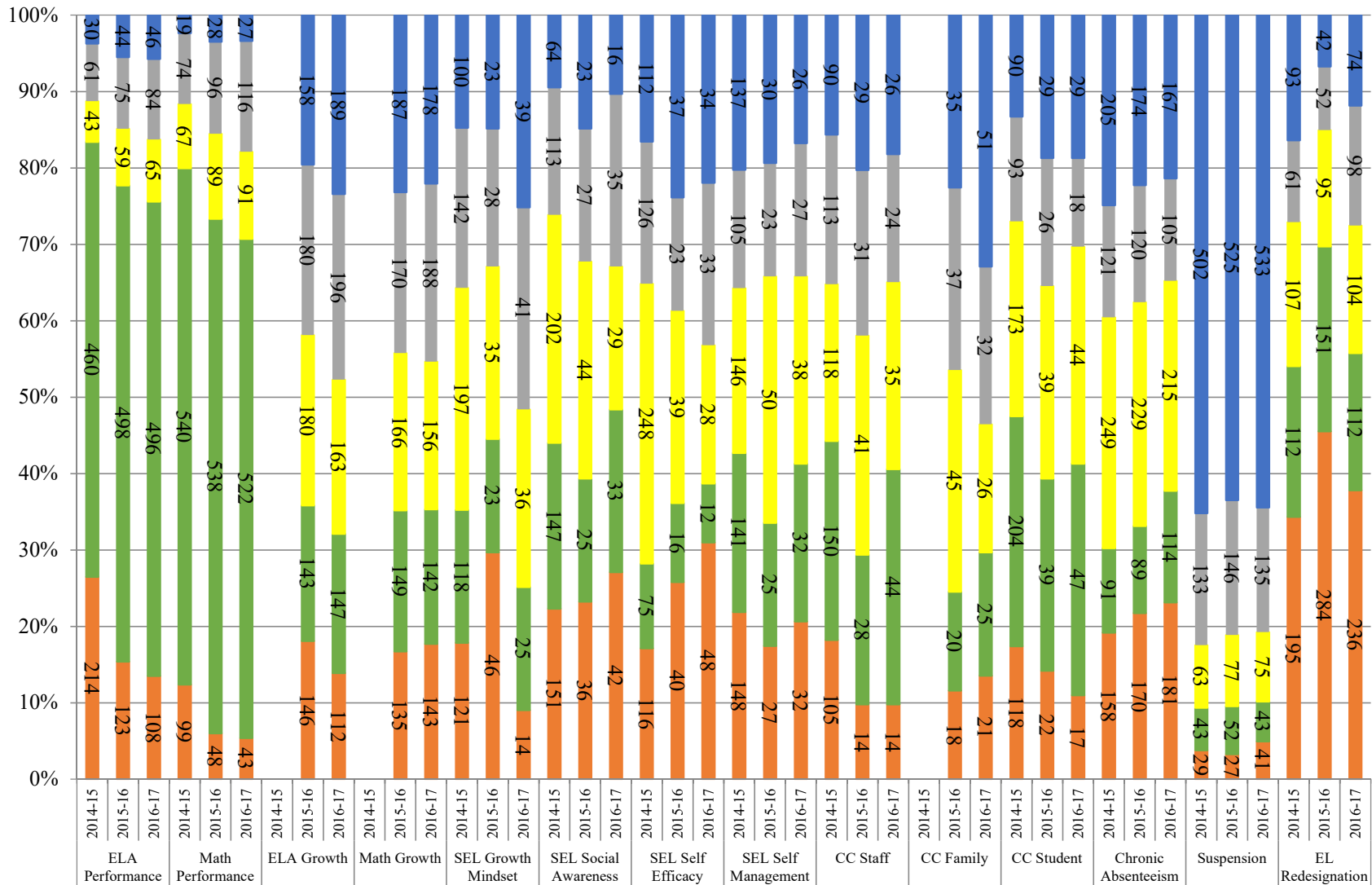Figure 2. Trends in elementary schools' CADL over time, by SQII measure, all students

Figure 3. Transition Tables

2016-17 ELA Performance

2015-16 ELA Performance	Red	Orange	Yellow	Green	Blue	
Red	78	45	0	0	0	123
Orange	30	438	25	5	0	498
Yellow	0	13	33	13	0	59
Green	0	0	7	60	8	75
Blue	0	0	0	6	38	44
0	108	496	65	84	46	799

Kendall's rank correlation coefficient τ_b=0.80

2016-17 Math Performance

2015-16 Math Performance	Red	Orange	Yellow	Green	Blue	
Red	28	20	0	0	0	48
Orange	15	486	35	2	0	538
Yellow	0	16	50	23	0	89
Green	0	0	6	87	3	96
Blue	0	0	0	4	24	28
0	43	522	91	116	27	799

Kendall's rank correlation coefficient τ_b=0.77

2016-17 ELA Growth   Orange

2015-16 ELA Growth	Red	32	Yellow	Green	Blue	
Red	38	39	31	26	19	146
Orange	25	33	21	35	23	143
Yellow	28	24	34	41	44	180
Green	12	19	54	44	46	180
Blue	9	147	23	50	57	158
	112		163	196	189	807

Kendall's rank correlation coefficient τ_b=0.22

2016-17 Math Growth   Orange

2015-16 Math Growth	Red	28	Yellow	Green	Blue	
Red	46	28	21	27	13	135
Orange	33	44	37	30	21	149
Yellow	21	22	27	38	36	166
Green	24	20	42	44	38	170
Blue	19	142	29	49	70	187
	143	x	156	188	178	807

Kendall's rank correlation coefficient τ_b=0.11

	2016-17 SEL Growth Mindset				Orange
2015-16 SEL Growth Mindset	Red	19	Yellow	Green	Blue
Red	10	1	12	2	3	46
Orange	3	3	6	12	1	23
Yellow	1	2	15	9	7	35
Green	0	0	2	12	12	28
Blue	0	25	1	6	16	23
	14		36	41	39	155

