Redesigning Systems of School Accountability:
A Multiple Measures Approach to
Accountability and Support

Soung Bae
Stanford Center for Opportunity Policy in Education
United States

Abstract: The challenges facing our children in the 21st century are rapidly changing. As a result, schools bear a greater responsibility to prepare students for college, career, and life and must be held accountable for more than just testing and reporting on a narrow set of outcomes aimed at minimum levels of competency. Thus, scholars, educators, and reform advocates are calling for a more meaningful next phase of school accountability, one that promotes continuous support and improvement rather than mere compliance and efforts to avoid punishment (Center for American Progress & CCSSO, 2014; Darling-Hammond, Wilhoit, & Pittenger, 2014). This paper reviews state and district level accountability systems that incorporate a multiple measures approach to accountability and highlights the following features that represent redesigned systems of accountability: 1) broader set of outcome measures, 2) mix of state and local indicators, 3) measures of opportunities to learn, 4) data dashboards, and 5) School Quality Reviews. The paper concludes with guidance for policymakers and practitioners on ways to support the development and implementation of a multiple measures system of accountability so that school accountability

becomes synonymous with responsibility for deeper learning and support for continuous improvement.
Keywords: school accountability; multiple measures; continuous improvement

Reestructuración de los sistemas de responsabilidad escolar: Un enfoque de medidas múltiples para la rendición de cuentas y el apoyo

Resumen: Debido a los desafíos que enfrentan nuestros niños en el siglo 21, las escuelas tienen una gran responsabilidad de preparar a los estudiantes para la universidad, la carrera profesional y la vida y deben rendir cuentas por algo más que las pruebas académicas y los reportes sobre una lista de resultados limitados a niveles mínimos de competencia. Por lo tanto, académicos, educadores y defensores de la reforma piden que la próxima fase de responsabilidad escolar sea más significativa, una que promueva el apoyo y el mejoramiento continuo en lugar de la conformidad y los esfuerzos para evitar el castigo (Center for American Progress & CCSSO, 2014; Darling- Hammond, Wilhoit y Pittenger, 2014). Este documento revisa los sistemas de responsabilidad a nivel estatal y del distrito que incorporan un enfoque de medidas múltiples para la rendición de cuentas y destaca las siguientes características que representan sistemas rediseñados de responsabilidad: 1) una lista más amplia de medidas de resultado, 2) combinación de indicadores estatales y locales, 3) medidas de oportunidades para aprender, 4) tableros de datos, y 5) Revisiones de Calidad Escolar. El documento concluye con una guía para los formuladores de políticas y profesionales sobre formas de apoyar el desarrollo y la implementación de un sistema de responsabilidad múltiple de medidas múltiples para que la responsabilidad escolar sea sinónimo con responsabilidad para un aprendizaje más profundo y apoyo para el proceso de mejora continua.
Palabras clave: responsabilidad escolar; medidas múltiples; la mejora continua

Reestruturação dos sistemas de responsabilização escolar: Uma abordagem multifacetada para a responsabilização e o apoio

Resumo: Devido aos desafios que nossos filhos enfrentam no século 21, as escolas têm uma grande responsabilidade para preparar os alunos para a faculdade, a carreira e a vida e devem ser responsabilizados por mais do que testes e relatórios acadêmicos. em uma lista de resultados limitados a níveis mínimos de competência. Portanto, os acadêmicos, os educadores e os defensores das reformas exigem que a próxima fase da responsabilidade escolar seja mais significativa, que promova o apoio e a melhoria contínua, em vez de conformidade e esforços para evitar a punição (Center for American Progress & CCSSO, 2014; Darling-Hammond, Wilhoit e Pittenger, 2014). Este documento analisa os sistemas de prestação de contas nos níveis estaduais e distritais que incorporam uma abordagem multi-medidas para a responsabilização e destaca as seguintes características que representam sistemas de responsabilidade redesenhados: 1) uma lista mais ampla de medidas de resultado, 2) combinação de indicadores estaduais e locais, 3) medidas de oportunidades para aprender, 4) tabelas de dados e 5) Revisões de qualidade escolar. O documento conclui com um guia para formuladores de políticas e profissionais sobre formas de apoiar o desenvolvimento e implementação de um sistema multi-ação de responsabilidade múltipla para que a responsabilidade escolar seja sinônimo de responsabilidade pela aprendizagem e suporte mais profundo para o processo de melhoria contínua.
Palavras-chave: responsabilidade escolar; medidas múltiplas; melhoria contínua

Introduction

The challenges facing our children in the 21st century are rapidly changing amidst demographic shifts and technological advances. As a result, schools bear a greater responsibility to prepare students for college, career, and life. To best meet that challenge, schools and districts must be accountable for more than just testing and reporting on a narrow set of outcomes aimed at minimum levels of competency. Students must have engaging, high-quality learning opportunities that help them acquire 21st-century content and skills, including deep content learning, critical thinking and problem-solving, communication, and collaboration abilities—knowledge and skills our students, our democracy, and our economy will need to thrive.

Thus, scholars, educators, and reform advocates are calling for a more meaningful next phase of school accountability, one that promotes continuous support and improvement rather than mere compliance and efforts to avoid punishment (Center for American Progress and the Council of Chief State School Officers, 2014; Darling-Hammond, Wilhoit, & Pittenger, 2014). Specifically, a new accountability system should be grounded in the notion that schools be reconfigured as learning organizations that are committed to continuous improvement and supportive of experimentation, ongoing evaluation, and self-reflection (Darling-Hammond & Plank, 2015). Examining the next phase of school accountability is particularly relevant given the passage of the Every Student Succeeds Act (ESSA) in 2015. Under ESSA, states are challenged to build new systems of accountability that "highlight and measure the things that matter most for student success and provide the most useful data for school improvement" (Darling-Hammond et al., 2016, p. 4). ESSA provides greater flexibility to states to design state accountability systems that reflect ambitious academic standards, use a variety of indicators to measure college- and career-ready outcomes for all students, and can direct resources and support to struggling students and schools.

To meet the challenge of ESSA and address the strengths, interests, and needs of our children and our communities, a new system of accountability should include the use of multiple measures that provide a holistic view of a student's learning and progress and goes beyond standardized test scores of English language arts (ELA) and math. These include measures of student engagement, social-emotional competency, and citizenship; the depth and breadth of meaningful learning opportunities for all students evidenced as rich curriculum opportunities, graduation rates, indicators of career readiness, and measures of successful transitions to post-secondary institutions; and the school's organizational strengths and resources such as data from school climate surveys, attendance rates, and teacher qualifications and distribution (Darling-Hammond, Wilhoit, & Pittenger, 2014; Mellor & Griffith, 2015).

Next phase accountability systems that employ a multiple measures approach to assessing student progress necessitate the use of data dashboards. Dashboards are data systems that track and monitor system and student progress and focus on continuous improvement by making the results of each metric visible and, therefore, more actionable (Rothman, 2015). With data dashboards, schools and districts can prioritize and pay attention to the most context-relevant indicators within the school accountability system to promote student learning and continuous system improvement.

Finally, the move from accountability systems with a compliance orientation to one centered on continuous improvement and learning requires the development and implementation of diagnostic review processes to better understand the quality of teaching and learning within schools and the capacity of systems to support them. These diagnostic reviews or School Quality Reviews (SQRs) provide contextual, qualitative information about the school that complements the quantitative information provided through the data dashboard. The SQR process engages experts and peers in school visits with the purpose of identifying the school's areas of strengths and

weaknesses to support ongoing improvement efforts. The results of the SQR often inform the development of school improvement plans and assist stakeholders in targeting resources to provide learning supports and build local capacity.

The goal of this paper is to review state- and district-level accountability systems that incorporate the concept of multiple measures and examine their potential to support deeper learning and continuous improvement. Specifically, the paper aims to highlight the following features that represent innovative developments in new systems of accountability that are currently being implemented in the United States and Canada:

- Broader set of outcome measures
- Mix of state and local indicators
- Measures of opportunities to learn
- Data dashboards
- School Quality Reviews

For each feature, an illustrative example will be presented to show how selected state and local education agencies incorporate the feature into their redesigned systems of accountability. These illustrative examples aim to provide guidance to education leaders and policymakers to take full advantage of the possibilities that ESSA provides as they redesign their next phase accountability systems.

The paper begins with a discussion of the theory of school-based accountability and summarizes the research literature on the effects of school-based accountability on students and teachers. Next, key features of next phase accountability systems are described and illustrated with examples of state and local education agencies that have incorporated these features into their new systems of accountability. The paper concludes with guidance to policymakers and practitioners on enabling conditions that support the development and implementation of next phase accountability systems focused on deeper learning and continuous improvement.

Theory of School-Based Accountability

School-based accountability is the process of evaluating school performance based on student performance measures and holding educators and school officials responsible for results. In the United States, school-based accountability came to the fore in the mid-1990s on the heels of the standards movement, which sought to identify clear, challenging performance standards for all students and to align curriculum to the standards as well as assessments (Figlio & Loeb, 2011). Contextually, the standards movement was prompted in response to the *A Nation at Risk* report that lamented the "rising tide of mediocrity" of educational performance in our nation's schools and urged the adoption of rigorous learning standards (National Commission on Excellence in Education, 1983, p. 7). The goal of standards-based reform was to identify the schools that were successful in helping students meet those high standards and to encourage schools that were less successful to improve student outcomes.

