Next Generation Accountability:
A Vision for School Improvement Under ESSA

Curt M. Adams, Timothy G. Ford, Patrick B. Forsyth, Jordan K. Ware,
Jentre J. Olsen, and John A. Lepine Sr.
Oklahoma Center for Education Policy at the University of Oklahoma

Laura L.B. Barnes, Jam Khojasteh, and Mwarumba Mwavita
Center for Educational Research and Evaluation at Oklahoma State University

LEARNING
POLICY
INSTITUTE

Next Generation Accountability:

A Vision for School Improvement Under ESSA

Curt M. Adams, Timothy G. Ford, Patrick B. Forsyth, Jordan K. Ware,
Jentre J. Olsen, and John A. Lepine Sr.
Oklahoma Center for Education Policy at the University of Oklahoma

Laura L.B. Barnes, Jam Khojasteh, and Mwarumba Mwavita
Center for Educational Research and Evaluation at Oklahoma State University

LEARNING POLICY INSTITUTE

FEBRUARY 2017

Acknowledgments

Foreword

In December of 2015, President Barack Obama signed into law the Every Student Succeeds Act (ESSA) and ushered out 14 years of high-stakes testing under the No Child Left Behind Act. Unlike the "one-size-fits-all" approach of NCLB, ESSA provides states with the opportunity to develop innovative strategies for advancing equity, measuring success, and developing cycles of continuous improvement. Specifically, the new law encourages states to use multiple measures to assess school and student performance. As the authors of this report state, "A broader spectrum of indicators, going well beyond a summary of annual test performance, seems necessary to account transparently for performance and assign responsibility for improvement."

This report, *Next Generation Accountability: A Vision for School Improvement Under ESSA*, comes at just the right time. The authors look at the existing body of evidence on old accountability models to imagine a 21st century education system that foregrounds the college and career readiness of its students. To achieve these ends, knowledge creation for students and teachers, and capacity building of school and district staff, must be the drivers of accountability.

ESSA empowers states to develop and implement accountability systems that drive continuous improvement for all students. For example, states can leverage ESSA to:
- develop useful dashboards of information that provide transparency and guidance for productive action;
- use additional indicators of school quality, beyond the four that are federally required, such as access to a rich curriculum, a productive school climate, and opportunities to learn;
- use continuous measures of achievement (such as scale scores and movement across performance categories) in order to better measure progress and equity gaps; and
- ensure that schools have sufficient time to implement thoughtful and effective accountability systems that incorporate stakeholder feedback and have the capacity to drive effective strategies for improvement in schools.

This report offers a robust conceptual framework for how to approach the new opportunities for creating an accountability system that supports school improvement, and offers detailed examples of the kinds of measures that can reveal needs and progress in actionable ways.

This kind of road map to school improvement is critically important at this moment in our history. Today's fast-paced technological advances are continually automating routine functions that once created low-skilled jobs. As the pace of global change continues, the old factory model of education that prepared students for Industrial Age employment is insufficient for the Information Age in which we now live. Instead of an economy focused on mass production of goods, today's economy demands such skills as the ability to make sense out of complex information and events, think creatively to solve real-world problems, work well with others, engage effectively in cross-cultural contexts, and manage many forms of media as well as quantitative data in sophisticated ways.

A next generation accountability must support the development of these skills and support a 21st century education system that enables a younger generation to gain productive employment, become engaged citizens, and invent new solutions to the pressing problems we face to maintain a robust and thriving society. This report provides a valuable framework for policymakers looking to develop systems of accountability that support and advance 21st century learning for all students.

<div align="right">

Linda Darling-Hammond, President and CEO, Learning Policy Institute
Charles E. Ducommun Professor Emeritus, Stanford University

</div>

Table of Contents

Figures

Tables

Executive Summary

Accountability, as the word implies, is about accounting and responsibility. But reliance on outcome information that is the product of multiple causes and influences cannot explain how and why schools perform as they do and who or what must change when improvement is needed. As school accountability systems evolve and mature, they should emphasize information that enables a clear understanding of school performance and suggests action by appropriate actors, be they legislators, school boards, school leaders, teachers, or parents. A broader spectrum of indicators, going well beyond a summary of annual test performance, seems necessary to account transparently for performance and assign responsibility for improvement.

So, what is the right accountability system now? Those that rely on regulatory compliance are insufficient, even for achieving a vision of improvement as myopic as one based on test scores. Next generation accountability aims at a loftier goal—universal college and career readiness—a goal that current accountability systems were not designed to achieve. To reach this higher level, next generation accountability must embrace a wider vision, distribute trustworthy performance information, and build support infrastructure, while eliciting the assent, support, and enthusiasm of citizens and educators. In this report, we explore this new goal and imagine an accountability approach suited to the task.

The timeliness of this exploration is evident. With the expiration of current No Child Left Behind (NCLB) waivers in August 2016 and implementation of the Every Student Succeeds Act (ESSA), states are about to gain considerably more authority and autonomy over the design of school accountability. These changes provide an opportunity to produce information that genuinely explains how schools and school systems are meeting students' learning and developmental needs. A few states have been working to put more supportive, humane accountability systems in place, but others remain stuck in a compliance mindset that undermines their ability to design effective accountability systems. This report can move both groups forward. For the former, this report serves as a reference point for design and implementation decisions; for the latter, it lays out a vision for next generation accountability that places knowledge creation and capacity building at the center of ongoing quality improvement.

The report has four parts. In part one, we evaluate the informational significance of a single composite accountability indicator. Part two advances the principles and conceptual framework that drive next generation accountability. Part three describes potential indicators of an Educational Quality and Improvement Profile. Part four includes recommendations to guide the design and implementation of next generation accountability systems.

Part One: The Information Significance of a Single Composite Index

As states transition to ESSA, the appropriateness of a single composite indicator to sort schools by effectiveness levels or to identify schools in need of comprehensive intervention is still in question. Of the 43 states that had received NCLB flexibility waivers, 35 applied some type of indicator weighting system in arriving at an overall assessment of school performance. Indeed, many states have relied on an accountability index for so long that policymakers may not understand how reducing school performance to a single indicator hides more than it reveals about teaching and

learning. With an opportunity to design a better system, but without a thorough understanding of measurement limitations, the response to ESSA may focus on the creation of a "better" indicator by merely adjusting the formula used to rank or sort schools. Such a response ignores problems inherent in these systems, and would not provide decision makers with valid information necessary to assess educational quality and improvement.

Before we lay out the argument for next generation accountability, we make explicit the limitations of composite indicators. Four problems stand out as jeopardizing the accuracy and legitimacy of decisions made about schools based on composite indicators: poor conceptual alignment, hidden variance in student performance, misleading accounts of student growth, and the absence of explanatory evidence for making sense of school outcomes.

1. **Poor conceptual alignment.** Some current accountability systems aspire to measure progress toward college and career readiness, but the data such systems use to calculate letter grades do not always align with the competencies associated with 21st century skills. Deeper learning and college and career readiness are multidimensional concepts that include a broader set of competencies than those measured by many state assessments.

2. **Hidden variation in student performance.** Variation in teaching and learning occurs naturally. Students differ in their learning and development for numerous reasons. Rather than collapsing variance into categorical rating schemes, educators need to study patterns in variance across many different factors and conditions. Composite indicators inhibit the construction of meaningful and useful data by combining multiple assessments into a composite indicator. For instance, evidence about achievement equity within and across schools and achievement growth across student groups is critical, but such distinctions are lost when combined into one composite indicator.

3. **Misleading accounts of student growth.** Under many accountability schemes, points are awarded when students change proficiency bands; this practice obscures actual changes in achievement from one year to the next. Further, composite indicators do not report achievement growth by student characteristics or across subjects. By far, trend data present more interpretable information for understanding changes in achievement over time.

4. **The absence of explanatory evidence for making sense of school outcomes.** Low student performance has multiple causes that are not reflected in aggregate test scores. In other sectors—for example, banking, healthcare, and manufacturing—it is unacceptable to make strategic decisions on outcome-only indicators. Yet school accountability systems dependent on composite indicators claim to enable state officials and school leaders to evaluate school effectiveness and the health of the educational system from a single summative outcome indicator.

Part Two: Principles and Framework for Next Generation Accountability

First generation accountability, in compliance with the prescriptions of the federal NCLB law, exposed vast, inequitable differences in student test scores across and within schools. Accountability systems, however, were not effective at informing capacity building within schools aimed at raising and equalizing achievement. As has been argued persuasively elsewhere, greater capacity lies at the heart of quality improvement; with the current emphasis on college and career

readiness, it seems imperative that next generation accountability policies be aligned with this goal. This new vision emphasizes the cognitive and noncognitive competencies expected for success in a postindustrial society and economy.

Next generation accountability is governed by three principles:

1. **Shared Accountability.** The principle of shared accountability reminds us that in a complex enterprise such as public education, performance responsibility is distributed across the system's components and not foisted upon any one group of actors or stakeholders. Control over each of the resources, processes, and/or outcomes essential for improvement resides almost exclusively within a particular stakeholder group; exerting pressure on stakeholders who do not have direct control over these elements is inappropriate and worse, harmful. The goal of "shared accountability" is to create an accountability environment in which all participants recognize their obligations and commitments in relation to each other.

2. **Adaptive Improvement.** First generation accountability assumed that districts and schools are uniform in their capacity to become effective and attain the goals set by accountability policy. Accountability impact studies, however, have documented that schools vary considerably in their capacity to respond coherently to the demands of external accountability. Next generation accountability acknowledges that school capacities differ greatly, and that effectiveness requires a system that is flexible and responsive to particular school conditions.

3. **Informational Significance.** A broad palette of functionally significant indicators to replace a single composite indicator will likely be regarded as informational rather than controlling, thereby motivating stakeholders to action. The information system designed to service next generation accountability should recognize the dual reporting needs of compliance with federal mandates and the particular improvement needs of a state's schools. It should also address the different information needs of state, district, school site leadership, teachers, and parents.

In keeping with these principles, next generation accountability features a two-stage framework. In Stage One, the design places emphasis on providing schools with useful information for their own improvement decisions; Stage One requires a process that is more formative than summative. Stage Two is aggressive, and takes seriously the need to identify and transform schools in catastrophic failure. Next generation accountability requires a significant shift in resource allocations to put in place a support infrastructure capable of serving the developmental and resource needs of schools with differing needs.

Both stages depend on a view of schools as systems whose outcomes are results of the quality of their resources and their processes. Primary resources are conceptualized as a school's organizational and home/community capacities. Essential school processes are conceptualized as instructional and learning capacities. Indicators of these capacities, we argue, will empower stakeholders at all levels, from the legislature to parents, to understand the meaning of school outcomes, and to design and implement effective interventions. A set of summative indicators, together with more formative capacity indicators, will enable the state to identify with confidence those schools in need of expert support and additional resources.

Part Three: An Educational Quality and Improvement Profile

Consistent with the information needs and guiding principles of next generation accountability, we propose for consideration the Educational Quality and Improvement Profile (EQuIP). By reporting data on school resources, processes, and outcomes, an EQuIP provides vital information for making judgments of school quality and helps focus school improvement efforts. It can aid responsible parties in the identification of schools in need of comprehensive support and improvement, and it points to school resources and conditions that need to change to improve observed outcomes. It establishes a level of transparency in reporting student outcomes far exceeding A–F letter grades or any summative index by presenting accurate and clear information on deeper learning and college and career readiness, tracking changes in achievement gaps, and displaying student growth.

EQuIPs benefit every school. Consistently low-performing schools desperately need nuanced data to inform deliberate strategies and focused support for improvement. Even the highest-performing schools cannot afford to be static since the needs of each new group of learners are quite different than the last. Thus, EQuIP is envisioned as a comprehensive profile with multiple uses, depending on the specific needs of each school community. All stakeholders can find useful information in EQuIPs. Parents seeking their children's highest well-being, policymakers seeking efficiency and effectiveness, and educators seeking to create and sustain thriving schools will find accurate and useful information about different aspects of the life and health of schools.

EQuIP provides accountability information to determine how well students are meeting deeper learning and college- and career-readiness standards, and improvement information to better understand potential reasons for observed outcomes. Six guidelines govern how data should be reported.

1. Outcome indicators should report achievement differences by student subgroup performance and changes in individual student performance over time.

2. Outcome indicators should be capable of identifying focus schools, priority schools, and reward schools consistent with criteria for federal waiver requirements.

3. Process and resource indicators should be scientifically defensible and tap conditions, attitudes, structures, and behaviors that can advance the goals of deeper learning and college and career readiness.

4. Indicators should be collected with appropriate frequency and minimal disruption to the learning process.

5. Indicators and measurement methods should have substantial evidence to support their validity and reliability, with the understanding that no measure is perfect in its ability to accurately capture the phenomenon under study.

6. Indicators and measurement methods should change over time in response to a continuous evaluation of the school accountability framework.

7. Indicators should be reported in ways which discourage gaming practices and distortion of school performance.

EQuIP examples for each of the resources, processes, and outcomes discussed are included in this section.

Part Four: Designing and Implementing Next Generation Accountability

In this final section, we translate next generation accountability into a set of recommendations for enacting this framework for both state and local education agencies interested in pursuing this path to educational quality and improvement. These recommendations have three distinct targets:

1. **Accountability Policy.** For many states, the design of next generation accountability will require revision of state statutes that prescribe the formulation and use of composite indicators of school performance for accountability purposes. Going forward, and in accordance with key provisions under ESSA, it seems reasonable that state law set clear and high expectations for an accountability system aligned with deeper learning and college and career readiness, while providing state education departments with the authority and managerial flexibility to design systems that fulfill these expectations. We propose the following recommendations and rationale for an accountability policy:

   a. Do not use a single summative index to report accountability information.

   b. Report outcome evidence in ways that clearly report student performance toward deeper learning and college- and career-readiness standards, changes in student performance over time, and achievement gaps.

   c. Include multiple indicators of capacity for quality improvement as part of a school profile.

   d. Adhere to the Standards for Educational and Psychological Testing, and write the policy in the least restrictive and prescriptive terms possible to allow for corrective action and improvement.

2. **Alignment of Standards, Assessments, and Accountability.** We establish in this report clear working definitions of deeper learning and college and career readiness, positioning them as critical educational outcomes of a next generation accountability system. As an essential first step, care must be taken to ensure that curricular, assessment, and evaluation systems all align with and/or serve these larger operational definitions of what it means to be a healthy, productive citizen.

3. **School, District, and State Capacity Building.** Success in using next generation accountability to elevate educational quality and improvement depends on the degree to which school professionals and stakeholders can unlearn some past ways of doing things. It also requires a support infrastructure exceeding that historically in place. A next generation framework, we argue, identifies five essential systemwide components of a support infrastructure for building capacity across the educational system: (a) state, district, and school leaders must create a systemwide culture grounded in "learning to improve"; (b) learning to improve using EQuIP necessitates the development of strong pedagogical data-literacy skills; (c) resources in addition to funding—including time, access to expertise, and collaborative opportunities—should be prioritized for sustaining these ongoing improvement efforts; (d) there must be a coherent structure of state-level support for learning to improve, including the development of a strong Longitudinal Data System (LDS) infrastructure; and (e) educator labor market policy in some states may need adjustment to support the above elements.

Next Generation Accountability: A Vision for School Improvement under ESSA

Introduction

We write this report when considerable evidence shows an educational system limping and sputtering under test-based accountability. Nationally, student achievement is not at the level it should be, and achievement gaps remain entrenched problems (Mintrop & Sunderman, 2009; National Center for Educational Statistics, 2015). New challenges lurk as well. Results on the Programme for International Student Assessment (PISA) exam, an assessment measuring deeper learning competencies, show no gains during the past decade in math, reading, and science scores for 15-year-olds in the United States. Moreover, U.S. students have lower average performance on test questions requiring creative thinking, problem solving, and knowledge application (OECD, 2014). The challenges confronting schools are great, and when considered alongside expectations of deeper learning and college and career readiness, it becomes even more urgent that we reconsider the function and use of school accountability.

Accountability, as the name implies, is about accounting for outcomes and accepting responsibility. In education, citizens want a system that supports the growth and future vitality of their economy. But outcome information, especially when it is the product of multiple causes and influences, cannot explain how and why schools are performing as they are, and who or what must change when improvement is needed. Whatever changes are made in the accountability system should produce significant information enabling an understanding of school performance and suggesting action by appropriate actors, be they legislators, school boards, school leaders, teachers, or parents. A broader spectrum of indicators, well beyond a summary of annual test performance, seems essential to transparently account for performance and assign responsibility for improvement (Hargreaves & Fullan, 2012; Harris, 2011; Mourshed, Chijoke, & Barber, 2010). The right accountability system can, as Darling-Hammond and colleagues argue, "raise the bar of expectations … and trigger the intelligent investments and change strategies that make it possible to achieve these expectations" (Darling-Hammond, Wilhoit, & Pittenger, 2014, p. 4).

Accountability systems reliant on regulatory compliance are insufficient, even for achieving a vision of improvement as myopic as one based on test scores. That evidence is plentiful (see, for example, Hamilton, Schwartz, Stecher, & Steele, 2013). More daunting still, the challenge for next generation accountability is aimed at an even loftier goal—universal college and career readiness—a goal that many state accountability systems are still not designed to achieve. So, what is the right accountability system now? The purpose of this report is to explore this new goal and imagine an accountability approach suited to the task. To achieve at this higher level, next generation accountability must embrace a wider vision, distribute credible performance information, and build support infrastructure, all while eliciting the assent, support, and enthusiasm of citizens and educators.

This report comes at an ideal time. With the expiration of current No Child Left Behind (NCLB) waivers in August 2016, and the impending implementation of the Every Student Succeeds Act (ESSA) during the 2017–18 school year, states will gain considerably more authority and autonomy to redesign their accountability systems. This is an opportunity for states to design

systems that enable a more complete and precise understanding of how schools and school systems are meeting the learning and developmental needs of all students. A few states have already been working to put more supportive, humane accountability systems in place. Undoubtedly, these states are ready for more challenges, whereas others remain stuck in a compliance mindset that has undermined their capacity to design a system that works for schools, students, and families. This report can move both groups forward. For the former, this report serves as a reference point for design and implementation decisions, while for the latter it lays out a vision for next generation accountability that places knowledge creation and capacity building at the center of ongoing quality improvement.

The report is organized in four parts. In part one, we evaluate the informational significance of a single composite accountability indicator. Part two advances the principles and conceptual framework that drive next generation accountability. Part three describes potential indicators of an Educational Quality and Improvement Profile (EQuIP). Part four includes recommendations that guide the design and implementation of next generation accountability systems.

Part One: The Informational Significance of a Single Composite Index

In the majority of states, accountability under NCLB, and more recently NCLB waivers, has relied primarily on a composite indicator like an index score, an A–F letter grade, a star system, or a qualitative descriptor (e.g., reward, focus, excellent, meets standards) to sort schools by their different effectiveness levels. In fact, 35 of the 43 states that had obtained NCLB flexibility waivers apply some type of weighting system to the indicators of school performance they currently employ in arriving at an overall assessment of school performance (Martin, Sargrad, & Batel, 2016).