Kendall's rank correlation coefficient τ_b=0.42

	2016-17 SEL Social Awareness				Orange
2015-16 SEL Social Awareness	Red	12	Yellow	Green	Blue
Red	23	5	1	0	0	36
Orange	9	14	6	4	1	25
Yellow	8	1	11	10	1	44
Green	1	1	9	11	5	27
Blue	1	33	2	10	9	23
	42		29	35	16	155

Kendall's rank correlation coefficient τ_b=0.51

Orange

2016-17 SEL Self-Efficacy

2015-16 SEL Self-Efficacy

	Red	Orange	Yellow	Green	Blue	
		4				
Red	27	2	5	4	0	40
Orange	7	5	5	2	0	16
Yellow	12	1	11	10	1	39
Green	2	0	4	9	7	23
Blue	0	12	3	8	26	37
	48		28	33	34	155

Kendall's rank correlation coefficient τ_b=0.43

2016-17 SEL Self-Management   Orange

2015-16 SEL Self-Management

	Red	Orange	Yellow	Green	Blue	
		5				
Red	20	12	2	0	0	27
Orange	5	13	5	2	1	25
Yellow	5	2	21	9	2	50
Green	2	0	9	7	3	23
Blue	0	32	1	9	20	30
	32		38	27	26	155

Kendall's rank correlation coefficient τ_b=0.54

2015-16 CC Family   Orange

2016-17 CC Family

	Red	7	Yellow	Green	Blue	
Red	8	5	2	1	0	18
Orange	6	10	4	3	2	20
Yellow	3	3	13	7	12	45
Green	3	0	5	13	13	37
Blue	1	25	2	8	24	35
	21		26	32	51	155

Kendall's rank correlation coefficient τ_b=0.45

2016-17 CC Students   Orange

2015-16 CC Students	Red	10	Yellow	Green	Blue	
Red	7	17	4	1	0	22
Orange	6	15	14	1	1	39
Yellow	4	5	15	3	2	39
Green	0	0	7	8	6	26
Blue	0	47	4	5	20	29
	17		44	18	29	155

Kendall's rank correlation coefficient τ_b=0.39

2015-16 CC Staff   2016-17 CC Staff   Orange

	Red	5	Yellow	Green	Blue	
Red	8	19	1	0	0	14
Orange	5	19	4	0	0	28
Yellow	1	1	17	2	2	41
Green	0	0	12	15	3	31
Blue	0	44	1	7	21	29
	14		35	24	26	143

Kendall's rank correlation coefficient $\quad \tau_b=0.36$

		2016-17 Chronic Absenteeism			Orange	
2015-16 Chronic Absenteeism	Red	20	Yellow	Green	Blue	
Red	139	30	10	0	1	170
Orange	32	60	23	4	0	89
Yellow	9	3	121	27	12	229
Green	1	1	49	45	22	120
Blue	0	114	12	29	132	174
	181		215	105	167	782

Kendall's rank correlation coefficient $\quad \tau_b=0.38$

		2016-17 Suspensions			Orange
2015-16 Suspensions	Red	7	Yellow	Green	Blue

Red	18	18	2	0	0	27
Orange	12	10	15	5	2	52
Yellow	8	5	28	23	8	77
Green	3	3	21	53	64	146
Blue	0	43	9	54	459	525
	41		75	135	533	827

Kendall's rank correlation coefficient τ_b =0.19

2015-16 EL Redesignation	\multicolumn{5}{c}{2016-17 EL Redesignation}	Orange
	Red	44	Yellow	Green	Blue	
Red	178	46	35	18	9	284
Orange	38	15	30	23	14	151
Yellow	15	5	20	33	12	95
Green	2	2	13	19	13	52
Blue	3	112	6	5	26	42
	236		104	98	74	624

Kendall's rank correlation coefficient τ_b=0.29

Characteristics of the Lowest Performing Schools

To examine the characteristics of the schools that may be predictive of which CADL categories a school receives for the various SQII measures, we employ OLS regression with district fixed effects in which we regress school-level characteristics such as the size of the school (i.e., enrollment) and the proportion of students that are (a) white, (b) African American, (c) Asian, (d) English learners (e) socio-economically disadvantaged, (f) students with disabilities on our dependent variable: CADL category.

Results of the 10 OLS regressions are presented in Table 7. The results illustrate that the relationships between schools' student demographics and CADL categories are quite consistent across nearly all SEL and culture-climate survey measures. In particular, controlling for school size and district variations, the more African American students, English learners, and students with disabilities in a school, the lower the school-level CADL category for SEL growth mindset, social awareness, self-management and student, staff and family culture-climate measures. This finding aligns with others that have found that white students report higher levels of SEL than African American and Latinx students (West et al., 2018). And, as West et al. (2018) note, it is important to keep in mind that students of different racial and ethnic backgrounds may respond to SEL survey items differently than their White peers due to out-of-school contexts. While the self-efficacy CADL category is also negatively associated with the proportion of students in a school that are English learners and students with disabilities, we find a positive association between the proportion of students in a school that are African American and a schools' self-efficacy CADL category.

When it comes to chronic absenteeism and suspension, the more African American students in a school, the lower the school-level CADL category for these two measures. Additionally, the proportion of students with disabilities in a school is negatively associated with the CADL categories for chronic absenteeism and EL redesignation.

One of the most significant findings is that one student demographic in schools is not significantly associated with schools' CADL categories for the SEL and culture-climate measures: socio-economic disadvantage. This finding suggests that, unlike school academic achievement levels, non-academic measures of school quality are not highly correlated with the socio-economic levels of students within the school, even after controlling for the size of the school and other student demographics (e.g., race, EL status, special education status). However, socio-economic disadvantage is negatively associated with schools' chronic absenteeism and suspension categories.