School-based accountability, as an outgrowth of the standards movement, intensified the stakes for improving student outcomes through the use of rewards and sanctions. Economic theory provides a rationale for its application to education. For example, the principal-agent problem (Holmstrom & Costa, 1986; Milgrom & Roberts, 1988) proposes that educators (agents) might not act in accordance with the interests of stakeholders (e.g., parents, community members, and policymakers – the principal) unless motivated to do so with incentives (both positive and negative) created through accountability systems. Incentive theory proposes that accountability systems will

motivate educators to work harder, cause parents to become more involved, and prompt administrators to implement more effective leadership to increase student achievement. In addition, the availability of independent, quality information regarding how well schools and districts are performing is thought to help educational stakeholders make informed decisions about their educational options (Figlio & Loeb, 2011; Rothstein, Jacobsen, & Wilder, 2008). Therefore, school-based accountability operates on the notion that incentives and public pressure from publicly reported information will result in improved student outcomes (Supovitz, 2009).

At the turn of the century, school-based accountability became the centerpiece of federal education policy with the No Child Left Behind (NCLB) Act of 2001 (Coburn, Hill, & Spillane, 2016; Figlio & Loeb, 2011). Accountability systems of the NCLB era relied on a narrow set of indicators—achievement scores on annual standardized exams in ELA and math—to evaluate school performance. This was motivated by the fact that NCLB required states to annually test all students in grades 3 to 8, and once in grades 10 through 12, in ELA and math, as evidence of student and school progress, with the goal of all students reaching proficiency by 2014. Standardized assessments do provide useful information regarding how students and schools are performing, especially by providing comparable scores that illuminate achievement gaps. Moreover, studies have shown that high-stakes testing has contributed to slight increases in student achievement (Au, 2007; Carnoy & Loeb, 2002; Jacob, 2002; Mintrop & Sunderman, 2009). However, research has also shown that for states that have posted large gains in student performance on their high-stakes state assessments, those gains have not translated into comparably large gains on the National Assessment of Educational Progress (Jacob, 2007; Klein, Hamilton, McCaffrey, & Stecher, 2000; Linn, 2000) or other low-stakes state assessments (Figlio & Rouse, 2006; Jacob, 2005).

The NCLB era, more importantly, has shown that standardized assessments, alone, should not make determinations of student progress or school quality. Research has shown that a myopic focus on student test scores in ELA and math did not provide equitable and meaningful learning opportunities for all students (Diamond & Spillane, 2004; Heilig & Darling-Hammond, 2008; Plank & Condliffe, 2013) and resulted in unintended consequences, such as educators narrowing the curriculum to focus on tested subjects (Darling-Hammond & Adamson, 2014; Hamilton, Berends, & Stecher, 2005; Linn, 2000), concentrating resources on the "bubble kids" or students on the cusp of passing the high-stakes test (Booher-Jennings, 2005), and teaching test-taking skills divorced from the content being tested (Jacob, 2005; Shepard, 1990).

Lessons learned from the NCLB era suggest that school-based accountability systems would be improved by focusing attention on a broader set of behavioral and attainment outcomes and balancing accountability with support for continuous improvement through the inclusion of input and process measures. Input measures refer to the social and fiscal resources available to the school, such as the level of funding, class size, teacher qualifications, course offerings, or the conditions of school facilities. Process measures refer to the activities that take place during the school day and in the learning environment, such as the quality of instruction, teacher and student interactions, school climate, and school safety. These additional measures can help stakeholders make valid inferences about school quality (what practices are effective or ineffective) and resource equity (what resources are available to the school) to support school improvement efforts (Darling-Hammond, Bae, Cook-Harvey, Lam, Mercer, et al., 2016; Darling-Hammond, Wilhoit, & Pittenger, 2014; Schwartz, Hamilton, Stecher, & Steele, 2011). That is, next phase accountability systems necessitate the use of multiple measures of student and school success. In fact, the joint statement on standards for appropriate test use by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, advises that high-stakes decisions about a student's continued education should not be based on "test scores as the sole indicator to characterize an individual's functioning, competence, attitude, and/or predispositions.

Instead, multiple sources of information should be used." (Standards for Educational and Psychological Testing, 2014, p.71). Similarly, the National Research Council Board on Testing and Assessment's study on appropriate and nondiscriminatory use of educational tests recommended that because a test score is never an exact measure of students' knowledge or skills, a single test score should not determine high-stakes educational decisions, and other relevant information must be taken into account (Heubert & Hauser, 1999). Notably, ESSA addresses the shortcomings of NCLB-era accountability systems by requiring that multiple measures of student and school success be incorporated into state accountability systems. The move to a multiple measures approach to school-based accountability broadens the focus of accountability to include the conditions and opportunities that can support student and school success as well as outcomes. As such, stakeholders are provided with more information to gauge progress on learning goals, to inform improvement efforts, and to create the positive conditions for learning.

Key Features of Next Phase Accountability Systems

Broader Set of Outcome Measures

New accountability systems of the post-NCLB era define student and school success along a broader array of outcomes that promote deeper learning and the attainment of relevant skills necessary for success in postsecondary education, the workplace, and life. Raising student achievement in two subject areas is no longer the sole focus of school-based accountability. Instead, this accountability expands educators' attention and energies on

- **developing student competencies beyond basic skills** to mastering 21st-century skills, such as creativity, problem-solving, critical thinking, and communication skills, and demonstrating mastery through the use of performance-based assessments and graduation portfolios;
- **promoting the development of noncognitive skills**, such as managing one's behaviors and emotions, empathy, self-efficacy, and a growth mindset; and
- **cultivating college *and* career readiness** in all students.

This broader array of outcomes is supported by ESSA, as the legislation specifies that, at a minimum, state accountability systems must include indicators of student achievement in ELA, math, and science; another valid and reliable statewide academic indicator for elementary and middle schools; graduation rates for high schools; English language proficiency rates; and at least one other measure of school quality or student success. ESSA's approach to state accountability systems encourages states to define student learning in a much more expansive way; states could consider incorporating this broader set of outcomes as additional measures of school quality or student success under the law. The following examples illustrate the inclusion of a broader set of student outcomes in redesigned school accountability systems in California and South Carolina.

Developing Student Competencies Beyond Basic Skills. In 2013, California's legislature enacted two key policies that are redefining K–12 education in the state: the Local Control Funding Formula (LCFF) and the Local Control and Accountability Plan (LCAP). The LCFF changes the way the state allocates money to school districts and the way the state supports underperforming districts (Taylor, 2013). Funding is now based primarily on student needs (with weights attached to funding allocations for poverty, English learner status, and foster child status). The LCAP accompanies the LCFF and outlines a process for establishing a school district's annual goals and tracks the district's progress toward those goals. The intent of the LCAP is to establish greater

transparency and accountability for school districts while also providing more flexibility to the districts to meet their goals and incentivizing schools to pay attention to important outcomes that are aligned to the state's eight priority areas (see Figure 1).

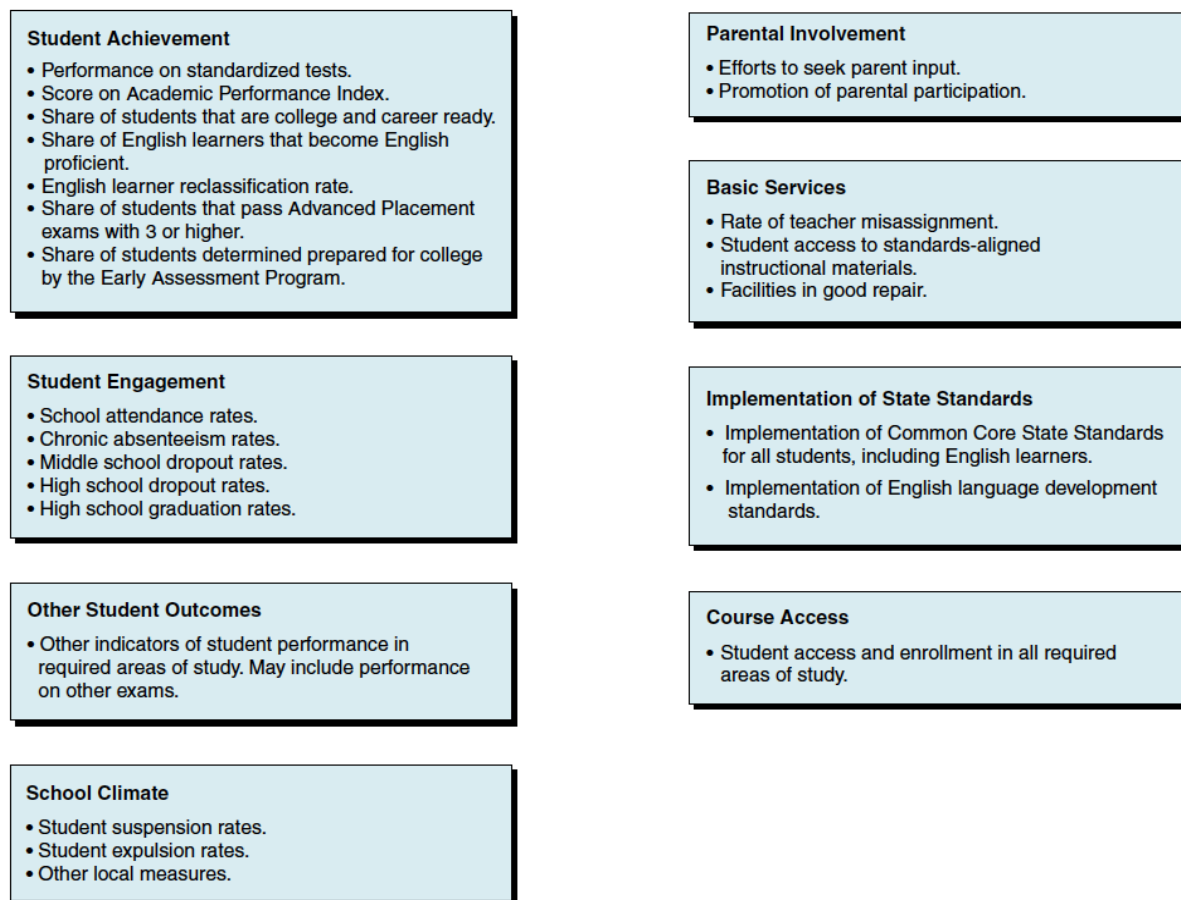## The Eight State Priority Areas and Associated Metrics

Student Achievement
- Performance on standardized tests.
- Score on Academic Performance Index.
- Share of students that are college and career ready.
- Share of English learners that become English proficient.
- English learner reclassification rate.
- Share of students that pass Advanced Placement exams with 3 or higher.
- Share of students determined prepared for college by the Early Assessment Program.