Proponents may laud the simplicity of using a recognizable indicator to evaluate schools, but this simplicity comes at the expense of meaningful knowledge about the quality of learning experiences available to students. One of the main issues with a single composite index is that subgroups, grades, and other measures of interest are all subsumed under one summative measure that obscures variation in performance as well as other information important to school improvement. Ironically, the effort to simplify outcome information has frustrated both meaningful assessment and evaluation of schools.

Issues of gaming and test perversion are prevalent in accountability systems that use single composite-type indicators (Booher-Jennings, 2005; Forte, 2010; Ho, 2008). There are several reasons for this. Typically, the primary component of these composite indicators is some function of aggregate student performance on standardized tests. Aggregate test performance is susceptible to gaming, especially through practices that do not improve student learning and achievement (Forte, 2010; Rothstein, Jacobson, & Wilder, 2008). Furthermore, reliance on a single annual test (usually end of year) does not provide information on learning growth that could be provided by more formative, longitudinal measures.

In addition to justifiable concerns about the validity of single composite-type indicators for improvement purposes, these indicators, because of their summative nature, are extremely susceptible to *Campbell's Law*. Donald Campbell warned of the use of a single indicator for decision making:

> The more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor (Campbell, 1979, p. 85).

Under test-based accountability, consequences and/or rewards are attached to *outcomes, not behaviors.* This is problematic because both desired behaviors (e.g., efforts to improve teaching practice) and undesired behaviors (teaching to the test, narrowing of the curriculum, cheating, etc.) are reinforced equally, so long as the outcome is the same (Ryan & Brown, 2005).

ESSA regulations provide states an opportunity to improve the nature and function of performance information. Accountability supportive of deeper learning and college and career readiness requires performance information that can be used in the service of knowledge creation and capacity building, in contrast to using it as a regulatory device meant to induce teachers and schools merely to work harder (Darling-Hammond, Wilhoit, & Pittenger, 2014). Useful performance information renders a complete picture of valued learning outcomes, as well as changes in these outcomes over time. Additionally, as businesses, healthcare, and industrial organizations have learned, process and resource information are just as critical for strategic decisions as outcome data (Langley, Moen, Nolan, Nolan, Norman, & Provost, 2009).

As states transition to ESSA, the appropriateness of a single composite indicator to sort schools by effectiveness levels or to identify schools in need of comprehensive intervention is still in question. States have relied on an accountability index for so long that even with an opportunity to design a better system, many policymakers may not understand how reducing school performance to a single indicator hides more than it reveals about teaching and learning. Without a thorough understanding of measurement limitations, the response to ESSA may focus on the creation of a "better" indicator by merely adjusting the formula used to rank or sort schools. Such a response effectively ignores problems inherent in current accountability measures, and would not provide decision makers with sufficient data to assess educational quality and improvement.

Limitations of Composite Indicators

Before we lay out the argument for next generation accountability, we will make explicit the limitations of composite indicators. For this, we turn to our evaluation of Oklahoma's A–F accountability letter grades, which is described in *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education* (Adams et al., 2015b). Although the evidence reported comes from one type of indicator, A–F letter grades, the methods used to derive any single summative indicator are similar (Domaleski & Perie, 2013; Howe & Murry, 2015; Polikoff, McEachin, Wrabel, & Duque, 2014). Problems detected in the function and analysis of an A–F letter grade approach are not likely to be unique. Four problems stand out as jeopardizing the accuracy and legitimacy of decisions made about schools based on composite indicators:

1. Poor conceptual alignment

2. Hidden variation in student performance

3. Misleading accounts of student growth

4. Absence of explanatory evidence for making sense of school outcomes

1. Poor Conceptual Alignment

The informational significance of any indicator depends on what it reveals about the object(s) of measurement (Ryan & Weinstein, 2009). To assess the usefulness of A–F letter grades, we need a conceptual understanding of college and career readiness. Considerable work has gone into defining essential knowledge, skills, competencies, dispositions, and mindsets for students to be prepared for an ever-changing and dynamic world. We draw on this professional literature to offer a definition that reflects growing expectations that schools will engage students in deeper learning so they are prepared for future life circumstances.

The conceptual roots of college and career readiness are consistent with the knowledge, skills, and dispositions associated with deeper learning. Deeper learning is not a new concept, nor does it recast the way many people think of learning and the purpose of education. In fact, its cognitive elements extend back to research of gestalt psychologists (e.g., Katona, 1942; Wertheimer, 1959) who differentiated between reproductive and productive thinking, with the latter processes involving application of knowledge to new problems and tasks (National Research Council, 2012). Deeper learning may not be a new idea, but it provides a useful distinction between the narrow focus on discrete knowledge and skills under NCLB, and the development of robust cognitive and

noncognitive attributes that many identify as necessary for the new economy (Bellanca, 2015). A call for schools to produce college- and career-ready graduates is a call for deeper learning processes and outcomes (National Research Council, 2012).

The National Research Council (2012) defines deeper learning as the ability to transfer knowledge and skills learned in one context to new situations, complex problems, and nonroutine tasks. Deeper learning outcomes, which are distinct from yet related to processes, include diverse cognitive, intrapersonal, and interpersonal competencies that guide human thinking and adaptive behavior. The Hewlett Foundation (2014) defines the outcome of deeper learning as the ability to master rigorous academic content through the application of higher order skills, including critical thinking and problem solving, communication, collaboration, learning to learn, and the development of an academic mindset. College and career readiness represents deeper learning competencies that high school graduates need for effective functioning in a postindustrial society. These competencies vary somewhat depending on definitions, but there is general agreement that college and career readiness is observable in mastery of integrated content and ideas, critical thinking and problem solving, creativity and imagination, effective written and oral communication, academic mindsets, and collaboration (Bellanca, 2015).

Accurate inferences about schools based on composite indicators depend on the alignment between state curricular tests, and the cognitive and noncognitive competencies associated with deeper learning and college and career readiness. At best, a summative indicator derived from student test scores may yield information about basic content knowledge and skills, but it does not capture other cognitive and noncognitive competencies required for productive functioning in the emerging economy. Composite indicators like A–F letter grades fail to capture adequately the mastery of concepts and ideas, creativity and imagination, critical thinking and problem solving, interpersonal abilities and effective communication, and learning mindsets.

With accountability indicators largely derived from student test scores, it would be reasonable to assume that a school grade would indicate mastery of content knowledge. This is actually not the case in Oklahoma and in many states. Deep understanding should not be conflated with basic knowledge and skills. Deeper knowledge assessments require students to apply knowledge and skills learned in one context to a specific problem or task in another context. Cognitive psychologists often refer to knowledge application as transfer, the ability to recall information and to use it for solving a complex problem (Glaser, 1984). Assessing for deeper knowledge depends on evaluating the cognitive complexity required of students to complete a task, solve a problem, or correctly answer a question (Darling-Hammond & Adamson, 2010).

Figure 1
Alignment Between A–F Letter Grades and College and Career Readiness

College and Career Readiness
• Mastery of Integrated Content Knowledge
• Critical Thinking and Problem Solving
• Creativity and Imagination
• Effective Written and Oral Communication
• Collaboration
• Academic Mindsets

A–F Grades
• Basic Content Knowledge

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education*.

The cognitive complexity of an assessment is often evaluated through Norman Webb's Depth of Knowledge framework. Webb (1997, 2002, 2007) advances four levels of knowledge that define different degrees of student understanding and application. Level one involves recall and reproduction of basic terms, facts, concepts, and/or procedures for routine tasks. Level two requires basic application of skills and concepts, wherein students use information for simple procedures, know which procedure to select, interpret and use simple graphs, and organize data and information. Level three requires strategic thinking as students work through abstract, complex, and nonroutine problems. Level four requires extended thinking to perform investigations of real-world tasks that require time to research and problem solve, and to process information from multiple sources.

Assessments capable of measuring dimensions of deeper learning and college and career readiness span the different knowledge levels. Tests should not be designed to assess only for strategic and extended thinking any more than they should be limited to recall and basic application. Tests need to represent and sample fairly knowledge and skills that cover the range of competencies associated with deeper learning (Webb, 2007). However, strategic and extended thinking are routinely underrepresented in many current state assessments (Herman & Linn, 2013), most of which merely require students to repeat what they learned in class (National Research Council, 2012). To that point, a 2012 RAND study evaluated test questions from 21 states known to have cognitively demanding assessments, and found that fewer than 2% of the items in math and 21% in English/Language Arts assessed deeper cognitive competencies (Yaun & Le, 2012).

2. Hidden Variation in Student Performance

Accountability grades that combine unlike and overlapping informational measurements into a single indicator may appeal to those who desire a simple signal of school quality, but the signal often comes at the cost of misrepresenting actual patterns in student performance. We demonstrate problems with the informational value of letter grades, or any gross composite indicator, by plotting distributions of reading scores based on A–F letter grades and student characteristics. Graphs in Figure 2 compare student test score distributions by A–F letter grades. The blue plot represents student scores within the specific letter grade, and the plot outlined in orange is the spread of scores in the population. Test score variation within letter grades has more in common than what is reasonably expected. Distributions of reading scores in B, C, and D schools are nearly identical. F schools had more low-performing students than the population, but arguably, the distribution of reading scores is not much different than that of the population.

With composite indicators, we lose sight of the fact that the grade does not reflect the performance of many students within schools. Many students in D and F schools did not perform as poorly as the grade suggests; they had reading scores as high as, and even better than, some students in A and B schools. Additionally, a large percentage of students in A and B schools scored lower than students in B and C schools, and many students in B and C schools scored lower than students in D and F schools. In many cases, judgments about student performance may be accurate, but there is also a high probability that judgments may misrepresent actual student achievement.

Figure 2
Distribution of 2014 Reading Scores for Elementary and Middle School Students
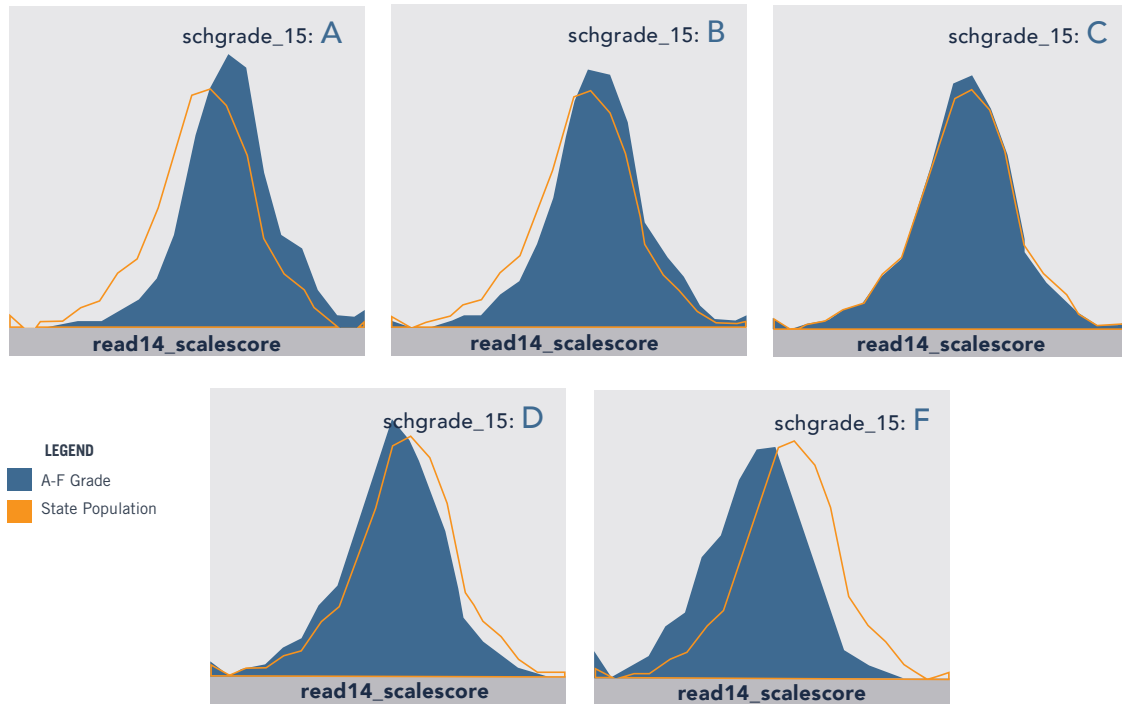
In Figure 3, we illustrate the lack of transparency in reporting achievement gaps. Notice the large difference in reading scores between Free and Reduced Lunch (FRL) and Non-Free and Reduced Lunch (Non-FRL) students in A and B schools. For 2012 and 2013, FRL students scored approximately .47 standard deviation units lower than Non-FRL students. In 2014, this gap increased to .54 standard deviation units. This amounts to a difference of approximately 45 scale score units, a sizable difference between these groups.

Figure 3
Average Achievement Gap in A and B Schools for the 2012–2014 School Years
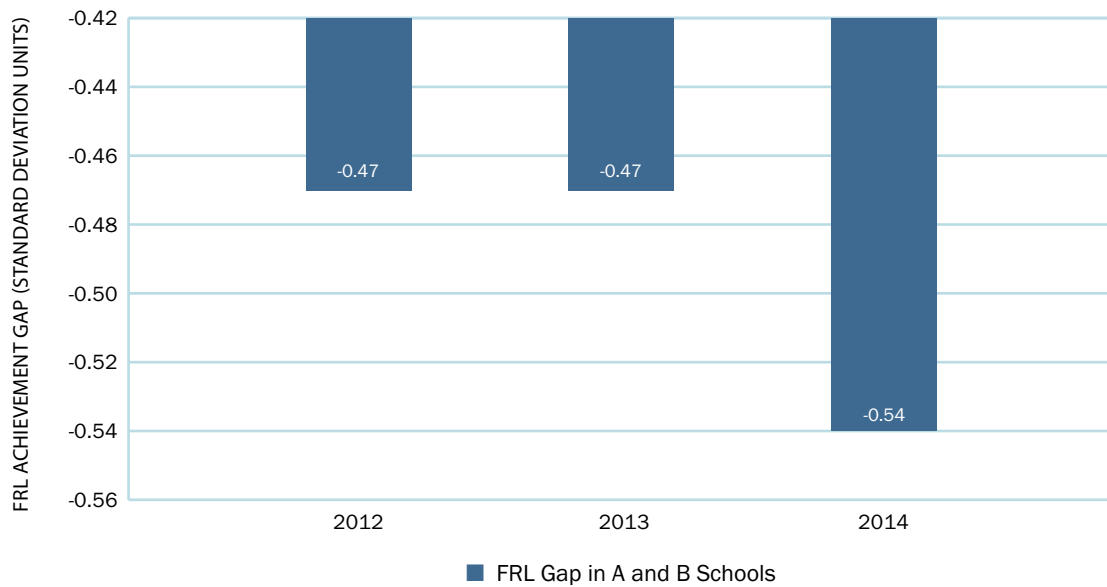
Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

3. Misleading Accounts of Student Growth

An essential question that has lingered unaddressed beneath the debates, opinions, and analyses of test-based accountability is the degree to which student performance is improving under the accountability system. We present graphs of achievement growth in math and reading over a 3-year period. Figure 4 reports 3-year growth rates for the math and reading exams. The bars are set at zero for no growth, and extend upward for average gains and downward for average declines. The blue bars are for math and the orange for reading. Given the resources, time, pressure, and costs associated with the A–F letter grade system, we would expect to see bars trending upward at a level to reflect significant gains in reading and math achievement. Keep in mind that scale scores range from 400 to 990 with a standard deviation around 90 points.

Figure 4

Three-Year Change in Math and Reading by 2011–12 School Grade, and Free and Reduced-Price Lunch Status

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

We do not find any meaningful gains in student achievement for 2012–2014. Instead of average "growth," achievement has declined for students in each letter grade category. In A and B schools, math and reading scores declined for Non-FRL and FRL students. In C schools, Non-FRL and FRL students had average declines in math, whereas in reading FRL students dropped and Non-FRL basically held steady. The only nominal gains for both Non-FRL and FRL students were in reading for D schools, and, even here, gains are so small they do not equal a difference of one additional question correct on a 50-item test. The most troubling finding is that Non-FRL and FRL students in F schools had the largest average drops in test scores, even though these students had the greatest room for growth.

Drops in test scores are also reflected in stagnant and declining achievement trends across the three testing periods. As seen in Figures 5 and 6, reading and math test scores trended downward for students who were in A and B schools in 2012. Reading scores for students in C schools remained flat, but math scores for these students had a slight downward trend. Students in D schools had a slight positive trend in reading, but their math scores remained flat. Students in F schools stand out for the largest decline in test scores. The achievement difference increased for students in F schools, even as test scores declined for students in schools with better grades.