STable 7. Results of the regression analysis examining predictors of SQII social-emotional & culture-climate domain measures, 2015-16 school year, district fixed effects

Dependent variable→	SEL - Growth mindset	SEL - Social awareness	SEL – Self-efficacy	SEL – Self-management	Chronic absenteeism	Suspension	Student culture-climate	Staff culture-climate	Family culture-climate	EL Redesignation
Constant	5.37	4.31	4.95	4.96	3.85	6.04	5.66	8.04	7.72	0.46
Log(enrollment)	-0.10	0.02	-0.12	-0.03	0.31***	-0.08	-0.19*	-0.59***	-0.52***	0.38***
Percent white	0.01*	0.01***	0.02***	0.01***	-0.01*	-0.01*	0.003	-0.001	0.01	0.02***
Percent African American	-0.02***	-0.02***	0.01**	-0.02***	-0.04***	-0.01***	-0.04***	-0.03***	-0.03***	-0.01**
Percent Asian	-0.01***	0.003	0.01	0.02***	0.01	0.001	0.01*	0.02***	0.02***	0.03***
Percent ELs	-0.02***	-0.01***	-0.01***	-0.03***	0.001	0.001	-0.02***	-0.02***	-0.02***	0.004
Percent students with disabilities	-0.02**	-0.03***	-0.03***	-0.02**	-0.02***	-0.0002	-0.01	-0.02**	0.001	-0.02***
Percent socio-economically disadvantaged	-0.01	-0.01	-0.01	-0.01	-0.03***	-0.01***	0.001	0.002	0.001	-0.01
District fixed effects	X	X	X	X	X	X	X	X	X	X
n	733	734	692	735	828	828	693	669	690	643
R^2 (overall)	0.38	0.30	0.34	0.55	0.36	0.08	0.31	0.18	0.24	0.09
R^2 (within)	0.37	0.32	0.32	0.53	0.38	0.11	0.27	0.21	0.26	0.22
R^2 (between)	0.63	0.44	0.59	0.83	0.33	0.03	0.91	0.03	0.001	0.40

*p<0.10; **p<0.05; ***p<0.01

Sensitivity to CADL Category Thresholds

Finally, we examined schools' sensitivity to the CADL category thresholds for the SEL and culture-climate domain measures. As shown in Figure 4, there are substantial differences in the proportion of schools in each CADL color category in the current system than in a system in which the color categories were shifted one unit to the right (widening the lowest CADL category). In particular, while a one-unit shift to the right does not alter the number of schools in any of the CADL color categories for the staff culture climate measure in either 2015-16 or 2016-17, the proportion of schools in each CADL color category for all other social-emotional and culture-climate domain measures would look substantially different under these slightly modified thresholds. The most sensitive measure is the family culture-climate, where the one unit shift to the right would have resulted in a 50 percent increase in the number of schools receiving a red CADL rating in 2015-16 (from 18 schools under the current thresholds to 27 schools under the shifted thresholds) and a 66 percent increase in the number of schools receiving a red CADL rating in the 2016-17 school year (from 21 schools under the current thresholds to 35 schools under the shifted thresholds). It should be noted, however, that some measures have much tighter variances than others; as such, one-unit shift on these 0-100 measurement scales will, by definition, experience a greater shift in the CADL color category.

Schools' CADL color categories are similarly sensitive when the color category thresholds are shifted one unit to the left, widening the highest CADL category (blue). As shown in Figure 5, family culture-climate is, again, the most sensitive measure. Other measures that are quite sensitive include the SEL growth mindset and the chronic absenteeism measures. For example, by widening the blue CADL color category by one unit, the percent of schools that would receive the blue rating would have increased by 35 percent in the 2015-16 school year (from 174 schools under the current system to 235 schools under the shifted thresholds) and 29 percent in 2016-17 school year (from 167 schools under the current system to 215 schools under the shifted thresholds).

Next, we examined year-to-year changes in schools' CADL categories after increasing the width of the red category threshold by one index level unit. Figure 6 displays the number of schools that would experience a positive, negative, or no change in their CADL measure between 2015-16 and 2016-17 as a result of this index level unit change. For example, under the current system, 40 schools made a positive CADL color category change for the SEL social awareness measure between 2015-16 and 2016-17 (e.g., movement from red to orange, orange to green, yellow to green, orange to blue, etc.). By increasing the width of the red category threshold by just one unit and shifting all other CADL color categories by one unit accordingly, 8 fewer schools would make a positive color category change. This represents a 25 percent reduction in the number of schools that would make a positive color category change for SEL social awareness between 2015-16 and 2016-17.

Next, we decreased the width of the red category threshold by one point on the index level for all SQII measures. Figure 7 displays the number of schools that would experience a positive, negative, or no change in their CADL measure between 2015-16 and 2016-17 under

the current system and under the shifted system. For example, under the current system, 29 schools made a positive shift and 46 schools made a negative shift in CADL color category for the SEL self-management measure. By decreasing the width of the red category threshold by one unit and shifting all other CADL color categories by one unit accordingly, 36 schools would have made a positive shift and 39 schools would have made a negative shift in CADL color category for the SEL self-management measure, representing a 19 percent increase in the number of schools that would make a positive color category change and a 17.9 percent decrease in the number of schools that would make a negative color category change for SEL self-management between 2015-16 and 2016-17.

In sum, these findings suggest that schools are quite sensitive to the color category thresholds: slight shifts in the CADL color category thresholds result in a substantial number of schools that made CADL color category improvements to no longer make those improvements, schools that made CADL color category regressions to no longer make those regressions, and schools that did not make CADL color category improvements to either make a positive or negative shift.

Conclusion

This paper represents a first examination of some characteristics of non-academic measures in performance measurement and accountability systems. Using data from the CORE districts, the report describes trends in non-academic performance measurement ratings over time, associations of non-academic performance measurement ratings with school characteristics, and the sensitivity of ratings to small changes in cut scores. Based on these analyses, we reached several conclusions.

First, we concluded that year-to-year changes in non-academic measures were not especially stable. Typically, they appeared more stable than growth-based measures of achievement but less-stable than achievement levels. This finding means that schools' ratings will tend to jump around somewhat from year to year, which may encourage states using these measures to consider running averages or other techniques to reduce year-to-year volatility. We also did not observe many obvious trends in non-academic measures across years, but this may have been due to a limited number of years of data.