Student Engagement
- School attendance rates.
- Chronic absenteeism rates.
- Middle school dropout rates.
- High school dropout rates.
- High school graduation rates.

Other Student Outcomes
- Other indicators of student performance in required areas of study. May include performance on other exams.

School Climate
- Student suspension rates.
- Student expulsion rates.
- Other local measures.

Parental Involvement
- Efforts to seek parent input.
- Promotion of parental participation.

Basic Services
- Rate of teacher misassignment.
- Student access to standards-aligned instructional materials.
- Facilities in good repair.

Implementation of State Standards
- Implementation of Common Core State Standards for all students, including English learners.
- Implementation of English language development standards.

Course Access
- Student access and enrollment in all required areas of study.

Figure 1. California's School Accountability Priority Areas
Source: Taylor, M. (2015, January). *Review of school districts' 2014–2015 Local Control and Accountability Plans* (p. 7). Sacramento, CA: Legislative Analyst's Office.

The state priorities include student achievement, student engagement, school climate, implementation of state standards including the Common Core State Standards, parental involvement, provision of basic services, curriculum access, and other student outcomes. The legislation also provided a list of potential metrics that districts could use to track their progress such as using suspension and expulsion rates and student surveys to measure school climate or dropout rates and attendance rates as measures of student engagement. Thus, California's redesigned school accountability system promotes the use of a more robust set of indicators to measure student achievement and school performance beyond the narrow focus on improving students' test scores in two academic subjects.

These indicators will be used to identify strengths within the system. For example, the state may choose to highlight successful policies and practices that schools and districts throughout the

state have shown support student learning and development. The indicators will also be used to identify areas needing improvement within the system and determine which schools and districts need support and intervention. As an example, ESSA regulations require that school accountability systems identify the bottom 5% of schools so that they can receive more targeted support. In California, the broader set of indicators will allow local and state education leaders to identify struggling schools and districts along a more holistic view of student success that aligns with the state's priorities and goals and provide them with targeted support and resources.

Promoting the Development of Noncognitive Skills. In 2013, the California Office to Reform Education (CORE)[1] districts submitted a waiver application to the United States Department of Education requesting flexibility with respect to certain requirements under the Elementary and Secondary Education Act (California Office to Reform Education, 2014). The CORE waiver centered on a new and differentiated accountability model called the School Quality Improvement System (California Office to Reform Education, 2013) (see Figure 2).
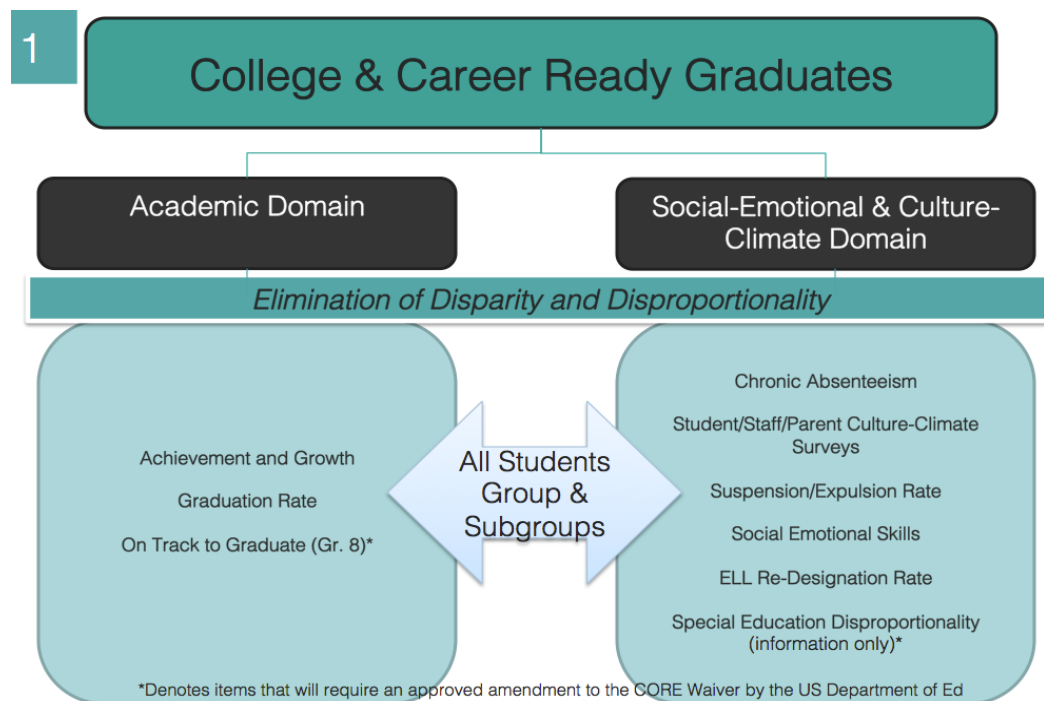
Figure 2. California Office to Reform Education (CORE) School Quality Improvement System Source: CORE's Accountability System (N. Bookman, personal communication, January 27, 2015).

The School Quality Improvement System emphasizes a multiple measures approach to accountability and includes the following indicators: academic achievement and growth, graduation rates, high school readiness of eighth-graders under the academic domain, as well as social-emotional and culture-climate factors such as chronic absenteeism, social-emotional skills, suspension/expulsion rates, student/staff/parent surveys of school culture and climate, and English

---

[1] CORE represents a partnership between eight unified school districts throughout California—Fresno, Long Beach, Los Angeles, Oakland, Sacramento City, San Francisco, Santa Ana, and Sanger—to share and learn together and implement education reforms. The Elementary and Secondary Education Act Flexibility waiver was approved and renewed in 2015.

Language Learner redesignation rates. The academic domain is weighted to account for 60% of a school's accountability score and the social-emotional and culture-climate factors account for the remaining 40%. The data from the School Quality Improvement System is designed to track students' progress related to those factors and to signal the types of interventions and continuous improvement strategies that could be implemented in the schools.

The CORE districts were one of the first to experiment with the inclusion of social-emotional learning skills as an important student outcome to track in school accountability systems (Klein, 2015). Drawing on social science research that highlights the role of noncognitive skills in student achievement in school (Duckworth & Carlson, 2013; Duckworth & Seligman, 2005) as well as in the labor market (Heckman, Stixrud, & Urzua, 2006), the CORE districts are focusing on four social-emotional learning competencies: growth mindset, self-efficacy, self-management, and social awareness.

These important skills are measured via a student self-report survey that assesses students on a series of behaviors as well as beliefs related to social-emotional learning. For example, the students are asked how often they come to class prepared, how much they care about other people's feelings, and whether they believe they are incapable of learning certain things. The inclusion of these social-emotional learning indicators incentivizes schools to prioritize and support the development of noncognitive skills, which helps expand the definition of student success and preparedness beyond test scores. However, Duckworth and Yeager (2015) caution educators on using currently available measures of personal qualities (e.g., self-control, grit, emotional intelligence) for accountability purposes. Because self-reported questionnaires rely on the subjective judgments of students or teachers, they are susceptible to reference bias, which refers to the frame of reference that is invoked to arrive at a judgment about one's relative standing, but differs systematically across respondents. Thus, differences in self-reports may reflect variation in normative expectations rather than true differences in skills, and this can render comparisons between schools ineffective. For example, West and his colleagues (2016) found that among eighth-grade students in their study, conscientiousness, self-control, and grit were positively correlated with test score gains between fourth grade and eighth grade at the student level. At the school level, however, those correlations disappeared, and students attending charter schools known for their high expectations for student success scored lower on the personal qualities scales than did students attending traditional public schools. The researchers concluded that the paradoxical results were influenced by reference bias. Moreover, Duckworth and Yeager (2015) warn that students or teachers may inflate their answers to look good, which can result in "superficial parroting" of personal qualities rather than actual, deep changes in beliefs or perceptions. Therefore, state leaders must be mindful of how these evolving measures of personal qualities are used and for what purposes, being hyper-vigilant about the limitations of the measures. Until the currently available self-reported measures of social-emotional learning become more reliable, state leaders could consider employing these measures for diagnostic and reporting purposes rather than for high-stakes purposes.