Figure 5
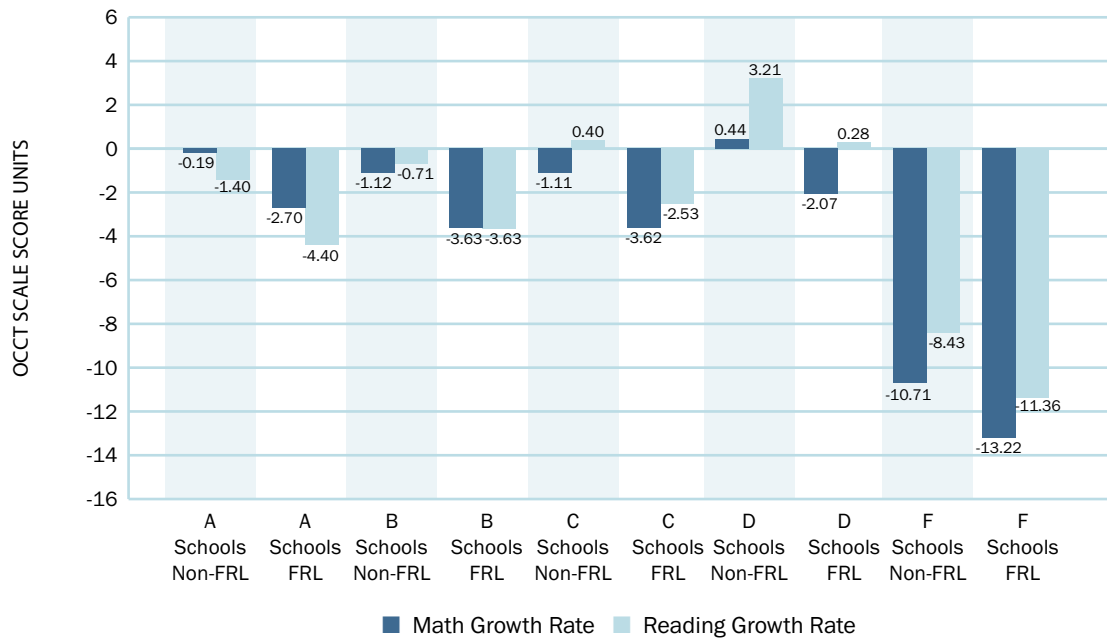**2012–14 Achievement Growth in Reading by 2011–12 Grade**

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

Figure 6
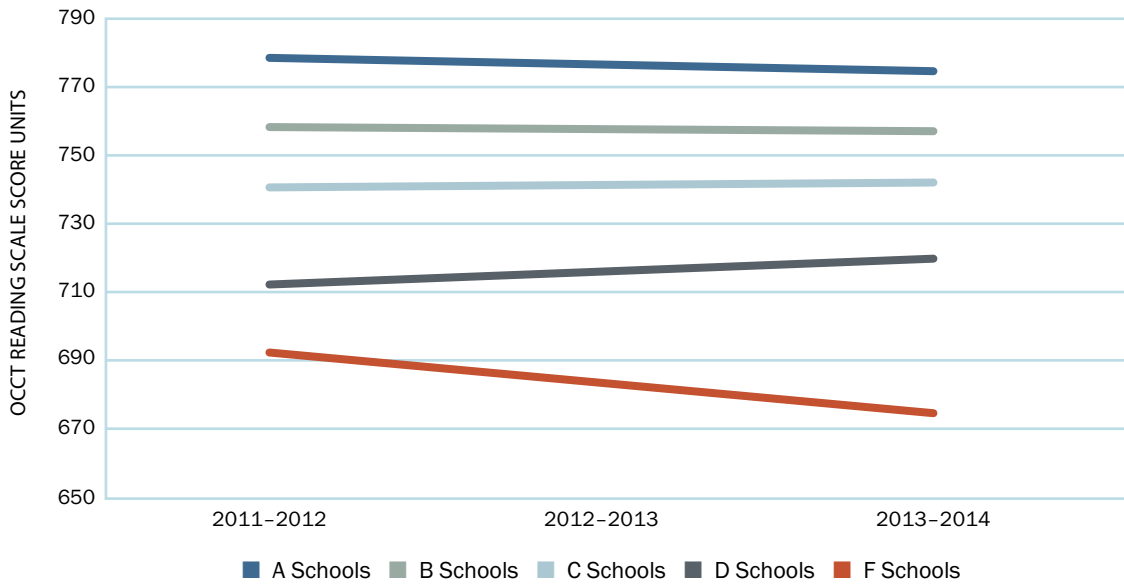**2012–14 Achievement Growth in Mathematics by 2011–12 Grade**
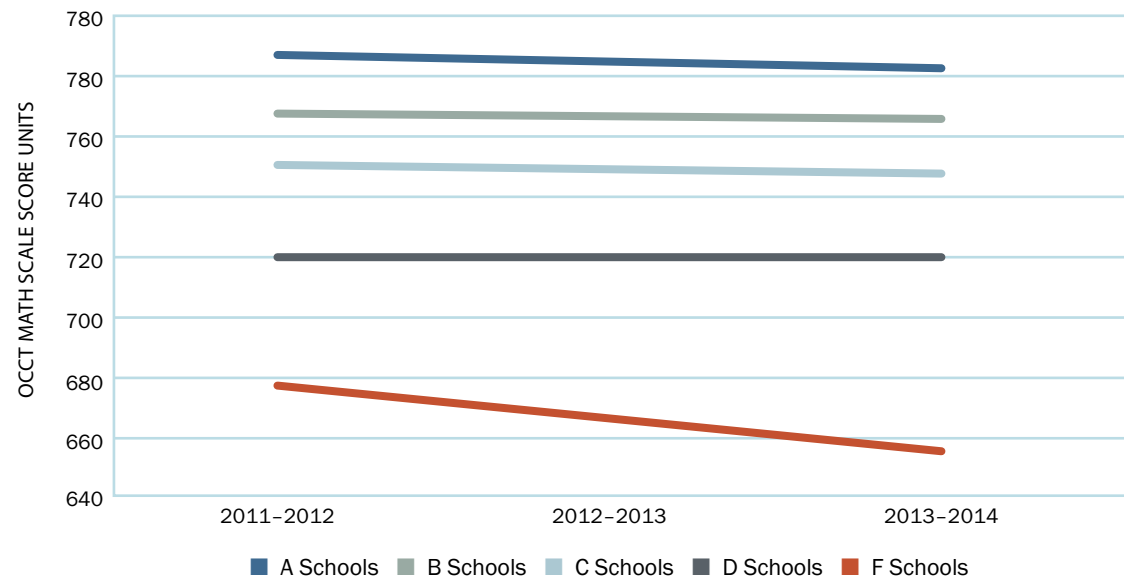
Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

Flat or declining test scores are remarkable findings given the time, money, and emphasis devoted to preparing students to pass state tests. With considerable attention to improving test scores, we would expect to find modest gains following the implementation of the A–F letter grade accountability system (Koretz & Hamilton, 2006). Practices such as coaching, use of practice tests, curriculum realignment, and focusing attention on borderline test-takers are common responses to the demands of high-stakes testing and tend to artificially inflate test scores without producing real gains in learning (Koretz & Hamilton, 2006). The lack of improvement in test scores suggests actual achievement decline may be greater than that measured by the tests and calls into question the validity of the motivational premise of the A–F letter grade accountability system. Stagnant and declining performance on tests of low-level knowledge and skills does not evoke confidence that the A–F grading system, as the cornerstone of Oklahoma's current accountability approach, can help schools reach standards well beyond mere proficiency.

4. The Absence of Explanatory Evidence for Making Sense of School Outcomes

An appeal of composite indicators is that they identify schools in need of state intervention. Low-performing schools do need pressure to improve, but pressure alone does not fix a pattern of distress and dysfunction. Instead, schools need resources, support, and expertise to design and develop changes that have a legitimate chance of enhancing student learning. The improvement problem is not with identifying low-performing schools; it is in knowing what and how to improve (Bryk et al., 2015).

Composed almost exclusively of test score results, composite indicators do not account for the resources and processes behind educational outcomes. The absence of information on factors contributing to healthy and thriving schools leads to misguided assumptions that more often than not add noise and waste to a system already plagued with burdensome mandates and regulations. Educators, policymakers, and the public alike deserve to know about the distribution of resources, processes, and conditions in schools so that improvement can target the likely causes of poor outcomes.

Part One: Conclusion

A single composite indicator has been one of the more controversial and problematic features of recent school accountability systems. Its appeal resides in the perceived ease by which judgments of school quality can be made. Behind the appeal of parsimony, however, is the high probability of distortion in the measurement of school effectiveness. We call attention to four specific measurement problems that limit the meaning and usefulness of composite indicators.

1. **There is a large conceptual gap between composite indicators like A–F grades and the goals of deeper learning and college and career readiness.** Current accountability systems purport to measure progress toward college and career readiness, but the data used to calculate letter grades do not align with the competencies associated with 21st century skills. Deeper learning and college and career readiness are multidimensional concepts that include a broader set of competencies than what many state assessments measure. Multiple formative and summative indicators are needed to adequately represent the spectrum of competencies that many analysts have identified as essential knowledge and skills for a postindustrial workforce.

2. **Composite indicators like A–F letter grades hide achievement differences within schools.** Variation is a natural aspect of teaching and learning. Students differ in their learning and development for numerous reasons. Rather than collapsing variance into categorical rating schemes, educators need to study patterns in variance across many different factors and conditions. Composite indicators actually inhibit the creation of meaningful and useful knowledge by combining multiple assessments into a composite indicator. For instance, we need evidence on achievement equity within and across schools, as well as accurate information about achievement growth for different groups of students. Such information gets lost when it is wrapped up with other metrics into one composite indicator. Given the importance of knowledge and skill transfer for success in life and the workforce, it would be useful for improvement purposes to know how students perform on tasks requiring different cognitive demands. Students may understand concepts and procedures in math—but can they apply quantitative thinking to new problems and situations?

   Related to the problem of hiding achievement differences is what social scientists refer to as an ecological fallacy. An ecological fallacy is the assumption that group indicators reflect the performance of individuals within the group. For example, an ecological fallacy occurs when we assume that individual students in an F school are failing. The distribution of scores in the previous graphs shows that this is not the case. Clearly, some students in F schools have low test scores, but assuming students in F schools are failing does not square with evidence showing many students excelling in their reading achievement. In fact, F schools were just as effective as A schools when accounting for the prior reading achievement of students. There is a great deal to learn about teaching and student progress from variation within schools.

3. **Composite indicators obscure achievement growth.** Paradoxically, student growth comprised 50% of the school grade in Oklahoma, yet the evidence we present suggests that achievement growth is not moving in the right direction. This happens because categorical data hide achievement changes that occur within proficiency bands. Points awarded for moving above proficiency, or remaining in the proficiency categories, mask actual changes in achievement from one year to the next. Letter grades do not report on the progress students below, or above, proficiency made from one year to the next. Further, grades do not report achievement growth by different student characteristics or across subjects. Trend data present better information for understanding actual changes in achievement over time.

4. **Test scores do not explain low or high performance.** We need good, comprehensive outcome data, but there is a limit to what outcome data reveal about weaknesses in the educational system. Low student performance results from numerous factors that are not reflected in aggregated test scores. In no other industry—banking, healthcare, manufacturing—would it be acceptable to make strategic decisions on outcome-only indicators. Yet accountability systems dependent on composite indicators expect policymakers and school leaders to diagnose the effectiveness of schools and the health of the educational system from a single summative outcome indicator.

Part Two: Next Generation Accountability

A New Goal: College and Career Readiness

First generation accountability policies, in compliance with the prescriptions of the federal NCLB law, exposed vast, inequitable differences in student test scores across and within schools. These policies were not effective at building capacity within school systems to raise achievement and equalize achievement distributions. As has been argued persuasively elsewhere, greater capacity lies at the heart of quality improvement (Darling-Hammond, 2005), and with the current emphasis on college and career readiness, it seems imperative that next generation accountability policies should align with this goal. College and career readiness is a significantly more challenging vision that sets our schools on a path well beyond test score proficiency. This new vision emphasizes the cognitive and noncognitive competencies expected for success in a postindustrial society and economy. What kind of accountability framework will facilitate progress toward this newly identified and specified vision? Three principles derived from the education policy and accountability literature provide guidance for the development of next generation accountability.

Principle One: Shared Accountability

Responsibility for school success is distributed.

The principle of shared accountability reminds us that in a complex enterprise such as public education, performance responsibility is distributed across the system's components, not foisted upon any one group of stakeholders. Control over each of the resources, processes, and/or outcomes essential for improvement resides almost exclusively within a particular stakeholder group; exerting pressure on stakeholders who do not have direct control over these elements is inappropriate and worse, harmful. Whereas the previous framework held schools alone responsible for student test scores, shared accountability recognizes a more fundamental set of student outcomes and identifies the critical contributions of the "whole village" to school effectiveness. So, for example, the governor and legislature are seen as responsible for creating a marketplace in which a state's schools can attract and retain their share of the most talented, appropriately trained, and competent teachers.

State political leaders are also responsible for ensuring that their state's educational system has sufficient funding for educational stakeholders to carry out their important work, while ensuring that this funding is equitably distributed across schools. State education departments are responsible for, among other things, identifying schools that have failed their communities catastrophically and providing supports so that children can realize their full potential. They must also provide the resources, expertise, and know-how to assist schools in achieving the readiness goal. School boards and superintendents share responsibility for maintaining a qualified, competent, and stable teaching corps in all schools, especially schools serving children from poverty. Principals and teachers are responsible for developing an instructional environment that meets the learning and psychological needs of students. Students and parents are responsible for collaborating and cooperating with teachers and school leaders in the pursuit of realistic but ambitious learning and life goals.

The goal of "shared accountability" is to create an accountability environment in which all participants recognize their obligations and commitments in relation to each other (Sullivan, 2009). In contrast with past accountability models, a shared accountability framework is designed

to render a comprehensive account of how each part of the educational system, and the system as a whole, is performing relative to the vision of college and career readiness. As Darling-Hammond and colleagues urge, "each level of the system should be held accountable for the contributions it must make to produce an effective system" (2014, p. 5).

Principle Two: Adaptive Improvement

Take schools where they are and move them forward.

First generation accountability assumed that districts and schools are alike in their capacity to become effective and attain the goals set by accountability policy. Accountability impact studies have documented that schools vary considerably in their capacity to respond coherently to the demands of external accountability (Shepard, Hannaway, & Baker, 2009). In contrast to high-performing schools that may only need to "redirect efforts" to improve, low-performing schools may lack resources and expertise to build from the ground up the kind of academic structures needed. Next generation accountability acknowledges that school capacities differ greatly, and that a one-dimensional continuum derived from student test results does not effectively describe school quality and capacity, explain performance, or identify targets for improvement. Embracing the concept of adaptive improvement necessitates a system that is flexible and responsive to particular school conditions, fitting interventions to the specific challenges.

Adaptive improvement acknowledges that schools are in different places on their paths to effectiveness, and that without essential resources and appropriate processes in place, schools will be unable to achieve even modest goals. A school lacking stable, quality leadership and teaching staff, for example, is at a different improvement stage than one whose teachers are experienced and have worked together successfully over time. It follows that schools have different information and resource needs, and their abilities to respond to and benefit from an accountability framework are different. Thus, a state's approach to working with schools is contingent on current assessed conditions at each school, and an individual school's particular need for support, expert assistance, and other resources.

Principle Three: Informational Significance

Put relevant and useful information in front of responsible actors.

What is sought in next generation accountability systems includes collecting and distributing significant information salient to the work and interests of all stakeholders—an ambitious goal. A broad palette of functionally significant indicators replacing a single composite indicator is likely to be regarded as informational rather than controlling, thereby motivating stakeholders to action (Ryan & Deci, 2001; Ryan & Weinstein, 2009). As Darling-Hammond and colleagues note:

> In a new system of accountability, multiple measures, coupled with thoughtful systems of judgment, should be used to inform decision making at each level. Transparency in providing information to the public and to educators and policymakers is a key aspect of the new accountability. Like businesses that use a dashboard of measures to provide a comprehensive picture of performance, we need a dashboard of indicators to inform key decisions … Full and timely reporting of a wide array of information to parents and community is a basic element of accountability (Darling-Hammond et al., 2014, p. 7).

The information system designed to service next generation accountability should recognize the dual reporting needs of compliance with federal mandates and the particular improvement needs of a state's schools. It should also address the different information needs of the state and district, school site leadership, teachers, and parents. Information about resources, processes, and a variety of outcomes can enable sensemaking by legislators who allocate school resources, by local school boards that develop school policy, by principals and teachers who design and adapt the instructional environment to fit local conditions, and by parents who want to make good choices for their children. However, it should be clear that the shape and "granularity" of information of interest to the state education department will be different from that of classroom teachers whose interests will be, by comparison, fine-grained. Next generation accountability focuses especially on shaping school improvement, and on movement toward goals of deeper learning and college and career readiness for all graduates, by making available relevant and useful information to those groups and individuals working to make schools effective.

A New Paradigm

Next generation accountability sets out to move beyond schooling "in which no child is left behind" to a system in which "children are healthy and thriving." This is a policy sea change—a focal shift from compliance with external mandates to strengthening schools and empowering the pursuit of standards for educational excellence. It shifts the task from test score gains to fostering an environment characterized by deeper learning, a condition foundational to the goal of college and career readiness (National Research Council, 2012). The framework casts a broad, inclusive net in recognizing those who share responsibility for building a state's capacity to achieve these lofty goals. It replaces a summative judgment of school performance rendered as a single indicator with multiple summative and formative indicators delivered as comprehensive school profiles. Adaptive interventions replace one-size-fits-all approaches, with the intent of assuring a high-quality, stable faculty for every school. The table below summarizes differences between these two accountability frameworks.

Table 1
Contrasting First and Next Generation Accountability

First Generation Accountability Framework	Next Generation Accountability Framework
No Child Left Behind	All children healthy and thriving
Focus on improving test scores	Focus on fostering deeper learning
Compliance to meet proficiency standards	Capacity building
Schools accountable to state	Shared accountability
Summative emphasis	Formative emphasis
Single composite indicator of student performance	Multiple indicators of whole system performance
One-size-fits-all interventions	Adaptive interventions
Replace teachers and leaders	Retain and support quality teachers and leaders

Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

A Two-Stage Framework

Next generation accountability has two primary stages and objectives. The second objective is embedded within the first, but it has a unique and precise function.

- **Objective one** is intended to provide all state schools with useful, longitudinal information profiles that monitor and inform each school's journey toward deeper learning, and the goal of universal college and career readiness. The framework joins inextricably the provision of significant information to an enhanced support infrastructure by matching resources, expertise, and other supports to the developmental needs of schools and districts.

- **Objective two** is narrowly focused on the identification of schools in need of fundamental change to disrupt enduring patterns of failure, and managing a mandatory process for infusing resources, expertise, and extended support as needed to see changes through to success.

Objective One: Supporting Improvement for All Schools

In next generation accountability systems, the central goals of deeper learning and college and career readiness are the foundation of this formative stage and constitute the premier work focus of state departments of education. This stage has states providing every school and school district with timely, longitudinal information about the quality of school resources, processes, and outcomes. It requires newly conceptualized and designed school profiles that can inform education stakeholders about the performance of student groups over time on measures of deeper learning and college and career readiness. It is essential that these profiles gauge key school and community resources and processes, thereby enabling explanations for school outcomes. Systemwide indicators spotlight weaknesses as well as suggest responsibility for action. Stage one will undoubtedly require state departments of education to broaden the scope of existing school support, and create infrastructure that expands and enhances the expert support teams assisting schools. As designed, then, a school's progress toward the foundational goals of deeper learning and college and career readiness will be the overarching criterion against which schools will judge their own performance and plan their development.

Thoughtful design of school profiles will empower citizens and school professionals alike to discern a school's strengths and weaknesses, and at the same time discourage the simplistic interpretation of school outcome measures. Transparency in reporting school conditions and effectiveness is achieved by displaying multiple indicators of significant information that can be understood and interpreted by all school stakeholders. In concert with state information-gathering processes, schools and districts will be encouraged to collect additional information of their choosing that is relevant to their particular needs for explaining and planning school improvement. Under Stage One, local schools are asked to take ownership of the school improvement process, including understanding their own challenges and managing their unique improvement journey. This approach stipulates that the state department of education is on call to provide prompt and expert assistance and support when it is requested by any school.

Objective Two: Identifying Schools for Intensive Assistance

The second stage of next generation accountability is designed to identify schools whose performance lags seriously over time and whose efforts appear unable to reverse a trajectory of failure. This stage is responsive to ESSA guidelines calling for the *identification of schools in need of comprehensive support and improvement.* It also is the basis for reporting objective measures of student outcomes in clear, concise, and easily understood indicators so that educators and

the public alike can accurately judge student progress toward deeper learning and college- and career-ready standards. Data reporting will comply with criteria established by education state departments for reporting school progress toward Annual Measurable Objectives (AMOs) for all students and student subgroups.