Second, we concluded that school demographics were associated with schools' non-academic ratings, though not always consistently across measures. For instance, the percent of students socioeconomically disadvantaged was strongly associated with chronic absenteeism and suspension rates, but not with SEL measures (controlling for student racial demographics). The opposite patterns were observed for racial demographic categories. Researchers should continue to probe the reasons for these associations and their implications for the fairness of performance measurement and accountability systems.

Third, we observed that school ratings were quite sensitive to even very small changes in the placement of category thresholds for many of the non-academic measures. In the most

extreme example, 66 percent more schools would have received a CADL color rating of red for the family culture-climate measure in 2015-16 if the red category was widened by one unit, and 63 percent more schools would have received a CADL color rating of blue for the family culture-climate measure in 2015-16 if the blue category was widened by one unit. This sensitivity also applies to year-to-year CADL color category changes. In the most extreme case, 30.8% fewer schools made a positive CADL color change between 2015-16 and 2016-17 on the family culture-climate measure when just a single point change in the threshold was applied. Of course, there is no "right" performance measurement or accountability rating, but these findings might prompt renewed consideration of the best approaches to finding performance measurement and accountability classification systems that are less sensitive to minor specification differences.

Figur 4. Percent change in number of schools in each CADL color category by widening the lowest CADL category (red) by one index level unit, 2015-16 and 2016-17
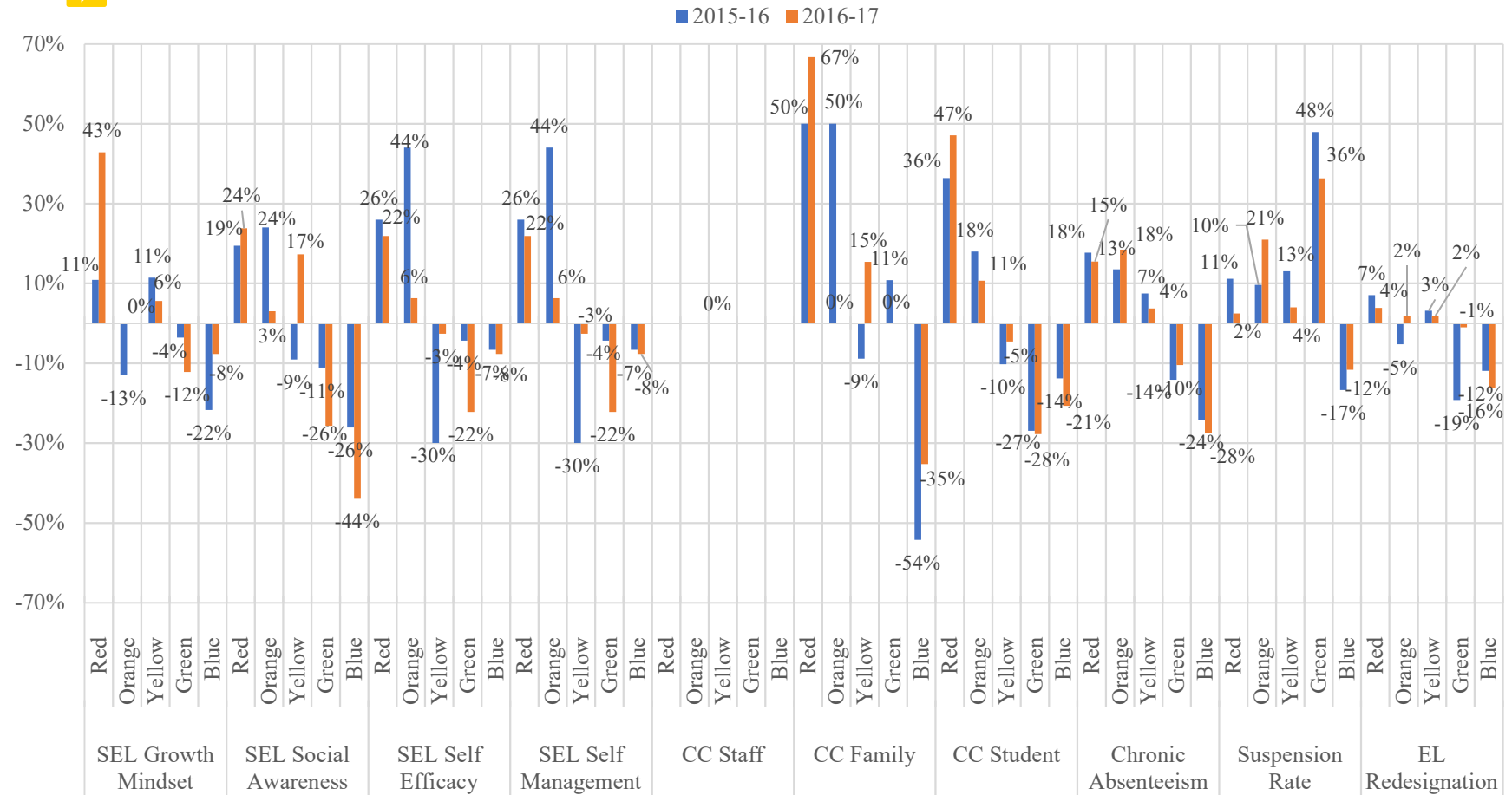
Figure 5. Percent change in number of schools in each CADL color category by widening the highest CADL category (blue) by one index level unit, 2015-16 and 2016-17
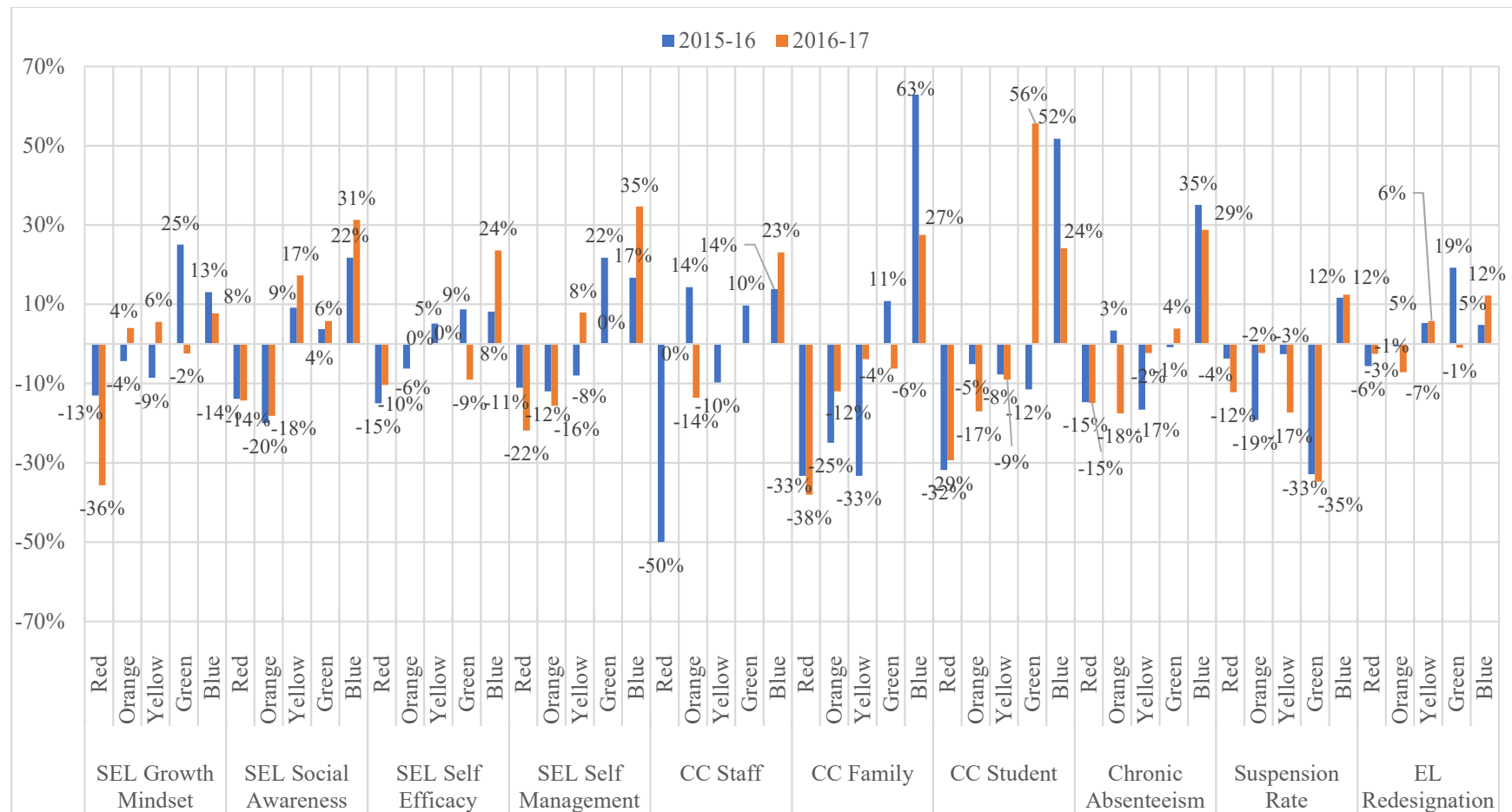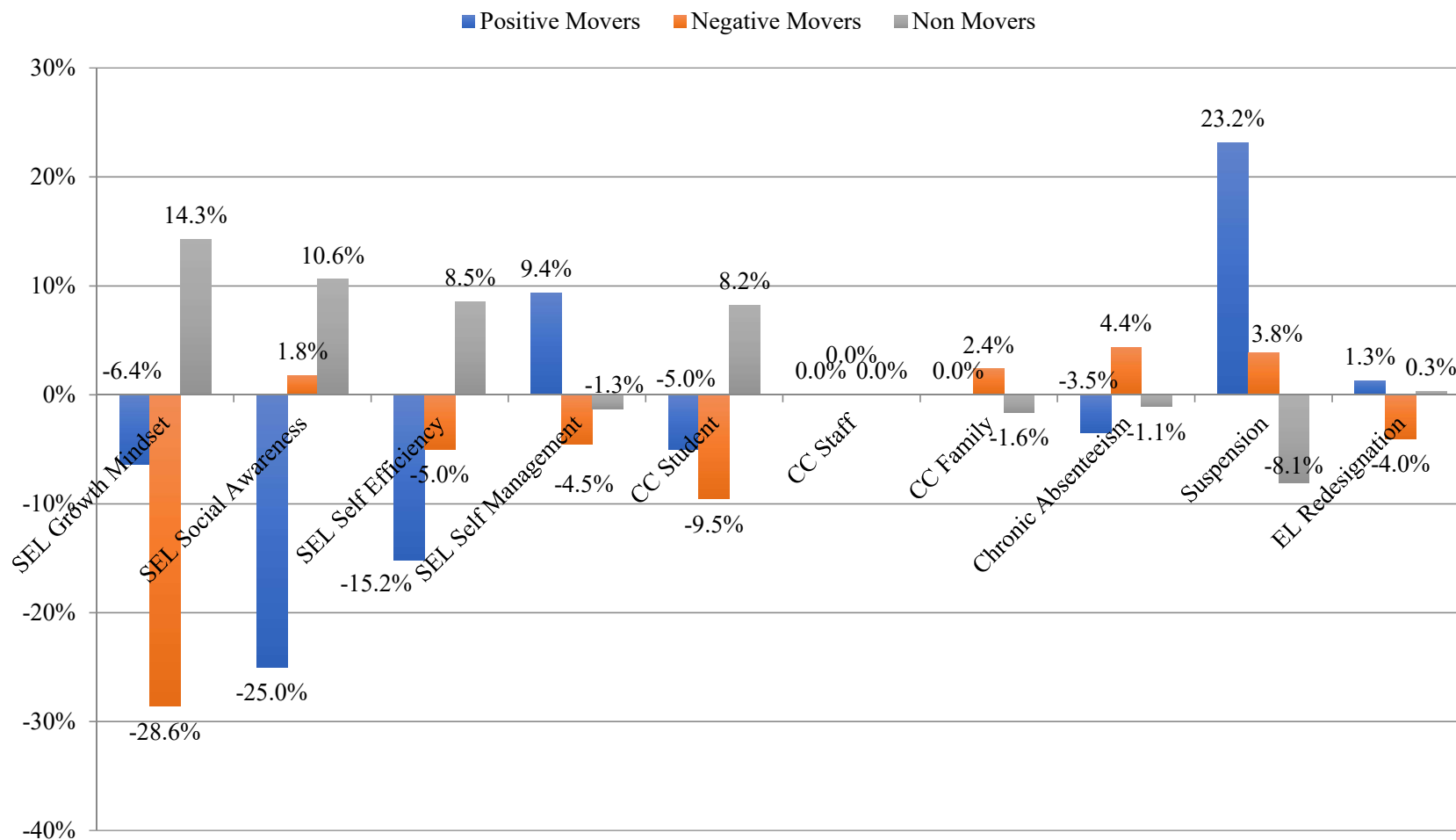
Figure 6. How widening the lowest CADL category (red) by one index level unit changes the number of schools that make positive, negative or no move in CADL category, 2015 -16 to 2016-17