Career Readiness Indicators. In 2005, the South Carolina Legislature passed the Education and Economic Development Act, which emphasizes career exploration throughout PK–12, including an individual graduation plan starting in grade 8 and the selection of a "major" from among career clusters by grade 10. Parents, students, and counselors meet annually to update the individual graduation plan. Part of the conference concerns the inclusion of "experience-based, career-oriented learning experiences including, but not limited to, internships, apprenticeships, mentoring, co-op education, and service learning" (SC EC, 2005, section 59-59-140). As of 2016, South Carolina was the only state that publicly reported the following career readiness indicators in conjunction with traditional college readiness indicators (e.g., the number of students enrolled in

Advanced Placement [AP]/International Baccalaureate [IB] programs and are successful in AP/IB programs) on their school report cards (Achieve and Advance CTE, 2016):

- The percentage of students who participate in approved work-based learning experiences
- The number of students enrolled in career technology courses
- The percentage of students attending career technology centers or comprehensive high schools who participate in cocurricular organizations[2] (South Carolina Education Oversight Committee, 2015)

Beginning in 2015, the state also included a college and career outcome in its state report card: the percentage of graduating students who enroll in a 2- or 4-year college or technical college to pursue a degree, certificate, or diploma. South Carolina's inclusion of career readiness indicators in its school accountability system demonstrates its commitment to prioritizing and supporting multiple pathways to postsecondary success for all students.

Mix of State and Locally Determined Indicators

The metrics and indicators that are chosen for inclusion in school accountability systems greatly shape how schools and districts prioritize important education goals for students. Meaningful indicators incentivize schools to pay attention to critical outcomes and can serve as a powerful lever to encourage schools to prepare students for a fuller range of postsecondary outcomes. Under NCLB, the law's narrow definition of school accountability created accountability systems with a myopic focus on student performance on standardized tests that, in the end, did not serve students well. In fact, scholars suggest that many existing state accountability systems lack flexibility and are not responsive to local needs and contexts when state accountability systems do not allow districts to include their own locally defined indicators (Rothman, 2015).

Take, for example, a district that is struggling to increase high school graduation rates, combat student disengagement, and prepare all its students for both college and career. As a response to those needs, the district may develop rigorous career technical education (CTE) pathways as a strategy to engage all students in the learning process and ensure they graduate with a high school diploma. However, if the school accountability system only permitted the inclusion of a limited number of state-sanctioned indicators, that district could be disincentivized to continue the difficult work of developing rigorous CTE pathways in favor of focusing solely on the required indicators. Moreover, the opportunity to incentivize districts to hold themselves accountable for education goals that are responsive to their local contexts would be lost. Thus, new school accountability systems are addressing this shortcoming by structuring ways for districts to include a mixture of state- and locally-developed indicators.

Next phase school accountability systems encourage districts to pursue their own priorities while maintaining state-level measures to ensure equity across districts throughout the state. In California, for example, the LCAP requires districts to set annual goals and targets that are aligned to the state's eight priority areas (see Figure 1). In addition, districts are encouraged to include annual goals that reflect local priorities and needs. For example, under the state's priority area of school

---

[2] Cocurricular organizations are school-related clubs and organizations such as DECA (a marketing, finance, hospitality, and management student organization), VICA (Vocational Industrial Clubs of America), FFA (Future Farmers of America), HOSA (student organization whose mission is to promote career opportunities in health care), and FBLA (Future Business Leaders of America).

climate, districts may include locally developed measures of school safety to address the school community's concerns around bullying and organizational health.

To support the development of local flexibility and innovation, California's accountability system will include three different types of indicators: (a) state-required, state-reported; (b) state-supported and locally reported; and (c) locally generated and reported (California Department of Education, 2016). State-required indicators are those that will be used for both state and federal accountability. The state-reported indicators are those that will be vetted and reported by the state and will be used to provide a more holistic view of performance, equity, and improvement. The state-supported, locally reported indicators are those that will be state developed and calibrated and made available for voluntary use and reporting by districts such as classroom-embedded, authentic performance assessments and assessments of social-emotional learning. Finally, the locally generated and reported indicators are those that are locally identified and vetted and used for LCAP design, implementation, and evaluation (e.g., increasing the hiring and retention rates of teachers of color). It is conceivable to imagine that as districts test and refine their locally developed indicators and grow the evidence base for reliability and validity, those indicators may be migrated into the bin of state-supported, locally reported indicators or even state-reported indicators.

New Hampshire's Performance Assessment for Competency Education (PACE) serves as another example of a redesigned accountability system that uses a mixture of state and locally derived measures. In contrast to accountability systems of the NCLB era that relied on annual statewide assessments in grades 3 through 8, the PACE system establishes a hybrid model that includes (a) the Smarter Balanced summative assessments given once within each grade span (i.e., grades 3–5, 6–8, 9–12) rather than every grade; (b) locally designed performance assessments that will be administered in all subjects and grades; and (c) common performance assessments that will be administered in grades not using Smarter Balanced. The use of both state- and locally-developed assessments allows district officials to better monitor student learning and progress throughout the year and provides district officials with the data to make informed decisions about necessary adjustments and improvements.

Measures of Opportunities to Learn

A major goal of a school accountability system should be to address and ameliorate the unequal educational outcomes that persist in the nation's schools and communities. The focus on multiple and more meaningful outcome measures of student learning and progress (as described above) is a necessary but insufficient condition for achieving greater equity because focusing solely on outcomes, without regard to the inputs that may enable or constrain student success, paints an incomplete picture of how the system is functioning, what it provides (or doesn't), and for whom. Therefore, measures of educational access or opportunities to learn are necessary, in conjunction with student outcome indicators. Together, they provide a comprehensive set of indicators with which all stakeholders can better assess how the educational *system* is serving all students and meaningfully supporting improvement.

Educational outcomes are highly related to and dependent on access to key resources. Research has documented the inequitable distribution of educational resources and access to knowledge that has contributed to the achievement gap (Adelman, 2006; Allensworth, Nomi, Montgomery, & Lee, 2009; Spielhagen, 2006). These resources include access to equitable funding and state-of-the-art equipment and materials for learning, access to highly qualified teachers,[3] access

---

[3] Although NCLB required state accountability systems to certify teachers as being highly qualified (as measured by teachers' years of teaching experience, licensure, and educational degrees), access to key

to a rich curriculum, and access to safe and orderly learning environments. These school characteristics or inputs can be operationalized as opportunities to learn about indicators that incentivize schools, districts, and states to emphasize policies and practices that support high-level learning for all students.

Measures of access to resources. Research has shown that a school's resource level, in itself, does not guarantee high-quality instructional programs (Coleman, Campbell, Hobson, McPartland, Mood, 1966). Rather, what matters is the way in which schools allocate their available resources into staffing, curriculum priorities, and organizational structures that provide students with access to high-quality learning experiences (Elmore & Fuhrman, 1993; Hanushek, 1981; Murnane, 1982). In addition, more recent research provides compelling evidence that increased school spending can improve the educational and lifetime outcomes of children from disadvantaged backgrounds (Jackson, Johnson, & Persico, 2014). Thus, states must pay attention to access to resources indicators including the level of school funding; staffing; availability of up-to-date textbooks, materials, and technology; class size; availability of professional learning experiences for teachers; and the like. These indicators may be assessed through parent, student, and staff surveys on the availability and use of key resources.

In California, the provision of basic services (e.g., student access to standards-aligned materials and well-maintained facilities) is one of the state's eight priority areas that must be addressed through the school accountability system (see Figure 1). In addition, the state's LCAP process is designed to make transparent how school funds will be used to implement specific strategies and practices to realize the state and district's goals. Thus, the inclusion of measures of access to resources in California's redesigned school accountability system signals to stakeholders that inputs matter in creating conditions for learning and ultimately student achievement.

Measures of access to highly qualified teachers. Research on teacher quality has demonstrated that teaching quality matters. In fact, teacher quality has been identified as the most significant school-based factor in student performance (McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2005), and teacher effects on student achievement are cumulative and long-lasting (McCaffrey, Lockwood, Koretz, & Hamilton, 2003). However, research has also shown that minority and poor students are more likely to attend schools staffed with the least qualified teachers (Darling-Hammond, 1990; Oakes, 2004). Therefore, ensuring that all classrooms, regardless of their zip codes, are staffed with highly qualified teachers is an imperative for enabling more successful learning for all students. To track and monitor this school characteristic, the following measures could be considered: teacher certification, concentrations of uncredentialed teachers, out-of-field teacher placements, and years of teaching experience.

In Vermont, the State Board of Education adopted the Education Quality Standards (EQS) in 2014 (Vermont State Board of Education, 2014). These standards were established to ensure that all students had access to high-quality educational programming that is substantially equal throughout the state. The standards focused on assuring that all schools delivered on proficiency-based learning, flexible pathways to graduation, safe school environments, high-quality staffing, and financially efficient practices. Then, to hold schools accountable for the standards, the Vermont Agency of Education developed Education Quality Reviews (EQRs), which are the tools with which the state and the public will measure student learning and school progress (Fowler, 2015).

The EQR consists of two complementary review processes: the Annual Snapshot Review and the Integrated Field Review, used to drive continuous improvement. The Annual Snapshot

---

resources in next phase accountability goes beyond that indicator to include curricula, learning materials, and teaching environments.

gathers quantitative data along the five dimensions of school quality, one of which is high-quality staffing. This dimension will be assessed through the quantification of staff credentials (e.g., percentage of staff working on a full license or a provisional license); staff stability (e.g., rate of staff turnover); staff experience (e.g., average years of experience); professional development (PD) (e.g., percentage of staff work schedules devoted to on-site PD, percentage of grant funds for PD, percentage of all expenditures for PD, percentage of staff participating in PD); staff-community connectedness (e.g., index of staff residential distance to community); and shared leadership (e.g., presence of leadership teams, diversity of membership in leadership teams).