The Stage Two identification process is set in motion by a longitudinal record of underperformance on key student outcomes, together with a pattern of resource instability or inadequacy. Departments of education will oversee an investigatory process that includes participation by community and school representatives. The rich profile of longitudinal system indicators reported to schools for Stage One simultaneously provides significant information needed by state departments to make informed determinations of catastrophic failure. The profile of system resources, process, and outcomes made available to failing schools provides the basis for designing appropriate interventions fitted to the specific problems and needs of the school. In cases of serious malfunction, the objective of the intervention is to take prompt action of sufficient scope and intensity to reset the school on a path to heightened performance. This process requires coherence and persistence, two conditions often in short supply in school reform (Bryk et al., 2015; Peurach & Neumerski, 2015).
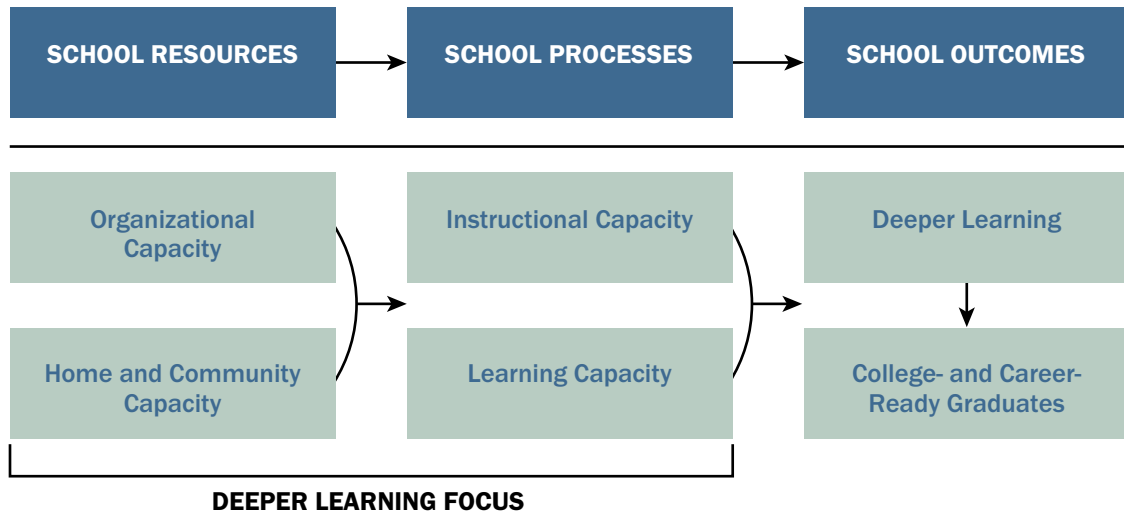
Next Generation Design

How will next generation accountability change the mindset and approach of stakeholders to address school failure and stimulate a more focused commitment to educational excellence? Shifting from a preoccupation with failure to success is not merely semantic; instead, the system and its momentum are directed away from deficits and toward assets in the form of school capacities that enable deeper learning and ultimately college and career readiness. A focus on capacity building simultaneously reveals the importance of resources and processes, as it sorts out responsibility for marshaling these in the service of school improvement (Forsyth & Tallerico, 1998). It is a vision of thriving schools rather than a vision of failure avoidance.

Figure 7 is a conceptual map of school function. The systems model depicts outcomes as a consequence of the quality of resources, together with the quality of school processes. Resource and process elements are defined as a set of critical capacities that ultimately focus instruction on the knowledge, skills, and dispositions that ready students for the workforce or advanced learning. When school outcomes are deemed unsatisfactory, the model and its constituent capacities enable stakeholders to identify, explain, and target needed change. Thus, while deeper learning and college and career readiness are ultimate goals, they also provide direction for capacity building throughout the system. As a result, policymakers and local school professionals (indeed all public education stakeholders) have a more precise understanding of what needs to be done. An accountability system "must attend to the inputs, processes, and outcomes that produce student learning: In other words, it must build capacity to offer high-quality education, while holding educators accountable for providing such education" (Darling-Hammond et al., 2014, p. 6).

If a fundamental goal for the public school system is college and career readiness, the logic of a systems approach urges that school resources and processes align with the deeper learning needed to participate effectively in the emerging economy. The commitment to a goal of college- and career-ready graduates requires embracing a public education system characterized by deeper learning. It means that, at every level and for all decision makers, criteria grounded in this commitment will shape decisions about the resources and processes responsible for school performance.

Figure 7
Key Concepts in Next Generation Accountability

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education*.

Table 2
Key Concepts in Next Generation Accountability

Resources	Organizational Capacity	School resources that are foundational to the development and maintenance of high-quality teaching and learning.
	Home/Community Capacity	The social and material supports that surround children, and describe relative opportunity that varies across families and communities.
Processes	Instructional Capacity	The ability and readiness of the school's teaching corps to design and deliver appropriate, challenging, and goal-related instruction to all students.
	Learning Capacity	The ability and readiness of the school's student cohorts to engage in the work of mastering the knowledge, skills, and dispositions necessary for college and career readiness.
Outcomes	Deeper Learning	The ability to transfer knowledge and skills developed in one setting to new situations, contexts, and problems through a set of cognitive, intrapersonal, and interpersonal competencies.
	College and Career Readiness	The preparation of high school graduates to enter a career, extended training, or specialized education without need for remediation.

Defining the Concepts

In this section, we define the concepts included in the accountability framework before discussing in detail each concept's proposed function. An abbreviated description of each concept may be found in Table 2.

Organizational Capacity

Organizational capacity is the first of two resource capacities included in the framework. It is defined as school resources that are foundational to the development and maintenance of high-quality teaching and learning. The stability and quality of a school's faculty is an example of organizational capacity whose importance is well documented (Ronfeldt, Loeb, & Wyckoff, 2013; Hanushek, Kain, O'Brien & Rivkin, 2005; Ingersoll, 2001). It seems necessary that organizational capacities be established at some minimal threshold before process capacities can effectively be developed and deployed. Thus, for example, some level of teacher corps stability must be achieved before a school can enact an effective instructional program matched to the needs of its students. Without a critical level of teacher corps stability, the shared organizational learning, social capital, specialized skills, and understandings about what works here are lost each year, requiring the rebuilding of the school's functional infrastructure.

Responsibility for the adequacy of organizational capacity does not reside solely in the school and its staff. Building these capacities clearly requires the active support and cooperation of state leaders, the local civic community, the local school board, and the superintendent, among others. Indicators of capacity and performance at the state and local levels should help determine if the threshold is met, and whether or not a school is ready to enact a sustaining and effective instructional program directed at college and career readiness.

Home and Community Capacity

This capacity is composed of the social and material support that surrounds children, and describes relative opportunity that varies across families and communities. The importance of this capacity for school success and the life chances of children is documented by overwhelming scientific evidence: " … all educational efforts that focus on classrooms and schools … could be reversed by family, could be negated by neighborhoods, and might well be subverted or minimized by what happens to children outside of school" (Berliner, 2006). Generally, schools have not excelled at developing this capacity, especially in high-poverty communities. Nevertheless, these environmental factors and resources, although often not under the school's control, need to be considered for their potential moderating influences. Therefore, a framework for school effectiveness and accountability ought to attend to, measure, and incorporate these factors (Hopson, 2014; Horsford & Sampson, 2014). Understanding family, neighborhood, and community capacity is central to sound educational policy and is a key to school improvement.

Instructional Capacity

Instructional capacity is defined as the ability and readiness of the school's teaching corps to design and deliver appropriate, challenging, and goal-related instruction to all students. Instructional capacity is concerned with proximal features of the teaching corps directly relevant to instruction, such as a school's affective climate, a coherent instructional program, and the health of principal-faculty relations. Equally important, instructional capacity also explores the levels of teacher

opportunity for instructional development, collaboration, and various supports to foster deeper learning and instructional vitality. Deeper learning will be the consequence of "deeper learning opportunities"; rising to this challenge is the work of the teaching staff in every school.

Learning Capacity

Learning capacity is defined as the ability and readiness of the school's student cohorts to engage in the work of mastering the knowledge, skills, and dispositions necessary for college and career readiness. This view of students is concerned with conditions and characteristics that are malleable within the school's process. Understanding the current quality of a school's learning capacity is essential for knowing how to ignite student motivation, curiosity, and engagement. Indicators of student affect and perception such as the teacher-student trust relationship are key, as is knowledge about the extent to which the school promotes a self-regulatory climate (Adams, Forsyth, Dollarhide, et al., 2015; Schunk & Zimmerman, 2008). Armed with these understandings, the school's staff can work to enhance the learning capacity of students, while adjusting the instructional system to meet them where they are.

Deeper Learning and College and Career Readiness

We consider two aspects of deeper learning. The first is the opportunities provided by the school for students to transfer knowledge and skills developed in one setting to new situations, contexts, and problems (National Research Council, 2012). The second is the deeper learning outcomes required for students to become college and career ready. Outcomes consist of cognitive, intrapersonal, and interpersonal competencies that individuals draw on. The three competency domains are foundational, interconnected, and connected as well to human development and growth. They also reflect malleable aptitudes, traits, and characteristics that are influenced by an individual's social surroundings and experiences.

The cognitive domain accounts for thinking and related cognitive processes like reasoning, synthesizing, evaluating, problem solving, encoding, and retrieving information. Both content knowledge and procedural knowledge (how to apply knowledge) associated with different disciplines fall within the cognitive domain. So, too, do more advanced cognitive processes and strategies such as synthesizing and evaluating information, analyzing data, reasoning and argumentation, drawing conclusions, and problem identification and problem solving. The intrapersonal domain includes malleable emotions, feelings, and personality traits that activate human behavior and lead to high adaptive functioning. Intrapersonal competencies are embodied in individual characteristics and mindsets that align with the continua of the big five personality traits of conscientiousness, openness to experience, extraversion, agreeableness, and neuroticism. The interpersonal domain includes malleable social competencies that facilitate both knowledge acquisition and transfer. Characteristics include, among others, the abilities to communicate and collaborate, trust, empathy, conflict management, leadership, social awareness, and leadership.

In a discussion of potential "noncognitive" outcomes of deeper learning, the warnings of *Campbell's Law* again bear repeating. At no time should such informational measures be used in high-stakes or otherwise summative decision making—to do so completely undermines their validity as measures of social processes (Duckworth & Yeager, 2015). A recent news report in *The New York Times* about the "testing of grit" in some California schools (Zernike, 2016) is a prime example of how easy it is for stakeholders to move from choosing an important indicator of student performance to using

such a measure inappropriately for accountability purposes. The tying of significant consequences or rewards to social-emotional skills is particularly dangerous because the most efficient way to capture these skills is through self-report survey; direct observation of desired behaviors is possible, but often too costly in terms of resources. If students and educators perceive that substantial consequences are tied to these skills, *Campbell's Law* predicts that they will alter their reports to better conform to desired expectations—likely without altering underlying behaviors or mindsets. This outcome is fundamentally different from one in which stakeholders' responses reflect genuine, autonomous effort to change these behaviors and/or mindsets.

Nationally, school accountability has shifted its sights to the dual goals of deeper learning and college and career readiness. The ultimate indisputable indicators of college and career readiness are admission to and progress in advanced education or appropriate and sustained employment. Such indicators need to report the degree to which students do not simply graduate from high school, but do so with the knowledge and skills for an effective transition to postsecondary opportunities. Exit indicators such as graduation rates, among others, are the final evidence warranting claims of college and career readiness.

Part Two: Conclusion

Building on conceptual inadequacies and technical difficulties found in current school accountability approaches, Part Two sought to justify a paradigm shift by introducing a set of goals much more ambitious than raising test scores. What is identified as next generation accountability commits states and public education systems to instruction that produces deeper learning and high school graduation standards that are aligned with the emerging world economy and readying graduates for advanced training or careers:

Three principles are derived from the accountability, organizational, and school reform literature to guide the design of a system focused on these goals.

1. **Shared Accountability** emphasizes the recognition that responsibility for school improvement is distributed widely across a state's institutions, professionals, and citizens.

2. **Adaptive Improvement** recognizes that any accountability system should account for differences in the capacities of schools to pursue standards reflective of 21st century skills and competencies.

3. **Informational Significance** points to the relative merits of information that informs knowledge of practice more than it seeks to control outcomes.

A two-stage systems framework is proposed as fitting the design needs for next generation school accountability. In Stage One, the design places emphasis on shepherding all public schools toward the dual goals referred to above. This shift requires a process that is more formative than summative, whereas Stage Two is aggressive, and takes seriously the need to identify and transform schools that are in catastrophic failure. Next generation accountability requires a significant reallocation of state resources , putting in place a support infrastructure capable of serving the developmental and resource needs of all schools.

Both stages depend on a view of schools as systems whose outcomes result from the quality of their resources and processes. Primary resources are conceptualized as a school's organizational and home/community capacities. Essential school processes are conceptualized as instructional and learning capacities. Indicators of these capacities, we argue, will empower stakeholders at all levels, from the legislature to parents, to understand the meaning of school outcomes, and to design and implement effective interventions. A set of summative indicators, together with capacity indicators, will enable each state to identify with confidence those schools in need of expert support and additional resources.

Part Three: An Educational Quality and Improvement Profile

Consistent with the information needs of next generation accountability, we have developed and propose for consideration an Educational Quality and Improvement Profile (EQuIP). EQuIP is envisioned as both an accountability and a school improvement tool. It provides accountability indicators of student performance, as well as resource and process indicators that are related to school outcomes. On the accountability side, EQuIP establishes a level of transparency in reporting summative student outcomes far exceeding a composite summative index. It presents accurate and clear accountability information on deeper learning and college and career readiness, tracks changes in achievement gaps, and reports student growth. On the improvement side, it points to formative conditions and resources in schools that need to change for observed outcomes to improve.

EQuIPs benefit every school. Consistently low-performing schools desperately need nuanced data to inform deliberate strategies and focused support for improvement. Even the highest-performing schools cannot afford to be static since each group of learners has quite different needs. Thus, EQuIP is envisioned as a comprehensive profile with multiple uses, depending on the specific needs of each school community. All stakeholders can find useful information in EQuIPs. Parents seeking their children's greatest well-being, policymakers seeking efficiency and effectiveness, and educators seeking to create and sustain thriving schools will find accurate and useful information about different aspects of the life and health of schools.

EQuIPs are committed to the following measurement guidelines:

1. Resource, process, and outcome indicators should report changes over time.

2. Indicators should be assessed with appropriate frequency and minimal disruption.

3. Credible scientific evidence substantiating the validity and reliability of measures should be clearly reported.

4. Resource and process indicators should measure conditions, attitudes, structures, and behaviors.

5. Outcome indicators should report achievement differences by student subgroups.

6. Outcome indicators should enable the identification of focus schools, priority schools, and reward schools consistent with criteria set for federal waiver requirements.

7. Indicators and measurement methods should change over time in response to the continuous evaluation of the school accountability framework.

8. Indicators should be reported and used to avoid gaming practices and distortion of school performance.

EQuIP begins with formative resource and process indicators to focus attention on the system's capacity to deliver high-quality learning opportunities. Resource and process indicators reflect improvement levers; they make known the conditions that are antecedent to observed outcomes.

Summative outcome indicators round out the profile so that educators and the public can judge how well schools are moving students toward proficiency in deeper learning and college and career ready standards.

Following the lead of improvement science in health care and education, the objective for using formative resource and process indicators is to report on a small number of high-leverage conditions that guide purposeful action toward future goals (Bryk et al., 2015). We expect resource and process indicators used in EQuIPs to change as conditions and needs in schools evolve, but for now, we see potential value in the indicators described below. These indicators align with deeper learning and college and career readiness; they respond to pressing problems of low teacher morale and increased student boredom, alienation, and disengagement (Fullan, 2015); and they call attention to persistent disparities between economically advantaged and disadvantaged communities (Putnam, 2015).

Resource Indicators

The appearance of a school and the quality of its facilities often come to mind when thinking of school resources. However, the most valuable resources of schools are not those seen from the parking lot or when walking the halls of an unoccupied building (Grubb, 2009). A school should not be perceived as a building—a school is a community of people. The essential function of schools—teaching and learning—depends not on the color of the school walls, the size of the football stadium, or the number of classrooms with SMART boards, but on the educators and students within. People are an indispensable resource (Forsyth, Adams, & Hoy, 2011). That is not to say that school appearance is not important; it certainly matters, but only to the extent that it affects the attitudes, mindsets, and opportunities of the school community. Therefore, the health of schools is manifest in the knowledge, motivation, and relationships of school stakeholders. For this reason, indicators of school resources reported in EQuIPs are people-oriented, and classified as organizational capacity and home and community capacity.

Organizational Capacity

Figure 8 displays an example of an EQuIP page with organizational capacity indicators. Organizational capacity calls attention to the essential function of human resources within schools. Quality schools invest in and retain highly effective educators, enabling them to deliver valuable learning opportunities to the children and families they serve. A highly effective and stable teaching corps, along with stable leadership, is especially critical for communities of economic disadvantage. Thus, EQuIP reports organizational capacity with human resource indicators. **Teacher Stability** is measured by the percentage of teachers returning to the school (tracked longitudinally), the percentage of teachers who remain in a school for a 5-year period, the distribution of teachers by experience, and a teacher substitution ratio (total days covered by subs/[total school FTE x total days]). **Leadership Stability** is the years of principal experience at the current school, experience and longevity of assistant principals in the school (if applicable), and number of principals in a 10-year period. Information is also reported on the number of instructional coaches, teacher-to-student ratio, counselor-to-student ratio, Special Education (SPED) teacher-to-student ratio, English as a Second Language (ESL) teacher FTE-to-English Language Learner (ELL) ratio, average class size, and largest and smallest class size.

Figure 8
An Example of an EQuIP Page for Organizational Capacity

ORGANIZATIONAL CAPACITY

SAMPLE ELEMENTARY SCHOOL

Organizational Capacity calls attention to the essential function of human resources in schools. Indicators of this capacity include the stability of the teacher corps, the stability of school leadership, and descriptive data about class sizes and the composition of support staff relative to the number of students in attendance.
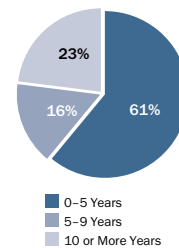
COMPOSITION OF SCHOOL

Number of Students..**594**

Gender

  Male: 47%   Female: 53%

Ethnicity

  Asian: 3%   Black: 27%

  Hispanic: 38%   White: 22%

  American Indian: 3%   Multiple Races: 7%

English Language Learners...**17%**

Students with Special Needs......................................**26%**

Free or Reduced-Price Lunch Rate **68%**

TEACHER STABILITY

The percentage of teachers who remain in a school annually, the percentage of teachers who remain in school over a 5-year period, and the distribution of teachers by years of experience.