■ Positive Movers   ■ Negative Movers   ■ Non Movers

Figure 7. How widening the highest CADL category (blue) by one index level unit changes the number of schools that make positive, negative or no move in CADL category, 2015-16 to 2016-17
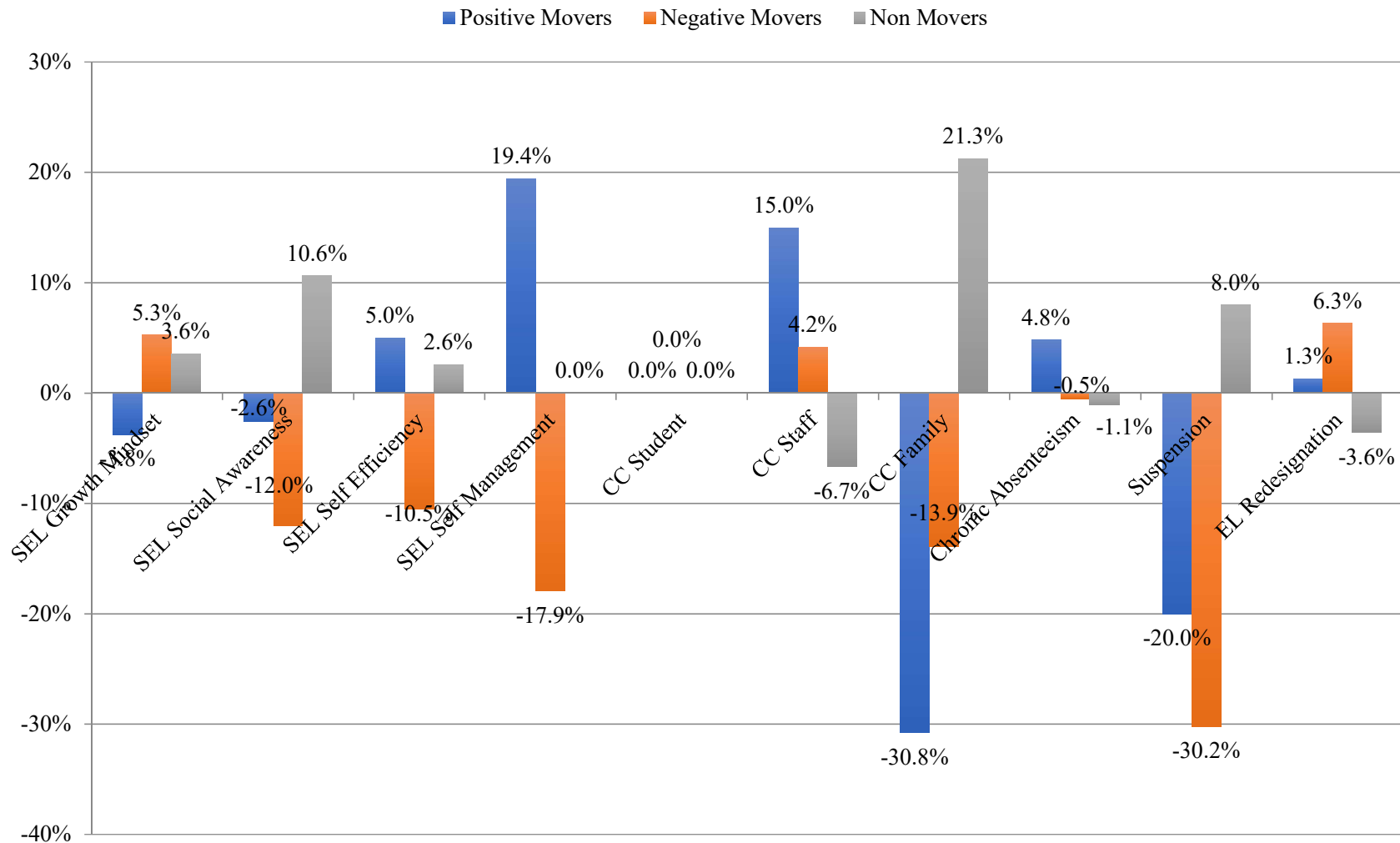
Table 1. Early state-level school performance management and accountability systems