Measures of access to rich curriculum. Ambitious educational outcomes such as preparing all students for success in college and career depend on students' access to a rich curriculum. Given the strong relation between enrollment in advanced mathematics and science courses and student achievement (Gamoran, 1992; Hoffer, 1997; Teitelbaum, 2003), access to a full curriculum matters. Unfortunately, research on secondary schools has shown that schools attended primarily by low-income and minority students offer fewer advanced and college preparatory courses. In addition, minority students are more likely to be assigned to nonacademic and remedial tracks (Oakes, 1985; Oakes & Guiton, 1995). Thus, states could consider collecting data on whether students have access to a full curriculum that includes ELA, mathematics, history, social studies, civics, art, music, world language, as well as physical education and career technical education. Other indicators could include the number of advanced courses (e.g., Advanced Placement, International Baccalaureate, dual enrollment) and career pathways offered, and the percentage of students enrolled in those courses and pathways, disaggregated by student subgroups.

Alberta, Canada's accountability system prioritizes high-quality learning opportunities for students as a meaningful goal for schools. To monitor this goal, students, parents, and teachers complete an annual survey that assesses the percentage of respondents who report that they are satisfied with the opportunity for students to receive a broad program of studies as well as the percentage of respondents who report that they are satisfied with the overall quality of basic education provided in schools. California has identified course access as a priority area and intends to collect data on student access and enrollment in all required areas of study.

Measures of school climate. School climate is "based on patterns of people's experiences of school life and reflects norms, goals, values, interpersonal relationships, teaching and learning practices, and organizational structures… that support people feeling socially, emotionally and physically safe" (Cohen, McCabe, Michelli, & Pickeral, 2009, p. 182). A growing body of correlational studies has demonstrated that a positive school climate is directly related to student achievement at the elementary school level (Cook, Murphy, & Hunt, 2000; Sherblom, Marshall, & Sherblom, 2006; Sterbinksky, Ross, & Redfield, 2006), middle school level (Ma & Klinger, 2000), and high school level (Lee & Bryk, 1989; Stewart, 2008). Monitoring and assessing school climate, therefore, is a critical endeavor for understanding, assessing, and supporting how school climate enables teaching and learning in classrooms.

The National School Climate Center conducted a scan in 2011 that revealed that 24 states have a school climate policy in place (Piscatelli & Lee, 2011). These policies range from addressing aspects of school climate within their school quality standards (e.g., Alaska and Vermont) and integrating school climate goals and assessments in their school improvement standards (e.g., Montana) to developing specialized school climate standards and guidelines (e.g., Ohio and Wisconsin, although both states' standards are voluntary).

Although related to measures of social-emotional learning, school climate differs in that it assesses stakeholders' perceptions of and experiences with the learning environment, while the former assesses students' perceptions of their own learning, attitudes toward healthy social

relationships, and conceptions about their ability to make sound judgments and decisions. School climate is typically measured through surveys of parents, students, and educators, which seek to measure stakeholders' perceptions of a safe and respectful climate, interpersonal relationships, and feelings of school connectedness, as well as satisfaction with how the school structures supports for learning. In addition, some school climate data can be derived from administrative data such as rates of school suspensions or expulsions, student enrollment in extracurricular activities, or character-development programs.

Numerous states, districts, and charter management organizations are collecting school climate data for diagnostic and school improvement purposes (Schwartz, Hamilton, Stecher, & Steele, 2011). For example, the California Department of Education required that all school districts that received federal Title IV (Safe and Drug-Free Schools and Communities Act) funding or a state Tobacco Use Prevention and Education grant to administer the California School Climate, Health, and Learning Survey at least once every 2 years to students in grades 5, 7, 9, and 11.[4] The survey assesses students' school-related attitudes, behaviors, and experiences including school safety, bullying, and feelings of school connectedness. The state of Rhode Island administers a climate survey to all students in grades 4–12, to all K–12 teachers and principals, and to all parents of students in grades K–12. The survey results are intended to illuminate the reasons for the school's academic performance indicators and be used as a component for school accreditation. Furthermore, the Consortium on Chicago School Research has long administered school surveys; since 2006 it has surveyed parents and students to produce indicators of school climate that are reported on the Chicago Public Schools' report cards. The school climate indicators do not influence the district's rating of the school but are provided to parents and students to inform school choice decisions (Schwartz, Hamilton, Stecher, & Steele, 2011).

To date, there are few school systems or consortia of school systems that incorporate school climate into the formal school accountability system. One that does is New York City. In New York City, each year, educators, parents, and students in grades 6–12 complete the New York City School Survey. The survey is designed to capture stakeholders' perceptions of the learning environment at each school, focusing on six core elements that are aligned to the New York City Department of Education's Framework for Great Schools: rigorous instruction, supportive environment, collaborative teachers, effective school leadership, strong family and community ties, and trust (NYCDOE, 2016). The survey results are used in conjunction with other measures (i.e., Quality Review, percentage of students with attendance rates of 90% or higher, and movement of students with disabilities to less restrictive environments) to produce a rating for each core element (NYCDOE, 2015). The ratings are reported in the School Quality Reports, which are the Department's tools that provide information about each school's practices and environment. The reports come in two forms: (a) a School Quality Snapshot geared towards families and community members attempting to provide a concise understanding of the quality of each school and (b) a School Quality Guide, a more detailed report that provides performance information over multiple years to track the school's progress over time.

Data Dashboards

A primary purpose of accountability is to monitor progress on meaningful goals to support continuous improvement. New school accountability systems use data dashboards to provide an accurate and up-to-date understanding of how well a school is progressing and how well a system is progressing in supporting schools (Rothman, 2015). To use an oft-cited analogy, a data dashboard is

---

[4] The federal Title IV funding ended in 2010, and since then the administration of the survey has become voluntary.

like a car dashboard, which provides critical information with which to gauge the car's functioning such as how fast the vehicle is moving, whether the car is in danger of running out of fuel, and if the engine is about to overheat. The dashboard provides indicators on critical measures of what is working and what is not as well as information on the location of the "problem." Thus, data dashboards create transparency, for educators and the local community, another key aspect of accountability (Darling-Hammond, Wilhoit, & Pittenger, 2014).

Data dashboards are rich information systems that facilitate insights into a school's areas of strength as well as areas for improvement. This information is essential for educators, parents, and community members to understand and interpret how well the school, and the system, is performing and to make informed decisions. For example, data dashboards can help schools and districts facilitate ongoing improvement by using the data from the dashboard to prioritize and target limited resources to support intervention and corrective action as well as to celebrate practices that have been proven successful in advancing teaching and learning.

States interested in moving toward the development of data dashboards to organize and report the information needed to guide improvement could look to the data system used in Alberta, Canada.

Alberta's data dashboard. The Alberta Results Report is an online reporting tool that contains data for the province's seven sets of indicators organized around three main goals in its multiple measures accountability system (see Figure 3):

1. **High-Quality Learning Opportunities**:
    a. Safe and Caring Schools – measured by the percentage of teacher, parent, and student agreement that students are safe at school, learning to be caring, learning respect for others, and being treated fairly in school.
    b. Student Learning Opportunities – measured by the percentage of teachers, parents, and students who are satisfied with the opportunity for students to receive a broad program of studies; the percentage of teachers, parents, and students who are satisfied with the overall quality of basic education; the percentage of students aged 14–18 registered in the K–12 system who drop out the following year; and the percentage of students in the grade 10 cohort who complete high school within three years.

2. **Excellence in Learner Outcomes**:
    a. Student Learning Achievement (Grades K–8) – measured by the percentage of students who achieve the acceptable or excellence standard on the Provincial Achievement Test.
    b. Student Learning Achievement (Grades 9–12) – measured by the percentage of students who achieve the acceptable or excellence standard on a diploma exam; the percentage of students in the grade 10 cohort who have taken four or more diploma exams by the end of their third year in high school; and the percentage of grade 12 students who have met the eligibility criteria for a Rutherford Scholarship based on course marks in grades 10, 11, and/or 12.
    c. Preparation for Lifelong Learning, World of Work, Citizenship – measured by the percentage of students in the grade 10 cohort who have entered a postsecondary-level program at an Alberta postsecondary institution or registered in an Alberta apprenticeship program within six years of entering grade 10; the percentage of teachers and parents who agree that students are taught attitudes and behaviors that

will make them successful at work; the percentage of teachers, parents, and students who are satisfied that students model the characteristics of active citizenship; the percentage of teachers and parents who are satisfied that students demonstrate the knowledge, skills, and attitudes necessary for lifelong learning.

3. **Highly Responsive and Responsible Jurisdiction**:
   a. Parental Involvement – measured by the percentage of teachers and parents who are satisfied with parental involvement in decisions about a child's education;
   b. School Improvement – measured by the percentage of teachers, parents, and students who indicate that their school and the schools in their jurisdictions have improved or stayed the same the last three years.

The Results Report details how schools in each province are performing along each metric, which is scored and color-coded to indicate performance levels (i.e., red indicates very low performance, orange indicates low performance, yellow indicates intermediate performance, green indicates high performance, and blue indicates very high performance). In addition, the report provides users with information on the trends associated with each metric. For example, educators, parents, and community members are given the numeric score for how the province schools fared on the safe and caring school metric for that year, along with the scores for the previous year and the previous three-year average.

Because the Results Report is an electronic tool, users can click on the safe and caring schools metric, which then drills down into more detailed information about the measure (e.g., tables and graphs display a breakdown of how each role group, such as teachers, parents, and students, agreed with the statement that the school is safe as well as a breakdown of how parents answered each survey item related to this measure for the last 5 years). All stakeholders use this information to understand how students and schools are progressing, make resource allocations, and strategically target areas in need of improvement. It is noteworthy that the seven sets of indicators do not roll up into a summative score; instead, the Results Report displays each individual measure and performance level. Administrators from Alberta state that this is deliberate so that schools can focus on improvement and prioritize their efforts on specific areas of need that are highlighted when the performance data are disaggregated (Mellor & Griffith, 2015).