64% of teachers have been in the school for at least 5 years
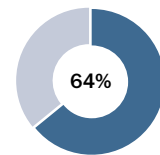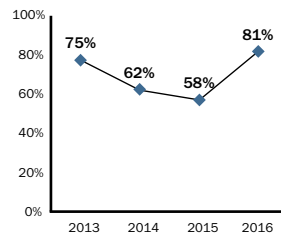
DISTRIBUTION OF TEACHING EXPERIENCE

- 0–5 Years
- 5–9 Years
- 10 or More Years

23%, 16%, 61%

LEADERSHIP STABILITY

CURRENT PRINCIPAL YEARS IN SCHOOL	**8**
NUMBER OF PRINCIPALS IN LAST 10 YEARS	**3**
AVERAGE ASSISTANT PRINCIPAL YEARS IN SCHOOL	**5**

SUPPORT STAFF

FULL-TIME INSTRUCTIONAL COACHES	**2**
COUNSELOR TO STUDENT RATIO	**1:150**
SPED TEACHER TO SPED STUDENT RATIO	**1:15**
ESL TEACHER FTE TO ELL RATIO	**1:18**

TEACHER ATTENDANCE

The average teacher attendance rate and a substitute teacher ratio (total days covered by subs)/(total school full-time equivalent allocations x total days).

TEACHER ATTENDANCE RATE	**97%**
SUBSTITUTION RATE	**6%**

TEACHER ATTENDANCE

SMALLEST CLASS SIZE	**8**
AVERAGE CLASS SIZE	**25**
LARGEST CLASS SIZE	**38**
TEACHER TO STUDENT RATIO	**1:21**

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

Home and Community Capacity

Figure 9 displays an example of an EQuIP page with home and community capacity indicators. Home and community capacity addresses the relational connections between schools and families as well as household and community resources available to children. **Parent Trust in School** measures the quality of relationships between teachers and parents. Parents are asked about teacher academic standards for all students, teacher concern for students, teacher communication with parents, teacher competence, teacher honesty, and teacher reliability in actions and commitments. Higher parent trust suggests that parents perceive teachers as being open, honest, reliable, competent, and benevolent. **School Outreach** measures the pattern of communication and interactions between parents and school authorities. Parents are asked about how well the school communicates information, parent opportunities to provide feedback to school authorities, and parent feelings of belonging in the school community. Higher perceived school outreach suggests that parents perceive school-parent communication and interactions as open and supportive.

Food Insecurity measures the percentage of children in the school from households with an inadequate food supply as a result of insufficient funds or resources. Using questions from the CPS Food Security Survey, parents are asked if anyone in their household skipped meals, cut the size of meals, ran out of food, or lost weight because there wasn't enough money for food within the last year (Coleman-Jensen, Gregory, & Singh, 2014). Affirmative responses to three or more questions indicate households where students experienced food insecurity. **Number of Parents in the Home** measures the percentage of children in the school who do not live with both parents in the home. Growing up in a single-parent family is linked to a range of negative outcomes in school and later in life (Sigle-Rushton & McLanahan, 2004). A higher percentage of children living with both parents tends to suggest increased supervision for children, and increased access to material, human, and social capital. **Community Relational Support** measures student reports of quality relationships with adults outside their homes or schools. Students thrive under the supervision of adults in the community who take special notice of them and believe in their potential. Students without community connections are more likely to feel socially isolated, and to experience bullying and depression (Murphey, Bandy, Schmitz, & Moore, 2013).

Figure 9
An Example of an EQuIP Page for Home and Community Capacity

HOME AND COMMUNITY CAPACITY

SAMPLE ELEMENTARY SCHOOL

Home and Community Capacity addresses the relational connections between schools and families as well as household and community resources available to children. Indicators include parent trust in the school, parent perceptions of school outreach, as well as student access to material, human, and social capital.

Strong Capacity
Moderate Capacity
Low Capacity

2013   2014   2015   2016
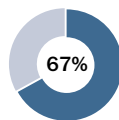
■ Parent Trust   ■ School Outreach
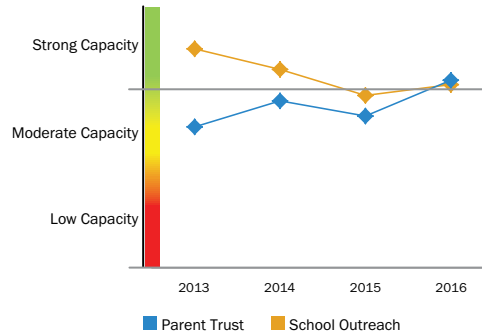
PARENT TRUST

Parent perceptions of teacher academic standards for all students, teacher concern for students, teacher communication with parents, teacher competence, teacher honesty, and teacher reliability in their actions and commitments.
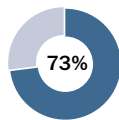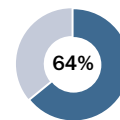
73% of parents believe the school has high standards for all students

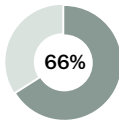**67%** of parents believe teachers in the school are always ready to help.
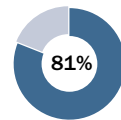
81% of parents believe students are well cared for while at school.
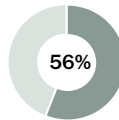
**64%** of parents believe the school does a terrific job.

SCHOOL OUTREACH

Parent perceptions of their sense of belonging in the school community, their opportunities to provide authentic feedback to school authorities, and the ability of the school to keep them informed of school happenings.
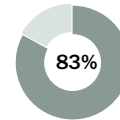
66% of parents believe teachers communicate regularly.

56% of parents believe the school works closely with parents to meet student needs.
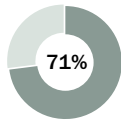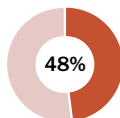
71% of parents believe the school regularly communicates with parents about how they can help.

83% of parents believe the school encourages parents to give feedback.

FOOD INSECURITY

48%

The percent of children in the school from households with an inadequate food supply as a result of insufficient funds or resources.

NUMBER OF PARENTS IN THE HOME
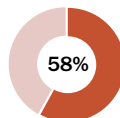
58%

The percent of children in the school who do not live with both parents in the home.

COMMUNITY RELATIONAL SUPPORT
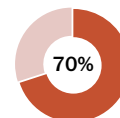
70%

The percent of children in the school who feel they have at least one quality relationship with an adult outside the home or school.

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

Process Indicators

After detailing resource indicators that take stock of school actors and the context within which they operate, EQuIP reports on structures, conditions, and behaviors that help maximize teaching and learning. Strong, effective school processes precede desired outcomes; they are the mechanisms through which teacher and student needs are supported. Self-Determination Theory posits that all humans have three innate psychological needs that, when satisfied, underlie intrinsic motivation and holistic well-being (Deci & Ryan, 2016). The school environment can either support or thwart need satisfaction for both teachers and students (Adams, Forsyth, Dollarhide, et al., 2015). Supportive, or nurturing, conditions underlie peak performance (Ryan & Deci, 2002). EQuIPs report process indicators in terms of instructional capacity (school structures and conditions that support teacher psychological needs) and learning capacity (school structures and conditions that support student psychological needs).

Instructional Capacity

Figure 10 displays an example of an EQuIP page with instructional capacity indicators. Instructional capacity exists in resources and social processes that enable teachers, individually and collectively, to reflect on and learn from their instructional practices. Useful indicators capture the degree to which a school is organized to activate the inner determination of teachers to learn and grow as professionals. Such conditions include teacher perceptions of professional learning opportunities, faculty trust, and the coherence of the instructional program.

Professional Learning reports the degree to which teachers experience the school environment as supporting them as professionals and enriching their development as expert practitioners. We propose the professional development opportunities scale (Rowan & Miller, 2009) as an effective measure to capture the quality of formal and informal learning opportunities available to teachers. **Faculty Trust** accounts for the quality of relationships within a teaching faculty, and among teachers and school leaders. Trust is both the glue that unites school members toward a shared vision and the lubricant that facilitates collective problem solving. The Omnibus Trust Scale (Forsyth, Adams, & Hoy, 2011) measures the strength of relational connections among teachers, and between teachers and the principal. A **Coherent Instructional Program** reflects a school that is organized around a shared and coherent instructional approach, has a vertically and horizontally aligned curriculum, and acts to make teaching and learning measurably better. We propose the instructional program coherence scale from The University of Chicago Consortium on Chicago School Research.
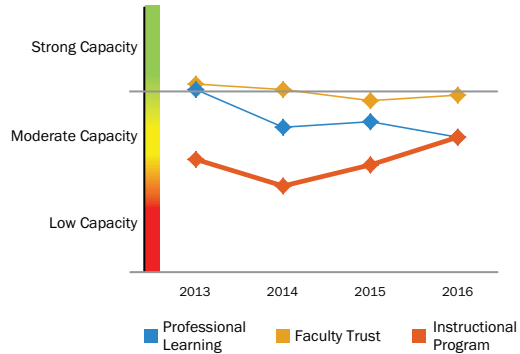
Figure 10
An Example of an EQuIP Page for Instructional Capacity

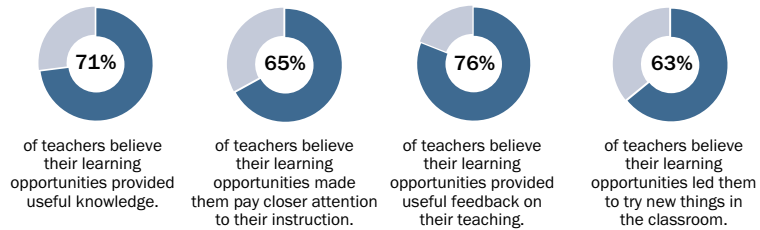## INSTRUCTIONAL CAPACITY

SAMPLE ELEMENTARY SCHOOL

Instructional Capacity refers to the aggregate ability and readiness of the school's teaching corps to design and deliver appropriate, challenging, and goal-related instruction to all students. Teachers' effectiveness depends on the learning opportunities they are provided, and the coherence of school curriculum within and across grades.

Strong Capacity

Moderate Capacity

Low Capacity

2013   2014   2015   2016

Professional Learning   Faculty Trust   Instructional Program
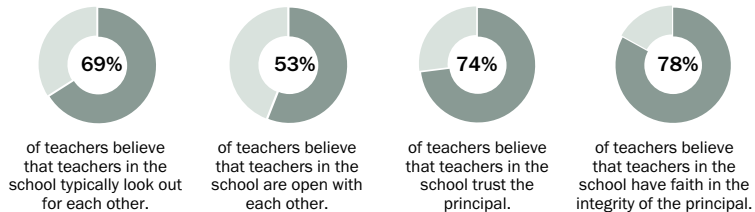
PROFESSIONAL LEARNING

The degree to which teachers believe the school environment supports them as professionals and enriches their development as expert practitioners.
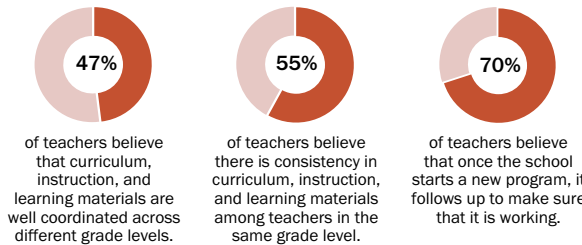
71% of teachers believe their learning opportunities provided useful knowledge.

65% of teachers believe their learning opportunities made them pay closer attention to their instruction.

76% of teachers believe their learning opportunities provided useful feedback on their teaching.

63% of teachers believe their learning opportunities led them to try new things in the classroom.

FACULTY TRUST

The quality of relationships within a teaching faculty, and among teachers and school leaders.

69% of teachers believe that teachers in the school typically look out for each other.

53% of teachers believe that teachers in the school are open with each other.

74% of teachers believe that teachers in the school trust the principal.

78% of teachers believe that teachers in the school have faith in the integrity of the principal.

COHERENT INSTRUCTIONAL PROGRAM

A school that is organized around a shared and coherent instructional approach has a vertically and horizontally aligned curriculum, and takes continual action to make teaching and learning measurably better.

47% of teachers believe that curriculum, instruction, and learning materials are well coordinated across different grade levels.

55% of teachers believe there is consistency in curriculum, instruction, and learning materials among teachers in the same grade level.

70% of teachers believe that once the school starts a new program, it follows up to make sure that it is working.

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

Learning Capacity

Figure 11 displays an example of an EQuIP page with learning capacity indicators. Learning capacity in a school environment is defined by student trust in their teachers, motivating instruction, and a positive student social climate. These conditions activate student interest, curiosity, and determination to persist in academic pursuits (Adams, Forsyth, Dollarhide, Miskell, et al., 2015). High levels of learning capacity do not guarantee positive educational outcomes, but it is hard to imagine deeper learning flourishing without these conditions.

Trust indicators account for the quality of relationships among students and teachers. Trusting relationships reflect a learning climate that brings out the best attitudes and behaviors in students, enabling them to persevere toward the challenging expectations of deeper learning (Adams, 2014). Trust indicators should be measured with data derived from a de-identified survey, and results should be aggregated to the school level. We propose using items from the Student Trust in Teachers Scale and Omnibus Trust Scale (Forsyth, Adams, & Hoy, 2011). Indicators of **Motivating Instruction** account for classroom contexts that students experience as supporting their psychological needs of autonomy and competence. Student-perceived autonomy support and competence support have been demonstrated in experimental and correlational studies to be related to persistence in school, higher interest in academic tasks, creativity and expression, and higher achievement (Adams, Forsyth, Dollarhide, et al., 2015; Niemic & Ryan, 2009; Kusurkar, Cate, Vos, Westers, & Croiset, 2012). Like trust, need support is best measured with a de-identified student survey with results aggregated to the school level. We propose items from the Autonomy-Enhancement Scale (Assor, Kaplan, & Roth, 2002) and the Academic Press Scale from The University of Chicago Consortium on Chicago School Research. **Student Social Climate** reports the strength of peer relationships. Peer relationships are related to classroom achievement and to student psychological health (Wentzel & Caldwell, 1997). We recommend items from the Student Trust in Students Survey and a general bullying survey (Forsyth, Adams, & Hoy, 2011). Again, these are best measured by a de-identified student survey with results aggregated to the school level.

Outcome Indicators

Summative outcome indicators conform to ESSA requirements for reporting school progress toward proficiency goals of all students and student subgroups, as well as high school graduation rates and progress of ELL students. Assuming states have developed a cognitively rigorous assessment system, EQuIPs lay out a comprehensive view of school progress toward deeper learning by reporting student performance, student growth, and achievement gaps in concise and easily understood descriptive graphs. Raw, descriptive data do not distort student outcomes by combining multiple data points into an index whose meaning is opaque (Adams et al., 2013). A rich profile of cognitive and noncognitive indicators enables educators, policymakers, and the public to determine definitively if a school's students are achieving proficiency standards.
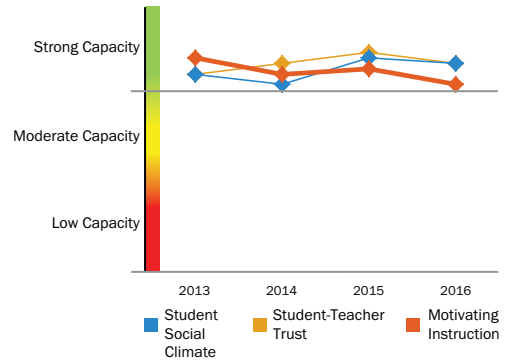
Figure 11
An Example of an EQuIP Page for Learning Capacity

LEARNING CAPACITY

SAMPLE ELEMENTARY SCHOOL

Learning Capacity refers to the quality of the learning environment, and its ability to support student psychological, emotional, and physical needs to activate their interest, curiosity, and determination to persist. Satisfying these needs allows students to focus, grow, and thrive.


Strong Capacity

Moderate Capacity

Low Capacity

2013　2014　2015　2016

■ Student Social Climate　■ Student-Teacher Trust　■ Motivating Instruction
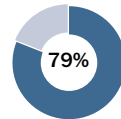

STUDENT SOCIAL CLIMATE

The quality of relational connections among students demonstrated by their trust of one another and the frequency with which they witness instances of bullying.
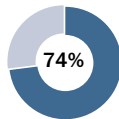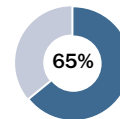
74% of students believe their peers are honest.
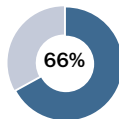
66% of students believe their peers are willing to help each other out.

79% of students say rumors have been spread about them in the school.

65% of students say they have witnessed school peers being threatened or physically hurt.


STUDENT-TEACHER TRUST

The quality of relational connections between students and teachers as measured by students' perceptions of the openness, honesty, reliability, benevolence, and competence of their teacher.
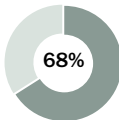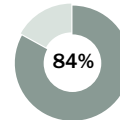
68% of students believe teachers in the school are honest with their students.
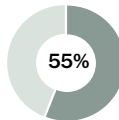
55% of students believe teachers in the school care about their students.
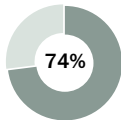
74% of teachers believe students in the school can be counted on to do their work.

84% of teachers believe students in the school care about each other.

MOTIVATING INSTRUCTION

The degree to which students experience instructional practices that support their autonomous motivation for learning and for achieving high academic standards.
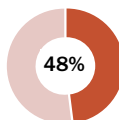
48% of students believe teachers show students how to solve problems themselves.

57% of students believe talk about the connection between what is studied in school and what happens in real life.
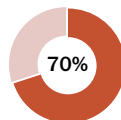
70% of students believe teachers in the school expect students to work hard.
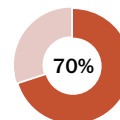
70% of students believe teachers in the school challenge students to achieve academic goals.

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

Cognitive Outcome Indicators: Elementary and Middle Schools

Figure 12 displays an example of an EQuIP page reporting cognitive outcome indicators for an elementary/middle school. Notice how **student achievement** on state exams is presented in easy-to-understand bar graphs that show the percentage of students scoring in each proficiency category.[1] Contrast this with A–F letter grades and other summative indexes utilized by states: A–F letter grades hide this information in a formula that masks the actual proficiency distribution of students. With EQuIPs, educators and the public know the actual percentage of students scoring advanced, proficient, limited knowledge, or unsatisfactory. Additionally, EQuIP provides an AMO target for 2020 so as to benchmark performance against a long-term goal.

Pie graphs are used to display data on **student growth**. For elementary schools, graphs report the changes in reading and math proficiency scores for a cohort of students from 3rd grade to 5th grade. The graphs describe the percentage of students who remained in proficiency, dropped out of proficiency, rose to proficiency, or never met proficiency. In middle schools, growth reflects changes in proficiency categories from 6th grade to 8th grade. Calculations can be easily adjusted to fit different grade spans for elementary and middle/junior high schools. With A–F letter grades (and other summative indicators), it is not clear what proportion of students moved across proficiency bands in either direction over time. Formulas for index scores hide this information from the public. In contrast, EQuIPs make known achievement changes, enabling discernments about the progress of lower- and higher-performing students alike.