State	Legal precedent	State performance management and accountability system details
Massachusetts	Administrative rules adopted by Board of Education June 16, 1997; Department guidelines approve September 28, 1999	**Under-performing schools and school districts.** 603 CMR 2.03 (1) The Board shall adopt, and the Department shall implement, a School Performance Rating Process to track the performance and improvement demonstrated by Massachusetts public schools on State assessments in core academic subjects. The School Performance Rating Process and its implementation shall be explained in written guidelines published by the Department. **Massachusetts Department of Education Guidelines for School and District Accountability System.** The Department will use the School Performance Rating Process to assess the extent to which all Massachusetts public schools are successfully preparing their students to demonstrate the skills and knowledge necessary to perform at an acceptable level on the MCAS tests in core academic subjects. While schools perform other important functions, such as supporting the healthy growth and development of our children, preparing students to meet State performance standards is at the core of every school's mission. The School Performance Rating Process will provide policy makers and the public with important information on the impact our State's education reform efforts have on student results. It will, at the same time, provide the Department with a valuable tool to identify: 1) schools with low performance that are not meeting improvement expectations; and 2) schools whose students have demonstrated impressive improvements or attained high levels of performance on MCAS tests.
North Carolina	Statute effective Date July 29, 1995 and modified to take into account recommendations of the Joint Legislative Education Oversight Committee on June 21, 1996	**School-based management and accountability program.** G.S. 115C-12(9) Among such duties [of the State Board of Education] are… To issue an annual 'report card' for the State and for each local school administrative unit, assessing each unit's effort to improve student performance based on the growth in performance of the students in each school and taking into account progress over the previous years' level of performance and the State's performance in comparison to other states. This assessment shall take into account factors that have been shown to affect student performance and that the State Board considers to be relevant to assess the State's efforts to improve student performance. […] c3. To develop a system of school building improvement reports for each school building. The purpose of the school building improvement reports is to measure improvement in the growth in student performance at each school building from year to year, not to compare school buildings. The board shall include in the building report any factors shown to affect student performance that the Board considers relevant to assess a school's efforts to improve student performance. Local school administrative units shall produce and make public their school building improvement reports. GS 115C-105.20 (b) In order to support local boards of education and schools in the implementation of this Program, the State Board of Education shall adopt guidelines, including guidelines to: […] (3) Identify low-performing schools. GS115C-105.21 (a) Local school administrative units shall participate in the School-Based Management and Accountability Program.

		GS115C-105.28. The School-Based Management and Accountability program shall (i) focus on student performance in the basics of reading, mathematics, and communications skills in elementary and middle schools, (ii) focus on student performance in courses required for graduation and on other measures required by the State Board in the high schools, and (iii) holds schools accountable for the educational growth of their students. GS115C-105.30. (a) The State Board of Education shall design and implement a procedure to identify low-performing schools on an annual basis. Low-performing schools are those in which there is a failure to meet the minimum growth standards, as defined by the State Board, and a majority of students are performing below grade level.
Texas	Statute effective Date June 30, 1984; Administrative rules effective date November 28, 1994;	**Texas Education Code Sec 21.753. Accreditation standards.** (a) The State Board of Education shall establish standards which a school district must satisfy to be accredited and shall adopt an accreditation process in accordance with this section. (b) The accreditation standards must include consideration of: (1) goals and objectives of the district; (2) compliance with statutory requirements and requirements imposed by rule of the State Board of Education under statutory authority; (3) the quality of learning of each of the district's campuses based on indicators such as scores on achievement tests; (4) the quality of the district's appraisal of teacher performance and of administrator performance; (5) the effectiveness of district principals as instructional leaders; (6) the fulfillment of curriculum requests; (7) the effectiveness of the district's programs in special education and for special populations; (8) the correlation between student grades and performance on standardized tests; (9) the quality of teacher in-service training; (10) paperwork reduction efforts; and (11) training received by board members. **19 TAC §61.1021** (b) The intent of the School Report Card (SRC) is to inform each student's parents or guardians about the school's performance and characteristics. Where possible, the SRC will present the school information in relation to the district, the state, and a comparable group of schools. The SRC will present the student, staff, financial, and performance information required by statute, as well as any explanations and additional information deemed appropriate to the intent of the report. (c) The SRC must be disseminated within six weeks after it is received from TEA (d) The campus administration may provide the SRC in the same manner it would normally transmit official communications to parents and guardians, such as: including the SRC in a weekly folder sent home with each student, mailing it to the student's residence, providing it at a teacher-parent conference, or enclosing it with the student report card. **Texas Education Code Sec 35.043. Campus report card.** (a) Each school year, the Central Education Agency shall prepare and distribute to each school district a report card for each campus. The campus report cards must be based on the most current data available. Campus performance must be compared to previous campus and district performance, current district performance, state established standards, and comparable campus group performance. (b) The report card shall include the following information where applicable: (1) student performance on state adopted assessment instruments; (2) attendance; (3) dropout rate; (4) student performance on college admissions tests; (5) student/teacher ratios; and (6) administrative and instructional costs per student. The commissioner of education shall adopt rules for requiring dissemination of campus report cards annually to the parent of or person standing in parental relation to each student at the campus. On written request, the school district shall provide a copy of a campus report card to any other party.

References

ABC's Plan (State Board of Ed). North Carolina State Law 1995-716. Retrieved from https://www2.ncleg.net/BillLookup/1995/S1139

Accreditation Standards. Texas Education Code Sec 21.753. 1994. Retrieved from https://lrl.texas.gov/legis/BillSearch/text.cfm?legSession=68-2&billtypeDetail=HB&billNumberDetail=72&billSuffixDetail=&startRow=1&IDlist=&unClicklist=&number=100

Brooks, S. (2000). *How states can hold schools accountable: The strong schools model of standards-based reform*. Seattle, WA: University of Washington Center on Reinventing Public Education.

Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. American Educational Research Journal, 42, 305–328.

Charbonneau, E., & Van Ryzin, G.G. (2012). Performance measures and parental satisfaction with New York City Schools. *American Review of Public Administration, 42*(1), 54-65.

Clotfelter, C., Ladd, H.F., Vigdor, J.L., & Diaz, R.A. (2004). Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management, 23*(2), 251-271.

CORE Districts. (2017, June 9). *Disproportionality in special education.* Retrieved from http://coredistricts.org/wp-content/uploads/2017/07/Disproportionality-SPED_6.9.17.docx

Cummins, J. (1980). Psychological assessment of immigrant children: Logic or institution? Journal of Multilingual & Multicultural Development, 1, 97–111.