Accountability Pillar Overall Summary
Annual Education Results Reports - Oct 2016
Province: Alberta

Measure Category	Measure Category Evaluation	Measure	Alberta			Measure Evaluation		
			Current Result	Prev Year Result	Prev 3 Year Average	Achievement	Improvement	Overall
Safe and Caring Schools	Excellent	Safe and Caring	89.5	89.2	89.1	Very High	Improved Significantly	Excellent
Student Learning Opportunities	Excellent	Program of Studies	81.9	81.3	81.4	Very High	Improved Significantly	Excellent
		Education Quality	90.1	89.5	89.5	Very High	Improved Significantly	Excellent
		Drop Out Rate	3.2	3.5	3.5	High	Improved Significantly	Good
		High School Completion Rate (3 yr)	76.5	76.5	75.5	High	Improved Significantly	Good
Student Learning Achievement (Grades K-9)	Good	PAT: Acceptable	73.6	72.9	73.4	Intermediate	Maintained	Acceptable
		PAT: Excellence	19.4	18.8	18.6	Intermediate	Improved Significantly	Good
Student Learning Achievement (Grades 10-12)	n/a	Diploma: Acceptable	85.0	85.2	85.1	Intermediate	Maintained	Acceptable
		Diploma: Excellence	21.0	21.0	20.5	High	Improved	Good
		Diploma Exam Participation Rate (4+ Exams)	54.6	54.4	53.5	Intermediate	Improved Significantly	Good
		Rutherford Scholarship Eligibility Rate	60.8	n/a	n/a	n/a	n/a	n/a
Preparation for Lifelong Learning, World of Work, Citizenship	Good	Transition Rate (6 yr)	59.4	59.7	59.3	High	Maintained	Good
		Work Preparation	82.6	82.0	81.1	High	Improved Significantly	Good
		Citizenship	83.9	83.5	83.4	Very High	Improved Significantly	Excellent
Parental Involvement	Good	Parental Involvement	80.9	80.7	80.5	High	Improved	Good
Continuous Improvement	Excellent	School Improvement	81.2	79.6	80.0	Very High	Improved Significantly	Excellent

Figure 3. Alberta Results Report
Source: Alberta Education. (2016). October 2016 Overall Provincial Summary. Retrieved July 5, 2017, from
https://education.alberta.ca/media/3273034/province_province-report-accountability-pillar-overall-summary.pdf.

School Quality Review

Redesigned school accountability systems should also include processes for evaluating school quality. It seems ineffective and unfair to hold schools and districts accountable for student learning and performance without knowing whether learners are provided the supports they need to be successful such as qualified teachers, a rigorous and meaningful curriculum, a positive school culture and climate, and strong school leadership. Without school context information, student performance scores are less meaningful and less actionable. SQRs can help generate the needed contextual information with which to better understand the quality of teaching and learning in schools as well as assess the system's support for school improvement.

The SQR is a formal process for assessing teaching and learning in schools. It typically involves a school self-assessment, which is followed by a school visit by formal and peer reviewers who develop qualitative conclusions about the quality of teaching and learning by observing teaching and learning; reviewing student work, teacher assignments, and school goals and documents; and interviewing stakeholders. The primary purpose of the SQR is to assist schools in developing a culture of ongoing review and continuous improvement (Ancess, 1996). SQRs foster a culture of review through two interacting components: (1) an ongoing school self-review and (2) an external review conducted by a team of diverse, external stakeholders who represent a mix of formal and peer reviewers (Ancess, 1996). The self-review encourages schools to engage in continual cycles of data gathering, reflection, examination of effective teaching practices, and critical conversations about how schools are meeting their goals of providing high-quality instruction so that all students achieve to their fullest potential (Ancess, 1996). SQRs provide direct observations of classroom or school activity by external expert reviewers and afford insights into the quality of instruction, curriculum, interpersonal relationships, and the learning environment. In addition, the data garnered from SQRs provide educators and administrators with actionable information to guide school improvement.

In New York City, the Quality Review is used to examine the quality of instruction in schools and to provide the school community an opportunity to reflect on and self-evaluate its progress and improvement efforts (NYCDOE, n.d.-a). The Quality Review focuses on instructional and organizational coherence as keys to improving student learning. The reviews evaluate the school's work as it relates to the following quality indicators:

- Rigorous, engaging and coherent curricula aligned to the Common Core Learning Standards
- Research-based, effective instruction that yields high-quality student work
- Curricula-aligned assessment practices that inform instruction
- Establishment of a culture of learning that communicates high expectations with supports
- Teacher teams engaged in collaborative practice using the inquiry approach to improve classroom practice (NYCDOE, n.d.-b)

The review process begins with the school completing a School Self-Evaluation Form. This form provides the contextual background for review and affords the school staff an opportunity to reflect upon and assess the school's functioning. For most schools, one external reviewer completes the review in one day.[5]

---

[5] Reviews of larger schools of 1,500 students or more and District 75 schools may engage an additional reviewer.

During the review, the reviewer conducts observations[6] of classrooms and teacher team meetings. In addition, the reviewer meets with the school administrator, students, parents, teachers, and the teacher union chapter leader, as well as examines curricular artifacts and other school-related documents. The reviewer must use specific tools such as the Quality Review rubric,[7] which describes the quality look-fors for each grade; the Classroom Visitation Tool,[8] an evidence-gathering document; and the Record Book, used to document and organize the evidence collected throughout the site visit. At the conclusion of the site visit, the reviewer provides verbal feedback to the principal and the school leadership team on the school's preliminary rating, emphasizing areas of celebration and areas of focus with specific examples gathered from the visit.

After a quality assurance process, a final report detailing the findings of the Quality Review is published on the New York City Department of Education's website. The information gathered from the Quality Review is used as another tool to evaluate school performance and as a lens with which to better understand and interpret the quantitative data (e.g., student test scores, school climate survey, gap scores).

Discussion

As exemplified by the state, district, and the Alberta province examples, next phase accountability systems embody the following critical features:

- Broader set of outcome measures
- Mix of state and local indicators
- Measures of opportunities to learn
- Data dashboards
- School Quality Reviews

Though only emerging out of the long shadow of NCLB and Race to the Top policies, these examples demonstrate that there is a will and a way to develop new systems of accountability that focus on promoting and supporting continuous improvement rather than compelling compliance through external mandates. Importantly, states can incorporate the first three of the critical features of next phase accountability systems as an indicator of school quality or student success under ESSA. Moreover, states can take advantage of the flexibility that the law provides for employing data dashboards and SQRs to diagnose and support continuous improvement. This opportunity encourages the development of more meaningful state accountability systems. To develop and implement a multiple measures system of accountability and continuous improvement, four enabling conditions should be considered: capacity, stakeholder engagement, organizational structures, and flexibility and local control.

---

[6] For schools with 1,500 students or more and District 75 schools, the reviewer conducts 12–15 classroom observations.

[7] Retrieved from http://schools.nyc.gov/NR/rdonlyres/8C11A001-7E78-469D-996F-B0C3703CEA81/0/201314QualityReviewRubric.pdf.

[8] See the NYCDOE Reviewer Handbook for the Quality Review for an example of the Classroom Visitation Tool. Retrieved from http://schools.nyc.gov/NR/rdonlyres/5CC3B3DF-49BF-4BA6-AFD0-0FE6BFBDC3E1/0/ReviewerHandbook201314CFarina.pdf.

Develop System Capacity

As new accountability systems reflect a broader definition of deeper learning and student and school success, a multiple measures approach to school accountability will require the use of new metrics and a greater number of them with which to gauge progress and support improvement. Thus, professional development on data literacy is needed to ensure that all stakeholders (e.g., educators, parents, community members, policymakers), first, understand what the data represent and say about how schools are progressing and, second, develop stakeholders' capacity to make reasoned, well-informed decisions from the data to support improvement efforts.

In addition, to incorporate school review processes like the one used in New York City, developing and honing the expertise of the reviewers is essential to ensuring that the school reviews are viewed as credible and the results are a valid, accurate reflection of the teaching and learning going on in the school. Growing the capacity of a cadre of expert and peer reviewers, however, can be a heavy lift for most systems new to this type of work. And in very large state systems like that of California's, the State Education Agency may not be able to conduct school reviews for all schools. Thus, consideration must be given to the development of a tiered system of support. In California, the County Offices of Education and the newly created agency, the California Collaborative for Educational Excellence (CCEE), will act as intermediaries between the California Department of Education and schools and districts. The County Offices of Education, which are regionally based, and the CCEE will provide direct support and interventions to struggling schools and districts. However, capacity within these agencies must be developed through extensive and ongoing training. Moreover, clear standards and criteria for school quality must be developed. In California, the State Board of Education recently adopted the LCFF evaluation rubrics, allowing schools and districts to self-assess how well they are meeting the state's priorities and goals as well as their own, which will then develop the capacity of the state to support its districts and schools.

Engage Stakeholders and Build Consensus

The new vision for accountability systems that is outlined by Darling-Hammond, Wilhoit, and Pittenger (2014) demands a shift in the way the public views the purpose of school accountability. Essentially, the purpose of school accountability must be reframed from an exercise in compliance to a way to focus and support continuous improvement throughout the entire educational ecosystem. To accomplish this, steps must be taken to build consensus around which educational goals and priorities best reflect the community's values and needs. ESSA and California's LCAP process exemplify the importance of soliciting public input from various stakeholder groups so that multiple voices are heard and are part of the development process.