Line graphs are used to track 5-year trends in **achievement gaps**. These graphs display changes in average achievement for subgroups with 30 or more students. This allows schools and the public to longitudinally track gains or losses in reading and math achievement for student subgroups and to determine if any observed achievement gaps are closing, widening, or remaining flat. With A–F letter grades (and other summative indicators), it is not clear if achievement gaps are closing, stagnating, or widening, leaving educators and the public unaware of achievement equity trends in schools.

---

1. With proficiency scores, it is important for the state to establish high and stable thresholds, and to make clear the criteria, methods, and rationale used to set, and when necessary change, cut scores.

Figure 12
An Example of an EQuIP Page Reporting Cognitive Outcome Indicators for an Elementary/Middle School

DEEPER LEARNING: COGNITIVE

SAMPLE ELEMENTARY SCHOOL

Deeper Learning refers to the expectation that schools develop in students cognitive and noncognitive competencies needed for effective participation in the workforce and active citizenship. Initially, proficiency rates, changes in proficiency rates over time, and achievement gaps serve as indicators of the cognitive component of deeper learning.

COMPOSITION OF SCHOOL

Number of Students...594

Gender
Male: 47%   Female: 53%

Ethnicity
Asian: 3%   Black: 27%
Hispanic: 38%   White: 22%
American Indian: 3%   Multiple Races: 7%

English Language Learners..17%

Students with Special Needs.....................................26%

Free or Reduced-Price Lunch Rate 68%

STUDENT PERFORMANCE

Bar graphs indicate the distribution of student achievement by proficiency status: advanced, proficient, limited knowledge, and unsatisfactory. The AMO target represents the goal of reducing by 50% the number of students scoring below proficiency by 2020.
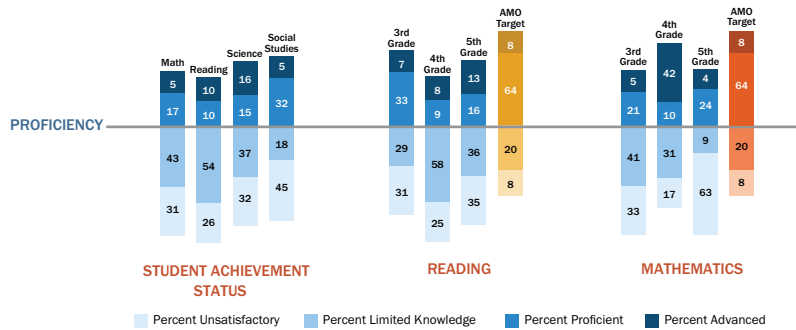
CHANGES OVER TIME

Pie graphs indicate the change in reading and math proficiency status for a cohort of students from 3rd grade in 2014 to 5th grade in 2016.
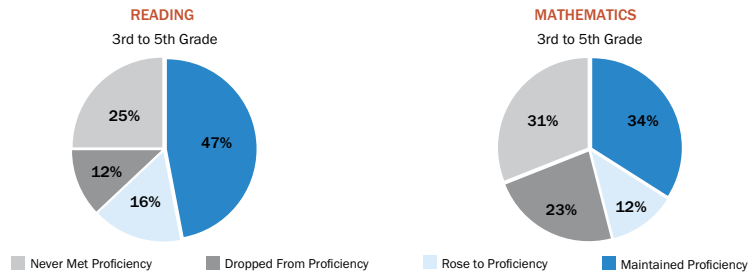
ACHIEVEMENT GAPS

Line graphs report changes in reading and math achievement gaps for student subgroups with 30 or more students.
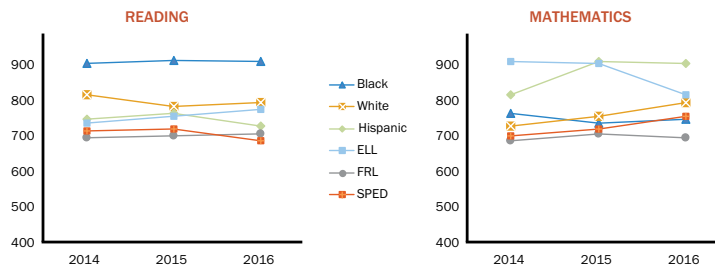
Cognitive Outcome Indicators: High Schools

Outcome indicators for high schools align with college- and career-ready expectations and are consistent with ESSA requirements. The set of outcome indicators includes the percentage of students graduating high school with knowledge and skills necessary for an effective transition to college and career training. Information is reported longitudinally and by student subgroups to assess both academic excellence and equity. The combined set of indicators can be used by educators, policymakers, and the public to evaluate school progress toward the goal of graduating all students college and career ready.

Figure 13 is a sample page displaying cognitive outcomes for high schools. A line graph is used to display **graduation rates** for the last 4 years. The average graduation rate is based on the 4-year adjusted cohort formula; it is reported for all students and all student subgroups with 30 or more students. The information can be adjusted to correspond to other reporting formulas. With Oklahoma's current A–F letter grades, graduation rate only counts for a few bonus points, and no information is provided on graduation rates for student subgroups or changes in rates over time. EQuIPs do both.

High school graduation is an essential outcome, but graduation should not be the basis of inferences about student readiness for college and careers. EQuIP uses some type of readiness indicator, which could involve a number of different assessments, from Smarter Balanced assessments to college entrance exams like ACT and SAT to PARCC or other state curricular exams that align with content standards. We are not advocating for any specific assessment. We use Smarter Balanced and ACT just as examples. The decision about what to use should be made at the state or local level with considerable consultation with assessment experts and stakeholders.

Notice that with the ACT example, a bar graph shows the percentage of students scoring at or above benchmark scores, within two points of the benchmark, or three or more points below the benchmark. Benchmark scores have been empirically established by ACT as the minimum score to indicate student preparedness for success in credit-bearing first-year courses in 2-year or 4-year colleges and career/technical schools (ACT, 2015). In addition to benchmark scores, EQuIP graphs student composite scores over 5 years by student subgroups with 30 or more students.

Graduation rates and scores on readiness assessments are useful indicators, but they leave open the question of access to and enrollment in postsecondary education. A true indicator of whether students are college and career ready after graduating would capture their endeavors and achievements after high school. EQuIPs report **postsecondary access** as the percentage of students enrolling in both 2-year and 4-year college, career training, or the military. Eventually, data about student completion of these programs would also be useful. In accounting for postsecondary access and completion, we hope to build better knowledge about the transition from high school to college and careers, as well as to encourage more purposeful efforts to bridge the gap between common education and higher education.

Figure 13
An Example of an EQuIP Page Reporting Cognitive Outcome Indicators for a High School

COLLEGE AND CAREER READINESS

SAMPLE HIGH SCHOOL

College and Career Readiness refers to the readiness of high school graduates to enter a career, extended training, or specialized education without remediation. Initially, school graduation rates, post-graduation options, ACT performance, and "on-track to graduate" measures serve as indicators of college and career readiness.
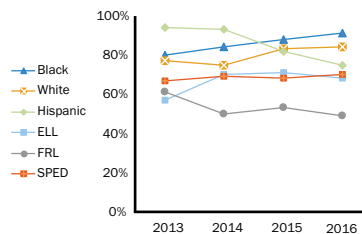
COMPOSITION OF SCHOOL

Number of Students..**594**

Gender

  Male: 47%   Female: 53%

Ethnicity

  Asian: 3%   Black: 27%

  Hispanic: 38%   White: 22%

  American Indian: 3%   Multiple Races: 7%

English Language Learners...**17%**

Students with Special Needs......................................**26%**

**Free or Reduced-Price Lunch Rate** **68%**

GRADUATION RATES AND POSTSECONDARY ENROLLMENT

The line graph tracks graduation rates by student subgroup over 4 years. The pie chart indicates postsecondary enrollment for the class of 2016.

GRADUATION RATES
(by subgroup)

POSTSECONDARY ENROLLMENT
Percent of graduating class of 2015 enrolled in:

- 2-Year University — 38%
- 4-Year University — 20%
- Career Training — 18%
- Military — 8%
- Other — 16%
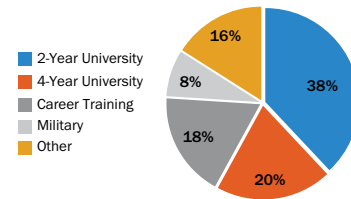
COLLEGE READINESS

The bar graph indicates the percentage of students achieving the benchmark scores for various sample assessments. The line graph tracks changes in composite scores by student subgroup.

SMARTER BALANCED
(by content area)

	ELA	Reading	Math
ALD, Levels 2 and below	37	32	56
ALD, Level 3	32	18	15
ALD, Level 4	31	50	29
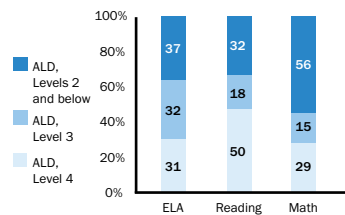
ACT COMPOSITE SCORES
(by subgroup)
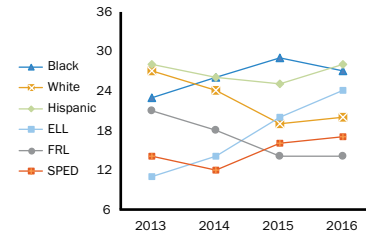
ON TRACK TO GRADUATE

The descriptive data report the percentage of 9th and 10th grade students on track for graduation. The bar graph reports the percentage of 10th grade students meeting the benchmark scores for ACT Aspire subject tests.

PROGRESS TOWARD GRADUATION
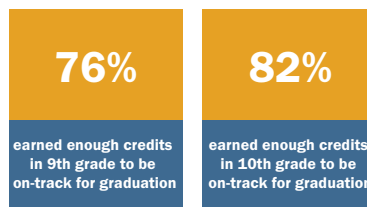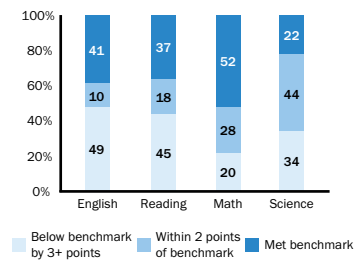
76% earned enough credits in 9th grade to be on-track for graduation

82% earned enough credits in 10th grade to be on-track for graduation

ACT ASPIRE

	English	Reading	Math	Science
Met benchmark	41	37	52	22
Within 2 points of benchmark	10	18	28	44
Below benchmark by 3+ points	49	45	20	34

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education*.

EQuIPs also provide an early warning signal to high schools by accounting for student progress toward graduating with college and career readiness. On-track to graduate reports the percentage of students earning the required number of credits in 9th and 10th grade, and the percentage of students meeting benchmark standards on the 10th grade ACT Aspire subject tests.

Formative Noncognitive Outcomes (Intra- and Interpersonal Indicators)

Figure 14 is a sample page displaying noncognitive outcomes for elementary/middle schools. Adequate levels of cognitive outcomes do not guarantee that deeper learning/college- and career-ready standards have been achieved. There are also intra- and interpersonal abilities and mindsets that schools are responsible for fostering. These indicators can inform improvement decisions by capturing elements of deeper learning not measured with state curricular tests. Proposed indicators should be used in a formative manner and would ideally come from self-report measures completed by students, authentic and problem-based district or school assessments, and/or observational data that are part of administrative records. Much of this information gathering can occur with minimal disruption, relying on administrative data when possible and otherwise integrating with existing data collection efforts. Figure 14 is an example of formative indicators for deeper learning and college and career readiness.

Deeper learning opportunity indicators focus on the degree to which students have opportunities to apply basic knowledge and skills to real-world situations. They are valuable for cognitive, intrapersonal, and interpersonal growth in students. Deeper learning opportunities can be demonstrated in different ways. One way is to follow the lead of the Organisation for Economic Co-operation and Development (OECD) through its PISA program. OECD uses a student survey to find out how often students engage in tasks requiring knowledge and skill application. So, for example, students are asked how often they work word problems in math, or how often they work on real-world problems in schools. A second way is to account for enrichment and learning opportunities available to students through advanced placement courses, career and technical training, fine arts programs, STEM, speech and debate, clubs, college trips, etc.

Intrapersonal indicators can be captured with authentic and performance-based assessments in districts and schools. Intrapersonal indicators report on social-emotional characteristics related to cognitive competencies, and successful school and workplace performance. EQuIPs rely on a self-report student measure of self-regulated learning and the percentage of students chronically absent. Self-regulated learning captures a trait that consistently shows up as predictive of educational outcomes and future life success (National Research Council, 2012; Kautz, Heckman, Diris, Weel, & Borghans, 2014). Chronic absences, a behavioral pattern that is detrimental to student learning and development, are easily calculated from administrative records and reports (Balfanz & Byrnes, 2012). Thriving students are engaged in their schools. Engagement indicators reflect a school environment where students are invested in deep learning, excited about school, and optimistic about their future. Many engagement surveys and items exist and can be used. We propose the engagement items from the Quaglia National Student Voice Survey.

Interpersonal indicators reflect competencies of collaborative problem solving, communication, and teamwork. These skills can be measured through deeper learning processes, authentic district assessments, or a direct measure of students' interpersonal skills (Stecher & Hamilton, 2014). Wang, MacCann, Zhuang, Lie, & Roberts (2010) developed a teacher rating scale of student teamwork that provides valid information about students' interpersonal competencies.

Figure 14
An Example of an EQuIP Page Reporting Noncognitive Outcome Indicators for an Elementary/Middle School

DEEPER LEARNING: NONCOGNITIVE

SAMPLE ELEMENTARY SCHOOL

Noncognitive indicators of deeper learning not measured with state curricular tests can help inform improvement decisions. Initially, noncognitive indicators include deeper learning opportunities, intrapersonal characteristics, and interpersonal dispositions.

Strong Capacity

Moderate Capacity

Low Capacity

2013　2014　2015　2016

■ Interpersonal Dispositions　■ Intrapersonal Characteristics
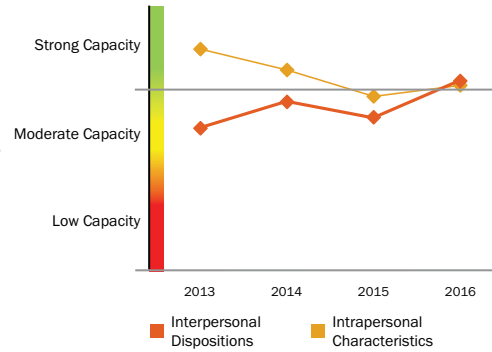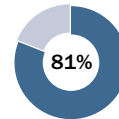

DEEPER LEARNING OPPORTUNITIES

The degree to which students have opportunities to apply basic knowledge and skills to real-world situations. Indicators include student self-reports of instructional tasks, and objective indicators about AP enrollment and concurrent enrollment in college/career training.
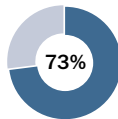
73% of students believe they regularly work on real-world problems in school.
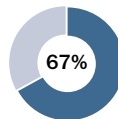
67% of students believe they regularly work in teams to find solutions to problems.

81% of students believe they have regular opportunities to present their work to peers and adults.

40% Students Chronically Absent

ONE ENRICHMENT PROGRAM AT SCHOOL STUDENTS WHO PARTICIPATE IN AT LEAST	60%

STUDENTS WHO PARTICIPATE IN A STEM PROGRAM AT SCHOOL	31%
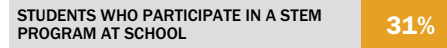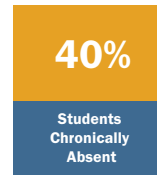
INTRAPERSONAL CHARACTERISTICS

The degree to which the regulation of learning processes and outcomes derives from the internal control of students, and the extent to which students are emotionally involved in their schoolwork and feel that it activates their curiosity, creativity, and passion.
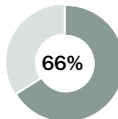
66% of students feel proud of being part of the school.
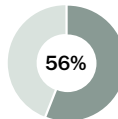
56% of students enjoy learning new things.
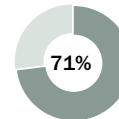
71% of students believe they keep trying after they fail.
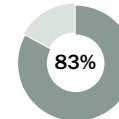
83% of students believe they get schoolwork done even when they don't feel like doing it.

INTERPERSONAL DISPOSITIONS

The degree to which students can work collaboratively and as active contributors to a team, and students' ability to listen to one another and communicate effectively.
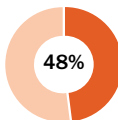
48% of students believe teachers show students how to solve problems themselves.
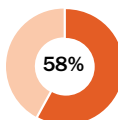
58% of students believe teachers talk about the connection between what is studied in school and what happens in real life.
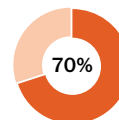
70% of students believe teachers in the school expect students to work hard.
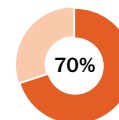
70% of students believe teachers in the school challenge students to achieve academic goals.

Source: Adapted from Adams et al. (2015b). *Next Generation School Accountability: A Report Commissioned by the Oklahoma State Department of Education.*

EQuIP Online

EQuIP comes with the potential to create electronic accessibility through an online platform that is compatible with personal computers and mobile devices. The online site would guide the user in accessing profile pages, graphs, and descriptions of the concepts captured in the reports. A fully activated online feature would enable the user to drill down to additional information on school outcomes, processes, and resources. Figures 15 and 16 present screenshots of a sample "landing page" and "profile page," respectively. Users looking for information about deeper learning, college and career readiness, or different capacities would simply click on the components of EQuIP they wish to see. Information graphs could then display descriptive data and information in a way that is clear and easy to understand.

Figure 15
Sample Electronic Landing Page for EQuIP Online

Figure 16
Sample Electronic Profile Page for EQuIP Online

Part Four: Designing and Implementing Next Generation Accountability

The previous three parts outlined some critical deficiencies in current state accountability systems and proposed a design for a next generation accountability system to address these deficiencies and establish alignment with the goals of ESSA. In this final section, we translate next generation accountability into a set of recommendations for enacting this framework for both state and local education. These recommendations have three distinct targets: (a) accountability policy; (b) alignment of standards, assessments, and accountability; and (c) school, district, and state capacity building in support of the accountability framework. We then provide some concluding remarks on next generation accountability under ESSA.

Accountability Policy

Any accountability policy that places disproportionate weight on a single composite indicator of school performance that is based on aggregate assessments of low-level knowledge and skills is certain to leave students behind. We envision effective accountability policy as establishing high expectations for schools and students while supporting meaningful investments in resources and processes that can move the entire educational system toward a desired state (Darling-Hammond et al., 2014). Many states have set a vision of deeper learning and college and career readiness for all students; it is now time for states and districts to invest in an accountability system that supports these goals, as well as innovation, transformation, and ongoing improvement, more broadly.