Dee, T.S., Jacob, B., & Schwartz, N.L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis, 35*(2), 252-279.

Education Commission of the States [ECS]. (2017). *ESSA requirements for non-academic indicators.* Retrieved from http://www.ascd.org/ASCD/pdf/siteASCD/policy/ESSA-Accountability-FAQ_May112016.pdf

Estrada, P., & Wang, H. (2013, May). Reclassifying and not reclassifying English learners to fluent English proficient, Year 1 findings: Factors impeding and facilitating reclassification and access to the core. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Every Student Succeeds Act of 2015. Pub.L. 115-141

Feng, L., Figlio, D.N., & Sass, T. (2010). *School accountability and teacher mobility.* (Working Paper 47). Washington, D.C.: CALDER, The Urban Institute.

Flores, S. M., & Drake, T. A. (2014). Does English Language Learner (ELL) identification predict college remediation designation? A comparison by race and ethnicity, and ELL waiver status. The Review of Higher Education, 38, 1–36.

Gagnon, D.J., & Schneider, J. (2017). Holistic school quality measurements and the future of accountability: Pilot-test results. *Educational Policy.* DOI: 10.1177/0895904817736631

Gehlbach, H., & Hough, H.J. (2018). *Measuring social emotional learning through student surveys in the CORE district: A pragmatic approach to validity and reliability.* Stanford, CA: Policy Analysis for California Education (PACE). Retrieved from https://edpolicyinca.org/publications/sel-validity

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

Harklau, L. (2002). ESL versus mainstream classes: Contrasting L2 learning environments. In V. Zamel & R. Spack (Eds.), Enriching ESOL pedagogy: Readings and activities for engagement, reflection, and inquiry (pp. 127–158). Mahwah, NJ: Lawrence Erlbaum.

Hutt, E.L. (2018). Measuring missed school: The historical precedents for the measurement and use of attendance records to evaluate schools. *Journal of Education for Students Placed at Risk (JESPAR), 23*, 1-2, 508.

Improving America's Schools Act of 1994. Pub.L. 103-382.

Jacobsen, R., Snyder, J.W., & Saultz, A. (2014). Informing or shaping public opinion? The influence of school accountability data format on public perceptions of school quality. *American Journal of Education, 121*(1), 1-27.

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

Kanno, Y., & Kangas, S. (2014). "I'm not going to be, like, for the AP": English language learners' limited access to advanced college-preparatory courses in high school. American Educational Research Journal, 51, 848–878.

Loeb, S., Christian, M., Hough, H., Meyer, R., Rice, A., & West, M. (2018). *School effects on*

social-emotional learning: Findings from the first large-scale panel survey of students. Stanford, CA: Policy Analysis for California Education (PACE). Retrieved from www.edpolicyinca.org/publications/sel-school-effects

Marsh, J. A., McKibben, S., Hough, H. J., Hall, M., Allbright, T., Matewos, A., & Siqueira, C. (2018). *Enacting social-emotional learning: Practices and supports employed in CORE districts and schools*. Stanford, CA: Policy Analysis for California Education (PACE). Retrieved from http://www.edpolicyinca.org/publications/sel-practices

Martinez-Garcia, C., LaPrairie, K.N., & Slate, J.R. (2011). Accountability ratings of elementary schools: Student demographics matter. *Current Issues in Education, 14*(1). Retrieved from https://cie.asu.edu/ojs/index.php/cieatasu/article/download/685/141/

Massachusetts School and District Accountability System. 603 CMR. 200, Regulations on Under-Performing Schools and School Districts, As Amended. (1999). Retrieved from http://www.doe.mass.edu/lawsregs/accountibility.pdf.

Meyer, R., Wang, C., & Rice, A. (2018). *Measuring Students' Social-Emotional Learning Among California's CORE Districts: An IRT Modeling Approach*. Stanford, CA: Policy Analysis for California Education (PACE). Retrieved from http://www.edpolicyinca.org/publications/sel-measurement

Mintrop, H., & Sunderman, G.L. (2009). Predictable failure of federal sanctions-driven accountability for school improvement and why we may retain it anyway. *Educational Researcher, 38*(5), 353-364.

Parrish, T. B., Merickel, A., Perez, M., Linquanti, R., Socia, M., Spain, A., . . . Delancey, D. (2006). Effects of the implementation of Proposition 227 on the education of English learners, K-12: Findings from a five-year evaluation. Washington, DC: American Institutes for Research.

Perie, M., Park, J., & Klau, K. (2007). *Key elements for educational accountability models.* Washington, D.C.: Council of Chief State School Officers Accountability Systems and Reporting State Collaborative.

School Report Cards. 19 TexReg 8979. 1994. Retrieved from https://texashistory.unt.edu/ark:/67531/metapth176940/m1/17/?q=8979

Sherblom, S.A., Marshall, J.C., & Sherlbom, J.C. (2006). The relationship between school climate and math and reading achievement. *Journal of Research in Character Education, 4*(1-2), 19-31.

Silver, D., Saunders, M., & Zarate, E. (2008). What factors predict high school graduation in the Los Angeles Unified School District? Santa Barbara: California Dropout Project,

University of California.

U.S. Department of Education. (2011). *Letter from the Education Secretary, September 23, 2011*. Retrieved from http://www2.ed.gov/policy/gen/guid/secletter/110923.html

U.S. Department of Education. (2012). *ESEA Flexibility.* Retrieved from https://www2.ed.gov/policy/eseaflex/approved-requests/flexrequest.doc

West, M. R., Buckley, K., Krachman, S. B., & Bookman, N. (2017). Development and implementation of student social-emotional surveys in the CORE districts. *Journal of Applied Developmental Psychology.*

West, M. R., Pier, L., Fricke, H., Hough, H., Loeb, S., Meyer, R.H., & Rice, A.B.. (2018). *Trends in student social-emotional learning: Evidence from the CORE Districts.* Stanford, CA: Policy Analysis for California Education.