In doing so, educators, administrators, and policymakers, who are responsible for implementing the redesigned accountability system, may come to view school accountability as something they are coconstructing rather than something that is being forced on them. In addition, parents and community members, who are often left out of the conversation altogether, may see next phase accountability as something that is responsive to their needs and aligned to the aspirations they have for their children and the community. The use of data dashboards to report the progress of schools aims to increase transparency in the system and ensure that the strengths and weaknesses that are revealed in the data are communicated and shared consistently. Moreover, the system's transparency will help to ensure reciprocal accountability. That is, when the data reveal areas in need of development, each level of the system (e.g., the state, the district, the school) must rise to the challenge and make contributions toward improvement efforts. It is only through consensus building and transparency that new systems of accountability can truly be viewed as systems of shared responsibility and continuous improvement.

Create Structures That Fast-Track Support to Schools

To ensure that accountability policies support continuous improvement and reciprocal responsibility, schools must receive the necessary support and resources to carry out improvement efforts. As the multiple measures system brings to light schools' areas of strengths and weaknesses, essential resources and support must be allocated quickly and in a fair and equitable manner (Darling-Hammond, Wilhoit, & Pittenger, 2014). Schools and districts would benefit from processes that streamline resource allocation to expedite and support school improvement efforts. Furthermore, the creation of entities that provide technical assistance and support to schools like the CCEE is critical. The CCEE is a state agency that plays a key role in supporting improvements in teaching and learning in schools as well as providing direct intervention to those that are struggling.

Support Flexibility and Innovation at the Local Level

As states incorporate locally developed or state-supported metrics into their accountability systems, State Education Agencies (SEAs) will have greater responsibility for providing resources and supporting innovation at the local level. In New Hampshire, for instance, the SEA provides ongoing training and technical assistance to the districts implementing the PACE pilot. This is essential for ensuring that all the pilot districts have the capacity to locally develop performance-based assessments and fully implement the state's vision of a balanced school accountability system. In California, the SEA bears responsibility for conducting validity studies to vet and validate new measures so that districts have a pool of state-supported indicators from which to choose. In doing so, districts are afforded the flexibility with which to address needs and goals that are responsive to their local communities and contexts. These locally determined and reported indicators will provide greater insight into the state-required and state-reported indicators. That is, stakeholders will have more information that either complements what is gleaned from state-required and state-reported indicators or that shows evidence of an alternative view of what the state-required and state-reported indicators suggest. This is the grist with which stakeholders can make reasoned and informed decisions about student and school success and tailor interventions and support to specific areas of need. Next phase state accountability systems that support innovation and flexibility at the local level necessitate a more interdependent role for SEAs to play in regard to the continuous improvement of schools and districts.

Conclusion

To fulfill the Elementary and Secondary Education Act's original intent—to ensure that students from disadvantaged groups receive a high-quality education that prepares them to succeed in their postsecondary choices and life—state accountability systems must move beyond reliance on standardized test scores in ELA and math to assess student and school success. A multiple measures approach to accountability can incentivize and support states to develop a broader range of knowledge, skills, and dispositions in all students. Moreover, the approach provides transparency to the system so that all stakeholders have access to the information with which to make informed decisions about how to support deeper learning and continuous improvement as well as ensure that all actors at all levels of the system, not only those on the front line, are held accountable. Doing so may incentivize the behaviors we seek to encourage (e.g., a focus on a holistic view of student learning and development, the incorporation of noncognitive indicators) and safeguard against the behaviors we seek to discourage (e.g., a shortsighted focus on standardized test scores in two subject areas, the narrowing of the curriculum to tested subject areas). As was seen and experienced with the state accountability systems in the NCLB era, unintended consequences of accountability can

proliferate when accountability systems fixate on a narrow view of student and school success and prioritize compliance and avoiding punishment over continuous improvement. ESSA presents new opportunities to design next phase accountability systems that concentrate on deeper learning, emphasize continuous improvement, and further equity goals. It is a call to action that must be heeded.

References

Achieve and Advance CTE. (2016). How states are making career readiness count: A 2016 update. Washington, DC: Authors. Retrieved March 21, 2017, from https://cte.careertech.org/sites/default/files/MakingCareerReadinessCountUpdate-2016.pdf

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college.* Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education.

Alberta Education. (2016, October). Overall Provincial Summary. Retrieved July 5, 2017, from https://education.alberta.ca/media/3273034/province_province-report-accountability-pillar-overall-summary.pdf

Allensworth, E., Nomi, T., Montgomery, N., & Lee, V. A. (2009). College preparatory curriculum for all: Academic consequences of requiring algebra and English I for ninth graders in Chicago. *Educational Evaluation and Policy Analysis, 31*(4), 367–391.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: Authors.

Ancess, J. (1996). *Outside/inside, inside/outside: Developing and implementing the School Quality Review.* New York, NY: National Center for Restructuring Education, Schools, and Teaching at Teachers College, Columbia University.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258–267.

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

California Department of Education. (2016, May). *Preparing all students for college, career, life, and leadership in the 21st century.* Sacramento, CA: Authors.

California Office to Reform Education. (2013). School Quality Improvement System. Retrieved January 22, 2015, from http://coredistricts.org/school-quality-improvement-system

California Office to Reform Education. (2014, August 27). Local educational agencies' request for waivers under section 9401 of the Elementary and Secondary Education Act of 1965. Retrieved January 22, 2015, from http://coredistricts.org/wp-content/uploads/2014/10/CORE-waiver-application-amended-fall-2014-v5-9-4-14-current-approved.pdf

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis, 24*(4), 305–331.

Center for American Progress and the Council of Chief State School Officers. (2014). *Next-generation accountability systems: An overview of current state policies and practices.* Washington, DC: Authors.

Coburn, C. E., Hill, H. C., & Spillane, J. P. (2016). Alignment and accountability in policy design and implementation: The Common Core State Standards and implementation research. *Educational Researcher, 45*(4), 243–251.

Cohen, J., McCabe, E. M., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record, 111*(1), 180–213.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education and Welfare, Office of Education.

Cook, T. D., Murphy, R. F., & Hunt, H. D. (2000). Comer's school development program in Chicago: A theory-based evaluation. *American Educational Research Journal, 37*, 535–597.

Darling-Hammond, L. (1990). Institutional policy into practice: The power of bottom over the top. *Educational Evaluation and Policy Analysis, 12,* 233–241.

Darling-Hammond, L. & Adamson, F. (Eds.). (2014). *Beyond the bubble test: How performance assessments support 21st century learning*. San Francisco, CA: Jossey-Bass.

Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). *Pathways to new accountability through the Every Student Succeeds Act*. Palo Alto, CA: Learning Policy Institute.

Darling-Hammond, L., & Plank, D. N. (2015). *Supporting continuous improvement in California's education system*. Stanford, CA: Policy Analysis for California Education and Stanford Center for Opportunity Policy in Education.

Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). Accountability for college and career readiness: Developing a new paradigm. *Education Policy Analysis Archives*, *22*(86), 1.

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

Duckworth, A., & Carlson, S. M. (2013). Self-regulation and school success. In B. W. Sokol, F. M. E. Grouzet, & U. Muller (Eds.), *Self-regulation and autonomy: Social developmental dimensions of human conduct* (pp. 208–230). New York, NY: Cambridge University Press.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*(12), 939–944.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237–251.

Elmore, R., & Fuhrman, S. H. (1993). Opportunity to learn and the state role in education. In S. L. Traiman (Ed.), *The debate on opportunity-to-learn standards: Supporting works* (pp. 73–102). Washington, DC: National Governors Association.

Figlio, D., & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbooks in Economics*, Vol. 3 (pp. 383–421). The Netherlands: North-Holland.

Figlio, D., & Rouse, C. E. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics, 90*(1–2), 239–255.

Fowler, A. (2015, September 15). *Education Quality Reviews: The whole picture*. The VTDigger. Retrieved July 5, 2017, from https://vtdigger.org/2015/09/15/amy-fowler-education-quality-reviews-the-whole-picture/

Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review, 57*, 812–828.

Hamilton, L., Berends, M., & Stecher, B. (2005). *Teachers' responses to standards-based accountability*. Santa Monica, CA: RAND Corporation.

Hanushek, E. (1981). Throwing money at schools. *Journal of Policy Analysis and Management, 7*(1), 1941.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482.

Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy*

Analysis, *30*(2), 75–110.

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Committee on Appropriate Test Use, National Research Council. Washington, DC: National Academy Press.

Hoffer, T. (1997). High school graduation requirements: Effects on dropping out and student achievement. *Teachers College Record, 98*, 684–707.

Holmstrom, B., & Costa, J. R. (1986). Managerial incentives and capital management. *The Quarterly Journal of Economics, 101*(4), 835–860.

Jackson, C. K., Johnson, R., & Persico, C. (2014). *The effect of school finance reforms on the distribution of spending, academic achievement, and adult outcomes*. Cambridge, MA: National Bureau of Economic Research. Retrieved from http://socrates.berkeley.edu/~ruckerj/Jackson_Johnson_Persico_SFR_LRImpacts.pdf

Jacob, B. (2002). *Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools* (Working Paper #8968). Cambridge, MA: National Bureau of Economic Research. Retrieved June 19, 2017, from http://www.nber.org/papers/w8968.pdf

Jacob, B. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics, 89*(5–6), 761–796.

Jacob, B. (2007). *Test-based accountability and student achievement: An investigation of differential performance on NAEP and state assessments* (Working Paper #12817). Cambridge, MA: National Bureau of Economic Research. Retrieved June 12, 2017, from http://www.nber.org/papers/w12817.pdf

Kentucky Department of Education. (2012). *Unbridled learning accountability model (with focus on the next-generation learners component)*. Frankfort, KY: Author. Retrieved March 5, 2015, from http://education.ky.gov/comm/UL/Documents/WHITE%20PAPER%20062612%20final.pdf

Klein, A. (2015, September 25). CORE districts have waiver renewed, shed 'high risk' label. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/campaign-k-12/2015/09/embargoed_core_districts_nclb_.html

Klein, S., Hamilton, L., McCaffrey, D., & Stecher, B. (2000). What do tests scores in Texas tell us? *Educational Policy Analysis Archives, 9*(49).