For many states, the design of next generation accountability will require revision of state statutes that prescribe the formulation and use of composite indicators of school performance for accountability purposes. As demonstrated earlier, many states that prescribe a system for rating schools and/or rendering summary judgment about a school's performance via a single indicator fail to deliver useful information about school progress toward the goals of deeper learning and college and career readiness. Moreover, this type of system does not support the intelligent investment in strategies that can build school capacity by targeting sources of underperformance. Going forward, and in accordance with key provisions under ESSA, it seems reasonable that state law set clear and high expectations for an accountability system aligned with deeper learning and college and career readiness, while providing state departments of education with the authority and managerial flexibility to design a system that fulfills these expectations.

We propose the following recommendations and rationale for an accountability policy.

1. Do not use a single summative index to report accountability information. Outcome evidence should clearly report student performance toward deeper learning and college- and career-ready standards, changes in student performance over time, and achievement gaps.

   - Single summative indices cannot be used to make valid and reliable judgments of school quality.
   - Single summative indices do not provide useful information for improvement.

- Outstanding and equitable outcomes should be the goal of every school.
- Variation in student outcomes needs to be studied and understood so targeted action can address performance gaps.
- Trend data provide a more accurate account of student and school performance compared to the instability of time point estimates.
- Evidence on achievement equity and performance trends allows for more reliable identification of schools in need of state intervention.

2. Multiple indicators of capacity for quality improvement should be part of a school profile.

- Knowledge formation includes understanding what, how, and why improvement is or is not happening.
- States and districts can better identify schools in need of intervention by understanding capacity differences among schools.
- We must learn the lessons of *Campbell's Law*. At no time should social measures be used in high-stakes or otherwise summative decision making—to do so completely undermines their validity as measures of social processes (Duckworth & Yeager, 2015).

Accountability policy should adhere to the Standards for Educational and Psychological Testing, which notes, "Those who mandate the use of tests in policy, evaluation, and accountability contexts and those who use tests in such contexts should monitor their impact and should identify and minimize negative consequences" (Standard 13.8, AERA, APA, NCME, 2014). Thus, accountability policy should be written in the least restrictive and prescriptive terms possible to allow for corrective action and improvement.

Alignment of Standards, Assessments, and Accountability

We establish in this report clear working definitions of deeper learning and college and career readiness, positioning them as critical educational outcomes of a next generation accountability system. As an essential first step, care must be taken to ensure that curricular, assessment, and evaluation systems all align with and/or serve these larger operational definitions of what it means to be a healthy, productive citizen. Toward this end, we advance five recommendations in the areas of curricular standards, assessment, and evaluation:

1. The development of a new, coordinated system of multiple assessments, both formative and summative, to measure student learning using the operational definitions we have constructed for deeper learning and college and career readiness. Such a system should be defined by the following major features (Darling-Hammond et al., 2014):

- Higher-order cognitive skills are assessed.
- Critical skills are assessed with high fidelity.
- Assessments are benchmarked to international standards.
- Assessments are instructionally sensitive and educationally valuable.
- Assessments are valid, reliable, and fair.

2. In addition to measures of student learning, indicators of dispositional and behavioral constructs associated with deeper learning and college and career readiness should be included.

3. The system should emphasize frequent use of formative assessments, particularly those embedded in instruction.

4. Consideration should be given to grade-span testing of achievement outcomes. It may not be necessary or even desirable to test every student in every subject every year.

5. Assessment results should be reported by student subgroups to highlight performance gaps.

Most importantly, the accountability assessment indicators used should not be combined to form a summative indicator of school performance. Assessments are selected because they provide useful and actionable information for schools regarding student progress. If the primary purpose of an accountability system is to improve the educational experiences for students, that is best accomplished when separate indicators retain their meaning and value.

School, District, and State Capacity Building

Success in using next generation accountability to elevate educational quality and improvement depends on the degree to which school professionals and stakeholders can unlearn some past ways of doing things. It also requires a support infrastructure that exceeds what has existed historically. Although professional learning opportunities for all school professionals are important, it is also necessary to provide time and space to understand the intent of the new policy, how it differs from past policy, and the significance these differences hold for changes to practice (Cohen, 1990; Cohen & Hill, 2001; Spillane, 2004). There are considerable constraints on the degree to which school professionals are able to take advantage of opportunities for learning in support of school improvement. A next generation framework, it is argued, identifies **five *essential* systemwide components** of a support infrastructure for building capacity across the educational system:

1. **A systemwide culture grounded in "learning to improve."** Change under next generation accountability is dynamic and context specific. A new support system needs to communicate its importance for enhancing local as well as statewide educational improvement. For genuine change to take root, it is essential that the purposes, processes, and goals of improvement be shared within and across sites. The ideas of science of improvement methodology and Networked Improvement Communities (NICs) proposed by the Carnegie Foundation for the Advancement of Teaching and Learning (Bryk et al., 2015) are two research-based frameworks that seem to have particular utility for accomplishing this.

   The Carnegie Foundation advocates a process model for improvement that is adaptable to any improvement context, yet adopts a highly disciplined approach to studying school problems and testing solutions. The model urges schools to implement slowly and learn fast, recognizing that initiatives often fail because organizations jump too quickly to large-scale implementation. The Plan, Do, Study, Act (PDSA) cycle of inquiry is a cornerstone of this process, as is the idea of NICs. The NICs concept recognizes social networks as valuable resources for change that facilitate the sharing of knowledge across improvement contexts. By connecting similar schools or districts, NICs institutionalize channels through which information on learning can be shared.

   States and districts that intend to implement a next generation accountability system should consider at the outset the scale at which these new initiatives should

be implemented: capacity and/or will for planned change among stakeholders vary considerably from site to site. Figure 17, adapted from conceptualizations by McLaughlin (1987) and Bryk and his colleagues (2015), provides some guidelines for determining the level at which stakeholders should begin implementation of any planned change, based on an assessment of stakeholder capacity and will for improvement. As this figure indicates, in no case do we believe it is appropriate to move immediately to systemwide implementation, even in instances where stakeholders exhibit strong capacity and will for change. Jumping straight to systemwide implementation reflects what we believe is the status quo of education reform. An overenthusiastic orientation to improvement exhibits not only a profound lack of respect for the complexity that accompanies any planned change within school systems, but also a lack of pragmatism in planning, given the significant scarcity of resources needed to make any change work at scale.

Figure 17
Guidelines for Next Generation Accountability Implementation Scale
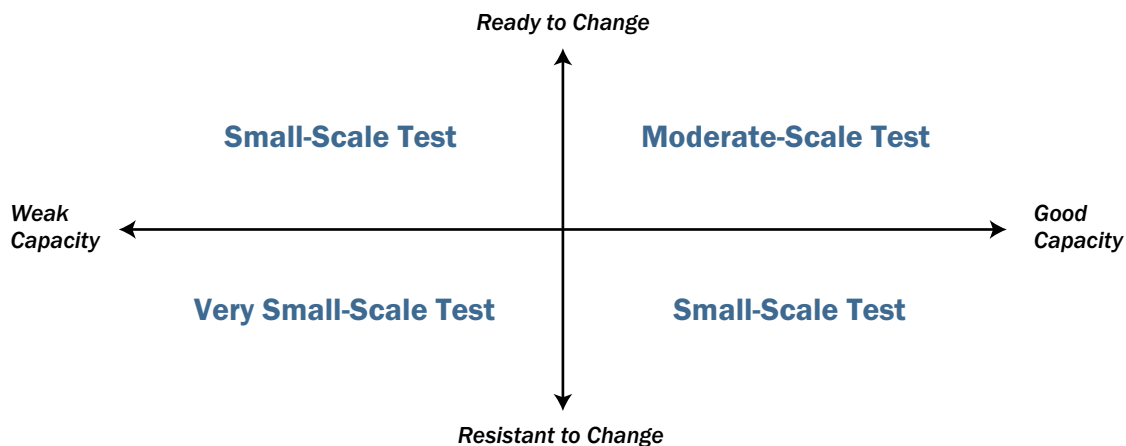
2. **Development of strong pedagogical data literacy skills.** In this framework, data are meant to enhance decision making, not be a substitute for it. "Pedagogical data literacy" (Mandinach, 2012) refers to technical ability related to one or more of the following areas: numeracy and statistical knowledge, facile use of data analysis software, and what might be considered general inquiry skills, such as the ability to formulate and test appropriate questions and to develop solutions based on findings (Kerr et al., 2006; Marsh et al., 2006). Research demonstrates that the demands we place on school professionals to select and use appropriate data for decision making have far outstripped our attention to the need for better training on how to go about this process in a meaningful way (Datnow & Park, 2014; Mandinach, 2012). A next generation accountability system would maintain focus on data-driven decision making, but would also ask school professionals to become expert on the use of data to explain outcomes and decide on appropriate interventions. This recommendation has strong implications for teacher- and leader-preparation programs. Preservice teachers and aspiring leaders need the knowledge and skill set to develop meaningful classroom assessments, to interpret assessment results, and to make meaning from student performance information.

3. **Prioritization of resources for sustaining ongoing improvement.** Systemwide availability of resources such as time, access to expertise, and collaborative opportunities is critical (Ingram, Louis, & Schroeder, 2004). Meaningful learning occurs in collaboration with others (Bandura, 1978), and having easy access to colleagues, instructional coaches, and other leaders, as well as outside experts, will ensure that school staff can see a wide range of possibilities in addressing issues of teaching, leading, and learning. Allocating time and improving access to expertise and collaborative opportunities will likely require increased (or some reallocation of) school funding. The intentional allocation of these key resources sends a signal about the importance of such endeavors and creates conditions in which schools and school personnel can achieve critical learning.

4. **A coherent structure of state-level support for learning to improve, including a strong Longitudinal Data System (LDS) infrastructure.** Learning to improve has to be part of a larger, coherent framework of state-led support. The alignment of a strong culture around learning to improve, as well as the allocation of key resources to support change, all need to be present, focused, and coordinated with the next generation accountability. This support structure could build on existing support resources in states, such as school support offices, but it will require significant expansion to accommodate disparate learning needs across states and districts. The current resources provided to most of these support networks are insufficient. States and their stakeholders are urged to partner with intermediate service agencies (including state and local universities) where capacity already exists to genuinely assist in efforts to develop a next generation accountability system.

   Moreover, strong, well-designed state LDSs play a crucial role in the development of a coherent structure of state-level support. Unless it is well organized, readily accessible, and produced in a timely fashion, much of the data to be collected under next generation accountability is useless to states, schools, and districts for the purposes of decision making (Conaway, Keesler, & Schwartz, 2015). Ideally, districts and schools under next generation accountability would have access to real-time, disaggregated data on their inputs, processes, and outcomes above and beyond what is provided in EQuIP reports.

5. **Educator labor market policy that supports the above elements.** Little progress in the preceding elements will be made without addressing key educator labor market challenges, and this includes a reexamination of current policy tied to the supply of experienced educators in particular states and regions of the United States. For example, a recent study highlights some disturbing current and future trends with respect to teacher supply and demand in Oklahoma. Currently, teacher salaries fall well below not only those of adjacent states but also in relation to their non-educator counterparts within the same Oklahoma labor market. In addition, the gap between the number of teachers entering and those leaving the profession is widening quickly (Berg-Jacobson & Levin, 2015). Recent evidence suggests that these challenges are not unique to any one state or region of the country. (In California, see, for example, Darling-Hammond, Fuger, Shields, & Sutcher, 2016; in Washington, see Association for Washington School Principals [AWSP], 2015; in South Carolina, see Garrett, 2016.) It stands to reason that attempting to address key challenges like teacher corps stability within schools without addressing these larger labor market issues would be a futile effort.

Furthermore, our proposed system of supports will require more professional educators who have had considerable past experience working in schools. In some cases, this may require an examination of teacher retirement incentive laws (i.e., "double-dipping" regulations) to determine if they preclude former teachers from participating in these new support positions. If so, then providing exemptions so that highly qualified former educators can participate in these support positions will be necessary. These policy changes are not without precedent: Some states, such as Michigan, have responded to labor market shortages by relaxing these regulations.

Just as the overreliance on extrinsic pressure, rewards, and/or punishments as a long-term approach to motivation is a flawed working assumption for improvement, so too is the assumption that schools, districts, and the working professionals who comprise them can make meaningful improvements with little to no support. Changing existing patterns of behavior is difficult work, even with strong motivation. As Spillane (2004), once noted, " … good intentions only go so far. When it comes to implementing new ideas about instruction, all the will in the world is not enough" (p. 168).

It is time we stop treating support for planned change as an afterthought in our educational policy and practice. Even well-laid plans are doomed to fail without proper support, and we would argue that in many ways attention to support for planned change should surpass that of the planned change itself. Attention to both the design of a better framework for improvement and the support it needs to thrive at the state and local level will determine the ultimate success of this next generation accountability framework.

Concluding Thoughts on Next Generation Accountability

This approach, unlike current test-based accountability, draws on the best social science evidence we have to date on how individuals and organizations are motivated, as well as how they learn, grow, and thrive. Although this is no guarantee of success, we believe it is a more fruitful starting point for school improvement policy and practice than one that is based on conjecture, "gut feeling," or ideas about motivation that have long since been debunked. The increased authority, autonomy, and flexibility given to states under ESSA represent a welcome opportunity for states to get improvement right. However, let us be clear: These changes to federal law by no means guarantee any measure of success in addressing the failings of past policies—far from it. As Milbrey McLaughlin (1987) once wrote, " … policy cannot always mandate what matters to outcomes at the local level" (p. 171). What state and local education agencies need now is a plan to leverage their newfound freedom into success in moving their educational systems forward.

There is no doubt that the work facing us is substantial. This report contains the broad strokes of a plan for states and other local education agencies that are serious about making schools work for all children, families, and communities. In the true spirit of improvement, however, this vision necessarily leaves the finer details about how to execute it up to local policy actors. We believe this is exactly as it should be.

Appendix A: Analysis of Oklahoma Student Achievement Growth from 2011–12 to 2013–14 in Math and Reading

This brief section summarizes the technical nature of the analyses of Oklahoma student achievement, which were reported in aggregate in Part One. Our two-level hierarchical linear models estimated the relationship of reading and math achievement growth over the academic years of 2011–12 to 2013–14 to important characteristics of students in grades 3–8.

We fit two-level HLM growth models (responses nested within students) separately to the Oklahoma achievement reading and math achievement data with respect to a set of student-level and response-level covariates. At the response level, whether or not the student experienced an online testing occasion was utilized. At the student level, the following covariates were used to predict achievement growth: free and reduced-price lunch status of student, ELL status, race/ethnicity, gender, and A–F letter grade of the school in which the student resided. Only 1% of students in the final sample were missing the A–F letter grade recorded for the 2011–12 school year for their school.

The final achievement models assumed an underlying linear pattern to the achievement data. Non-linear (i.e., quadratic patterns) were investigated as well, but the linear model had better overall fit compared to the non-linear model. The effective sample sizes for each analysis were 799,981 responses nested within 392,692 students for reading and 792,356 responses nested within 392,930 students for math. Median and modal number of responses per student was 3 for both reading and math. The final achievement growth model, in which Time was centered at Year 1, had the following structure:

Level 1 (Response Level):

$$Y_{ti} = \pi_{0i} + \pi_{1i}*Time_{ti} + \pi_{2i}*(OnlineTest) + e_{ti}$$

Level 2 (Student Level):

$$\pi_{0i} = \beta_{00} + \sum_{q=1}^{Q} \beta_{0q}X_q + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \sum_{q=1}^{Q} \beta_{1q}X_q + r_{1i}$$

$$\pi_{2i} = \beta_{20}$$

Where:

Y_{ti} represents the reading or math achievement score for student i at time t.

π_{0i} represents the intercept of the true change trajectory for student i in the
2011–12 school year.

π_{1i} represents the linear yearly rate of growth in achievement in reading or math for student i.

$Time_{ti}$ was measured in academic year and centered on Year 1 (the 2011–12 school year).

π_{2i} represents the response-level effect of taking the test online versus on paper on reading/math
achievement.

e_{ti} is the within-student random effect (error term) assumed to be normally distributed with a
mean of 0 and a constant variance σ^2 across students.

β_{00} represents the average reading or math score for a white, non-ELL, non-free or reduced-price
lunch student in an A school during the 2011–12 school year (the first year of this study,
i.e., their "initial status").

$\beta_{0q}X_q$ signals that q number of student-level covariates were entered into the achievement
models predicting achievement initial status in the 2011–12 school year. As mentioned
before, these were: free and reduced-price lunch status of student, ELL status, race/
ethnicity, gender, and A–F letter grade of the school in which the student resided.

r_{0i} represents the between-student random effect (error term) with respect to student
initial status.

β_{10} represents the average rate of linear growth for white, non-ELL, non-free or reduced-price
lunch students in an A school.

$\beta_{1q}X_q$ signals that q number of student-level covariates were entered into the achievement
models predicting linear growth rate in achievement during the 3 years under study. These
were the same variables entered to predict student achievement initial status: free and
reduced-price lunch status of student, ELL status, race/ethnicity, gender, and A–F letter
grade of the school in which the student resided.

r_{1i} represents the between-student variation in the rate of linear growth.

β_{20} represents the aggregate, fixed effect of online test taking versus paper on reading/math
achievement.

References

NOTE: References marked with an asterisk indicate resources consulted but not cited.

ACT. (2015). *The condition of college and career readiness 2015*. ACT, Inc. Retrieved from http://www.act.org/newsroom/data/2015/index.html.

Adams, C.M. (2014). Collective student trust: A social resource for urban elementary students. *Educational Administration Quarterly, 50*(1), 135–159.

Adams, C.M., Forsyth, P.B., Dollarhide, E.A., Miskell, R.C., & Ware, J.K. (2015). Self-regulatory climate: A social resource for student regulation and achievement. *Teachers College Record, 117*(2).

Adams, C.M., Forsyth, P.B., Ford, T.G., Ware, J.K., Barnes, L.B., Khojasteh, J., et al. (2015). *Next generation school accountability: A report commissioned by the Oklahoma State Department of Education.* Stillwater, OK: Oklahoma Center for Education Policy (The University of Oklahoma) and the Center for Educational Research and Evaluation (Oklahoma State University). Retrieved from http://okedpolicy.org/wp-content/uploads/2015/12/Next-Generation-School-Accountability-Report-Final.pdf.

Adams, C.M., Dollarhide, E.A., Forsyth, P.B., Miskell, R.C., Ware, J.K., Barnes, L.L.B., Khojasteh, J., & Mwavita, M. (2013). *Oklahoma school grades: Hiding "poor" achievement.* Tulsa, OK: Oklahoma Center for Education Policy.

*Almlund, M., Duckworth A., Heckman J. J., & Kautz, T. (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education, Volume 4* (pp. 1–181). Amsterdam, The Netherlands: Elsevier.

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

*Annenberg Institute for School Reform at Brown University. (2002). *School communities that work for results and equity.* Retrieved from http://www.annenberginstitute.org/sites/default/files/product/283/files/Results.pdf.