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62*, 172–192.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4–16.

Ma, X., & Klinger, D. A. (2000). Hierarchical linear modeling of student and school effects on academic achievement. *Canadian Journal of Education, 25*, 41–55.

McCaffrey, J. R., Lockwood, D. F., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability* [Monograph]. Santa Monica, CA: RAND Corporation.

Mellor, M., & Griffith, D. (2015). *Multimetric accountability systems: A next-generation vision of student success*. Alexandria, VA: ASCD.

Milgrom, P., & Roberts, J. (1988). An economic approach to influence activities in organizations. *American Journal of Sociology, 94*, S154–S179.

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

Murnane, R. J. (1982). Interpreting the evidence on school effectiveness. *Teachers College Record, 8*(1), 19–35.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: Author. Retrieved June 19, 2017, from

http://www.mat.uc.pt/~emsa/PMEnsino/ANationatRisk.pdf

New York City Department of Education. (n.d.-a). Reviewer handbook for the quality review 2015–2016. Retrieved March 3, 2016, from http://schools.nyc.gov/NR/rdonlyres/5CC3B3DF-49BF-4BA6-AFD0-0FE6BFBDC3E1/0/ReviewerHandbook1415.pdf

New York City Department of Education. (n.d.-b). 2015–2016 Quality review rubric. Retrieved March 3, 2016, from http://schools.nyc.gov/NR/rdonlyres/8C11A001-7E78-469D-996F-B0C3703CEA81/0/QualityReviewRubric1516.pdf

New York City Department of Education. (2015, October 30). Final changes to the School Quality Reports for 2014–15. Retrieved March 3, 2016, from http://schools.nyc.gov/NR/rdonlyres/4F6B6F94-9DE0-4AD6-8F34-7D900E13E5C0/0/201415FinalChanges10302015.pdf

New York City Department of Education. (2016). New York City School Survey.. Retrieved March 3, 2016, from http://schools.nyc.gov/Accountability/tools/survey/default.htm

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale.

Oakes, J. (2004). Investigating the claims in Williams v. State of California: An unconstitutional denial of education's basic tools? *Teachers College Record, 106*(10), 1889–1906.

Oakes, J., & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal, 32,* 151–181.

Piscatelli, J., & Lee, C. (2011). *State policies on school climate and bully prevention efforts: Challenges and opportunities for deepening state policy support for safe and civil schools.* New York, NY: National School Climate Center. Retrieved March 4, 2016, from http://www.schoolclimate.org/climate/documents/policyscan.pdf

Plank, S. B., & Condliffe, B. F. (2013). Pressures of the season: An examination of classroom quality and high-stakes accountability. *American Educational Research Journal, 50*(5), 1152–1182.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrics, 73*, 417–458.

Rothman, R. (2015). *Data dashboards: Accounting for what matters.* Washington, DC: Alliance for Excellent Education.

Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right.* NY: Teachers College Press and the Economic Policy Institute.

Schwartz, H. L., Hamilton, L. S., Stecher, B. M., & Steele, J. L. (2011). *Expanded measures of school performance.* Santa Monica, CA: RAND Corporation.

Shepard, L., (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement, Issues and Practice, 9*(3), 15–22.

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

South Carolina Education and Economic Development Act, S.C. Code Ann. § 59-59-140 (2005).

South Carolina Education Oversight Committee. (2015). *2014–2015 Accountability manual: The annual school and district report card system for South Carolina public schools and school districts.* Retrieved September 29, 2015, from http://www.eoc.sc.gov/Reports%20%20Publications/Acct%20Manual%202014-15/FINAL%202014-15%20Manual.pdf

Spielhagen, F. R. (2006). Closing the achievement gap in math: Considering eighth grade algebra for all students. *American Secondary Education, 34*(3), 29–42.

Sterbinksky, A., Ross, S. M., & Redfield, D. (2006). Effects of comprehensive school reform on student achievement and school change: A longitudinal multi-site study. *School Effectiveness and School Improvement, 17*, 367–397.

Stewart, E. B. (2008). School structural characteristics, student effort, peer associations, and parental

involvement: The influence of school- and individual-level factors on academic achievement. *Education & Urban Society, 40*, 179–204.

Supovitz, J. (2009). Can high-stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change, 10*, 211–227.

Taylor, M. (2013). *Updated: An overview of the Local Control Funding Formula.* Sacramento, CA: Legislative Analyst Office. Retrieved July 3, 2014, from http://www.lao.ca.gov/reports/2013/edu/lcff/lcff-072913.pdf

Taylor, M. (2015, January). *Review of school districts' 2014–2015 Local Control and Accountability Pla*ns (p. 7). Sacramento, CA: Legislative Analyst's Office.

Teitelbaum, P. (2003). The influence of high school graduation requirement policies in mathematics and science on student course-taking patterns and achievement. *Educational Evaluation and Policy Analysis, 25*(1), 31–57.

Vermont State Board of Education. (2014). Education Quality Standards: State Board Rule 2000. Retrieved October 2, 2015, from http://education.vermont.gov/sites/aoe/files/documents/edu-state-board-rules-series-2000.pdf

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., & Gabrieli, J. D. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis, 38*(1), 148–170.

Appendix

Abbreviations and Definitions

CCEE – California Collaborative on Educational Excellence is a public agency established in the California Education Code section 52074 that serves to advise and provide technical assistance to school districts, county superintendents of schools, and charter schools in implementing a local control and accountability plan.

CORE – California Office to Reform Education; a partnership between eight California school districts: Fresno, Garden Grove, Long Beach, Los Angeles, Oakland, Sacramento, San Francisco, and Santa Ana. Their partnership mission is to work together to innovate, implement, and scale successful strategies and tools to prepare all students for college and career. The CORE districts applied for and received an Elementary and Secondary Education Act waiver in 2013 to implement their own school quality improvement system.

LCFF – Local Control Funding Formula; established in the California Education Code section 42238 and enacted in 2013–14, it replaced California's finance system and established base, supplemental, and concentration grants to fund school districts and charter schools as well as county offices of education for oversight activities and instructional programs. The new funding formula is primarily based on student needs, with weights attached to funding allocations for poverty, English learner status, and foster child status. The law also gave responsibility to school districts to make resource allocation decisions to achieve desired results.

LCAP – Local Control and Accountability Plan, established in the California Education Code section 52060 and a critical component of California's Local Control Funding Formula, is a planning tool that school districts, charter schools, and county offices of education are required to use to develop, adopt, and annually update a 3-year plan that outlines annual goals and the progress toward those goals. The goals must address the state's eight priority areas and document how grant funds are spent to increase and improve education for targeted students (e.g., students living in poverty, of English learner status, and of foster child status).

PACE – Performance Assessment for Competency Education is New Hampshire's vision for a new accountability system. It relies on a competency-based approach to instruction, learning, and assessment as well as multiple measures of local and common performance-based assessments to support deeper learning, in addition to standardized assessments.

EQS – Education Quality Standards were adopted by the Vermont State Board of Education in 2014. The standards were established to ensure that all students had access to high-quality educational programming that is substantially equal throughout the state. The standards focused on assuring that all schools delivered on proficiency-based learning, flexible pathways to graduation, safe school environments, high-quality staffing, and financially efficient practices.

EQR – Education Quality Review is a tool by which stakeholders in Vermont will be able hold schools accountable for the Education Quality Standards and drive continuous improvement efforts. The review consists of two complementary processes: the annual snapshot, which gathers quantitative data along five dimensions of school quality, and the integrated field review, which

gathers qualitative data through observations, interviews, and document review on the same five dimensions of school quality.

IFR – Integrated Field Review is the complement to the annual snapshot in Vermont's Education Quality Review and is the tool used to provide stakeholders with an in-depth review of school quality, illuminate potential reasons for that which is revealed by the quantitative data, identify promising practices in the field, and provide support and interventions for school improvement.

About the Author

Soung Bae
Stanford University
soungb@stanford.edu
Soung Bae is a Senior Learning Specialist and UDL Innovation Studio Manager at the Schwab Learning Center at Stanford University. Formerly, she was a Senior Research and Policy Analyst at the Stanford Center for Opportunity Policy in Education. Her research interests focus on school accountability, student engagement, and designing learning environments that appreciate and support learner variability.

Guest Editors

Elizabeth Leisy Stosich
Fordham University
estosich@fordham.edu
Elizabeth Leisy Stosich is an Assistant Professor in Educational Leadership, Administration, and Policy at Fordham University. Previously, she was a Research and Policy Fellow at the Stanford Center for Opportunity Policy in Education. Her research interests include education policy, assessment and accountability, school and district leadership, school improvement, and teachers' professional learning.

Soung Bae
Stanford University
soungb@stanford.edu
Soung Bae is a Senior Learning Specialist and UDL Innovation Studio Manager at the Schwab Learning Center at Stanford University. Formerly, she was a Senior Research and Policy Analyst at the Stanford Center for Opportunity Policy in Education. Her research interests focus on school accountability, student engagement, and designing learning environments that appreciate and support learner variability.

Jon Snyder
Stanford University
jdsnyder@stanford.edu
Jon Snyder is the Executive Director of the Stanford Center for Opportunity Policy in Education (SCOPE). His research interests include teacher learning, conditions that support teacher learning, and the relationships between teacher and student learning.