Association of Washington School Principals (2015). *The teacher shortage: Key findings of the 2015 OSPI/AWSP teacher shortage survey.* Office of the Superintendent of Public Instruction: Author. Retrieved from http://www.k12.wa.us/LegisGov/2016documents/TeacherShortageDashboard.pdf.

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in school work. *British Journal of Educational Psychology, 72*, 261–278.

Balfanz, R., & Brynes, V. (2012). T*he importance of being in school: A report on absenteeism in the nation's public schools.* Baltimore, MD: Johns Hopkins University Center for Social Organization of Schools.

*Ball, D.L., & Cohen, D.K. (1999). Developing practice, developing practitioners. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco, CA: Jossey-Bass.

Bellanca, J.A. (2015). *Deeper learning: Beyond 21st century skills.* Bloomington, IN: Solution Tree Press.

Berg-Jacobson, A., & Levin, J. (2015). *Oklahoma study of educator supply and demand: Trends and projections.* Washington, DC: American Institutes for Research.

Berliner, D. (2006). Our impoverished view of educational reform. *The Teachers College Record, 108*(6), 949–995.

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

*Borman, G.D., Benson, J., & Overman, L.T. (2005). Families, schools, and summer learning. *The Elementary School Journal, 106*(2), 135–150.

Bryk, A.S., Gomez, L.M., Grunow, A., & LeMahieu, P.G. (2015). *Learning to improve: How America's schools can get better at getting better.* Cambridge, MA: Harvard Education Press.

*Bryk, A.S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement.* New York, NY: Russell Sage Foundation.

*Bryk, A.S., Sebring, P.B., Allensworth, E., Easton, J.Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago.* Chicago, IL: University of Chicago Press.

Campbell, D.T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*, 67–90.

Cohen, D.K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12*, 327–345.

*Cohen, D.K. (2014). *Teaching and its predicaments.* Cambridge, MA: Harvard University Press.

*Cohen, D.K., & Barnes, C. A. (1993). Pedagogy and policy. In D. K. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice* (pp. 207–239). San Francisco, CA: Jossey-Bass.

Cohen, D.K., & Hill, H. (2001). *Learning policy: When state education reform works.* New Haven, CT: Yale University Press.

Coleman-Jensen, A., Gregory, C., & Singh, A. (2014). *Household food security in the United States in 2013.* United States Department of Agriculture, Economic Research Report 173.

Conaway, C., Keesler, V., & Schwartz, N. (2015). What research do state education agencies really need? The promise and limitations of state longitudinal data systems. *Educational Evaluation and Policy Analysis, 37*, 16S–28S.

Darling-Hammond, L., Hightower, A.M., Husbands, J.L., Lafors, J.R., & Young, V.M. (2005). *Instructional leadership for systemic change: The story of San Diego's reform.* Lanham, MD: Scarecrow Education.

Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning.* Stanford, CA: Stanford Center for Assessment, Learning, and Equity.

Darling-Hammond, L., Furger, R., Shields, P., & Sutcher, L. (2016). *Addressing California's emerging teacher shortage: An analysis of sources and solutions.* Palo Alto, CA: Learning Policy Institute. Retrieved from http://www.learningpolicyinstitute.org/addressing-ca-teacher-shortage.

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.

Datnow, A., & Park, V. (2014). *Data-driven leadership.* San Francisco, CA: Jossey-Bass.

Deming, E.W. (2000). *The new economics for industry, government, and education*. Cambridge, MA: MIT Press.

*Dittus, M.S., & Epstein J. (2007). Family and community involvement in schools: Results from the School Health Policies and Programs Study 2006. *Journal of School Health, 77*, 567–579.

Domaleski, C., & Perie, M. (2013). *Promoting equity in state education accountability systems*. Lawrence, KS: National Center for the Improvement of Educational Assessment, Center for Educational Testing and Evaluation, University of Kansas.

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

Forsyth, P.B., Adams, C.M., & Hoy, W.K. (2011). *Collective trust: Why schools can't improve without it.* New York, NY: Teachers College Press.

Forsyth, P.B. & Tallerico, M. (1998). Accountability and city school leadership. *Education and Urban Society, 30*(4), 546–555.

Forte, E. (2010). Examining the assumptions underlying the NCLB federal accountability policy on school improvement. *Educational Psychologist, 45*(2), 76–88.

Fullan, M. (2010). *All systems go: The change imperative for whole system reform*. Newbury Park, CA: Corwin Press.

Garrett, J. (2016). *A report on the fall 2015 supply and demand survey*. Rock Hill, SC: Center for Educator Recruitment, Retention, and Advancement, Winthrop University. Retrieved from https://www.cerra.org/uploads/1/7/6/8/17684955/2015_supply_demand_report__1_.pdf.

*Glaser, R. (1984). The role of knowledge. *American Psychologist, 39*(2), 93–104.

*Grubb, W.N. (2006). When money might matter: Using NELS:88 to examine the weak effects of school funding. *Journal of Education Finance, 31*(4), 360–378.

*Grubb, W.N. (2009). *The money myth: School resources, outcomes, and equity*. New York, NY: Russell Sage Foundation.

Hall, D. (2013). *A step forward or a step back? State accountability in the waiver era.* Washington, DC: The Education Trust. Retrieved from http://files.eric.ed.gov/fulltext/ED543222.pdf.

Hamilton, L.S., Schwartz, H.L., Stecher, B.M., & Steele, J.L. (2013). Improving accountability through expanded measures of student performance. *Journal of Educational Administration, 51*, 453–475.

Hanushek, E.A., Kain, J.F., O'Brien, D.M., & Rivkin, S.G. (2005). *The market for teacher quality* (No. w11154). Cambridge, MA: National Bureau of Economic Research.

*Hatch, T. (2005). *Into the classroom: Developing the scholarship of teaching and learning.* Indianapolis, IN: Jossey-Bass.

Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school.* New York, NY: Teachers College Press.

Harris, A. (2011). Reforming systems: Realizing the fourth way. *Journal of Educational Change, 12*(2), 159–171.

Heckman, J.J., Pinto, R., & Savelyev, P.A. (2013). Understanding the mechanisms through which an influential early childhood boosted adult outcomes. *American Economic Review, 103*(6), 1–35.

Herman, J., & Linn, R. (2013). *On the road to assessing deeper learning: The status of smarter balanced and PARCC Assessment Consortia*. CRESST Report 823. Los Angeles, CA: University of California Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Hewlett Foundation (2014). Deeper learning. Retrieved from http://www.hewlett.org/programs/education/deeper-learning.

*Hiebert, J., Gallimore, R., & Stigler, J.W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher, 31*(5), 3–15.

Ho, A.D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher, 37*(6), 351–360.

Hopson, R. (2014). Why are studies of neighborhoods and communities central to education policy and reform? *Urban Education, 49*(8), 992–995.

Horsford, S.D., & Sampson, C. (2014). Promise neighborhoods: The promise and politics of community capacity building as urban school reform. *Urban Education, 49*(8), 955–991.

*Howe, K.R., & Murray, K. (2015). *Why school report cards merit a failing grade*. Boulder, CO: National Education Policy Center. Retrieved from http://greatlakescenter.org/docs/Policy_Briefs/Howe-StateReportCards.pdf.

*Hoy, W. (2012). School characteristics that make a difference for the achievement of all students: A 40-year odyssey. *Journal of Educational Administration, 50*(1), 76–97.

Ingersoll, R.M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38*(3), 499–534.

Ingram, D., Louis, K.S., & Schroeder, R. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *The Teachers College Record, 106*, 1258–1287. Retrieved from http://www.tcrecord.org/content.asp?contentid=11573.

Katona, G. (1942). The role of the order of presentation in learning. *American Journal of Psychology, 55*(3), 328–353.

Kautz, T., Heckman, J.J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and noncognitive skills to promote lifetime success* (No. w20749). Cambridge, MA: National Bureau of Economic Research.

Kerr, K.A., Marsh, J.A., Ikemoto, G.S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*, 496–520.

Koretz, D., & Hamilton, L.S. (2006). Testing for accountability in K–12. In R. L. Brennan (Ed.), *Educational measurement* (pp. 531–578). Westport, CT: American Council on Education/Praeger.

Kusurkar, R.A., Cate, T., Vos, C.M., Westers, P., & Croiset, G. (2012). How motivation affects academic performance: A structural equation modeling analysis. *Advances in Health Sciences Education, 18*(1), 57–69.

Langley, G.J., Moen, R.D., Nolan, K.M., Nolan, T.W., Norman, C.L., & Provost, L.P. (2009). *The improvement guide: A practical approach to enhancing organizational performance.* San Francisco, CA: Jossey-Bass.

*Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends.* Cambridge, MA: The Civil Rights Project at Harvard University. President and Fellows of Harvard College.

*Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal comparative and correlational studies. *Review of Educational Research, 78*(3), 608–644.

*Linn, R.L. (2005). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives, 13*(33), 1–20.

*Linn, R.L. (2008). Educational accountability systems. In K. Ryan and L. Shepard (Eds.), *The future of test-based educational accountability* (pp. 3–24). New York, NY: Routledge.

*Lortie, D. (1975). *Schoolteacher: A sociological study.* Chicago, IL: University of Chicago Press.

Louis, K.S., Leithwood, K., Wahlstrom, K., Anderson, S.A., & Michlin, M. (2010). *Learning from leadership: Investigating the links to improved student learning: Final report of research findings.* New York, NY: Wallace Foundation.

Mandinach, E.B. (2012). A perfect time for data-use: Using data-driven decision making to inform practice. *Educational Psychologist, 47*(2), 71–85.

*Marsh, J.A., Sloan, J.M., & Martorell, F. (2010). How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools. *Educational Policy, 24*(6), 872–907.

Marsh, J.A., Pane, J.F., & Hamilton, L.S. (2006). *Making sense of data-driven decision making in education* (RAND Occasional Paper #OP-170-EDU). Santa Monica, CA: RAND. Retrieved from http://www.rand.org/pubs/occasional_papers/OP170.html.

Martin, C., Sargrad, S., & Batel, S. (2016). *Making the grade: A 50-state analysis of school accountability systems*. Washington, DC: Center for American Progress.

*McLaughlin, M.W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9*, 171–178.

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

Niemiec, C., & Ryan R.M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education, 7*(2), 133–144.

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

Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better.* London, United Kingdom: McKinsey.

Murphey, D., Bandy, T., Schmitz, H., & Moore, K.A. (2013). Caring adults: Important for positive child well-being. *Child Trends*. (Research brief.)

National Center for Education Statistics. (2013). *NAEP 2012: Trends in academic progress, reading 1971–2012, Math 1973–2012*. U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/subject/publications/main2012/pdf/2013456.pdf.

National Research Council. (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. In J.W. Pellegrina & M.L. Hilton (Eds.), *Committee on Defining Deeper Learning and 21st Century Skills*. Washington, DC: National Academics Press.

Office of Economic Co-Operation and Development (2014). *Program for international student assessment: Results from 2012*. Retrieved from http://www.oecd.org/pisa/keyfindings/PISA-2012-results-US.pdf.

Oklahoma State Department of Education. (2015). *Oklahoma ESEA Flexibility Renewal Form.* Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/okrenewalreq7282015.pdf.

*Oklahoma State Department of Education. (2015). *A to F Report Card Quick Reference Guide.* Retrieved from http://www.ok.gov/sde/sites/ok.gov.sde/files/documents/files/AFReportCardQuickReferenceGuide.pdf.

Peurach, D.J., & Neumerski, C.M. (2015). Mixing metaphors: Building infrastructure for large-scale school turnaround. *Journal of Educational Change*.

Polikoff, M., McEachin, A., Wrabel, S., & Duque, M. (2014). The waive of the future? School accountability in the waiver era. Educational Researcher. Retrieved from http://edr.sagepub.com/content/43/1/45.

Putnam, R.D. (2015). *Our kids: The American Dream in crisis*. New York, NY: Simon & Schuster.

*Quality Counts. (2014). *District disruption and revival: School systems reshape to compete and improve*. Education Week. Retrieved from http://www.edweek.org/ew/toc/2014/01/09/index.html.

*Reed, E., Scull, J., Slicker, G., & Winkler, A.M. (2012). *Defining strong state accountability systems: How can better standards gain greater traction, a first look*. Washington, DC: Thomas B. Fordham Institute.

*Reeve, J. (2006). Extrinsic rewards and inner motivation. In C. M. Evertson & C. S. Weinstein (Eds.), *Classroom management: Research, practice, and contemporary issues* (pp. 645–664). Mahwah, NJ: Lawrence Erlbaum.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal, 50*(1), 4–36.

Rothstein, R., Jacobson, R., & Wilder, T. (2008). *Grading education: Getting accountability right*. New York, NY: Teachers College Press.

Rowan, B., & Miller, R. J. (2009). *SII Multi-component survey data files user's guide.* Ann Arbor, MI: Consortium for Policy Research in Education.

Ryan, R.M., & Brown, K.W. (2005). Legislating competence: The motivational impact of high-stakes testing as an educational reform. In. C. Dweck & A. Elliot (Eds.), *Handbook of competence and motivation* (pp. 354–372). New York, NY: Guilford Press.

Ryan, R.M., & Deci, E.L. (2002). An overview of self-determination theory: An organismic-dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3–33). Rochester, NY: University of Rochester Press.

*Ryan, R.M., & Weinstein, N. (2009). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. T*heory and Research in Education, 7*(2), 224–233.

*Schneider, A., & Ingram, H. (1990). Behavioral assumptions of policy tools. *Journal of Politics, 52*, 510–529.

*School Leaders Network (2014). *Churn: The high cost of principal turnover*. School Leaders Network: Author.

Schunk, D.H., & Zimmerman, B.J. (Eds.)(2008). *Motivation and self-regulated learning: theory, research, and applications*. New York, NY: Lawrence Erlbaum Associates.

*Schumacher, K.S. (2015). Informing early childhood policy: An analysis of the sensitivity of a school readiness risk index to changes in indicator selection (Doctoral dissertation). Stillwater, OK: Oklahoma State University.

Shepard, L., Hannaway, J., & Baker, E. (2009). *Standards, assessments, and accountability. Education Policy White Paper.* Washington, DC: National Academy of Education. Retrieved from http://eric.ed.gov/?id=ED531138.

*Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–23.

Sigle-Rushton, W., & McLanahan, S. (2004). Father absence and child wellbeing: A critical review. In D. P. Moynihan, T. M. Smeeding, & L. Rainwater (Eds.), *The future of the family* (pp. 116–155). New York, NY: Russell Sage Foundation.

Spillane, J.P. (2004). *Standards deviation: How schools misunderstand education policy*. Cambridge, MA: Harvard University Press.

Standards Company LLC. (2008). Analysis of the enacted curriculum for the Oklahoma State Department of Education for the collection period February–March, 2008. Unpublished report.

Stecher, B.M., & Hamilton, L.S. (2014). *Measuring hard-to-measure student competencies*. Santa Monica, CA: Rand Corporation.

Sullivan, P. (2009). Reciprocal accountability. *International Journal of Public Sector Management, 22*(1), 57–72.

*Tschannen-Moran, M. (2004, November). *What's trust got to do with it? The role of faculty and principal trust in fostering student achievement.* Presented at the annual meeting of the University Council for Educational Administration, Kansas City, MO.

*Tschannen-Moran, M. (2014). *Trust matters: Leadership for successful schools* (2nd ed.). San Francisco, CA: Jossey-Bass.

Wang, L., MacCann, C., Zhuang, X., Lie, O.L., & Roberts, R.D. (2009). Assessing teamwork and collaboration in high school students. *Canadian Journal of School Psychology, 24*(2), 108–124.

Webb, N.L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education (Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6).* Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Webb, N.L. (2002). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states.* Washington, DC: Council of Chief State School Officers.

Webb, N.L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education, 20*(1), 7–25.

*Wertheimer, M. (1959). *Productive thinking*. New York, NY: Harper.

*Wilson, S.M., Rozelle, J.J., & Mikeska, J.N. (2011). Cacophony or embarrassment of riches: Building a system of support for quality teaching. *Journal of Teacher Education, 62*, 383–394.

*Yuan, K., & Le, V. (2012). *Estimating the percentage of students who were tested on cognitively demanding items through the state achievement tests*. Santa Monica, CA: RAND Corporation.

Zernike, K. (2016, February 29). Testing for joy and grit? Schools nationwide push to measure students' emotional skills. *New York Times*. Retrieved from http://www.nytimes.com/2016/03/01/us/testing-for-joy-and-grit-schools-nationwide-push-to-measure-students-emotional-skills.html?_r=0.

About the Authors

Curt Adams is an Associate Professor of Educational Leadership and Policy Studies at the University of Oklahoma and co-director of the Oklahoma Center for Educational Policy. In 2014, he was awarded the Linda Clarke Anderson Presidential Professorship for outstanding contribution to the university, field, and community through research, teaching, and service. He conducts research on the social psychology of school systems, performance measurement, accountability, and improvement science. Recent books include: *Quantitative Research in Education* with Wayne Hoy, and *Collective Trust: Why Schools Can't Improve Without It* with Patrick Forsyth and Wayne Hoy. Recent journal articles include: *Self-regulatory climate: A positive attribute of schools* (Journal of Educational Research); *Self-regulatory climate: A social resource for student regulation and achievement* (Teachers College Record); *Revisiting the collective trust effect in urban elementary schools* (Educational Administration Quarterly); and *Collective trust: A social indicator of instructional capacity* (Journal of Educational Administration).

Timothy G. Ford is an Assistant Professor of Educational Leadership and Policy Studies and a Senior Research Scientist for the Oklahoma Center for Education Policy at the University of Oklahoma. Broadly speaking, he explores the relationship among policy, the social organization of schooling, and school improvement. He specializes in the use of advanced quantitative methods in examining the role of school leadership in creating work environments conducive to teacher collaboration, learning, and development. Tim earned his Ph.D. in Curriculum, Teaching, and Educational Policy at Michigan State University.

Patrick B. Forsyth is a Professor at The University of Oklahoma and co-director of The Oklahoma Center for Education Policy. For 15 years he was Executive Director of the University Council for Educational Administration (UCEA) and simultaneously Executive Secretary for the National Policy Board on Educational Administration. He has served the American Education Research Association (AERA) in various capacities, including on the Executive Council and as Division A Vice President. His policy interests include state accountability models and teacher corps stability. Recent books include two volumes on trust in education: *Collective Trust* (2011) with Curt M. Adams and Wayne K. Hoy (Teachers College Press), and *Trust and School Life* (2014) with Dimitri Van Maele and Mieke Van Houtte (Springer). Other research interests include application of self-determination theory to school problems, work alienation, professionalization, and collective efficacy.

Jordan K. Ware is a postdoctoral fellow for the Oklahoma Center for Education Policy and the University of Oklahoma. He holds a Ph.D. in Education Leadership and Policy Studies from the University of Oklahoma. His research seeks to understand and mitigate the effects of poverty on child development. He is also a former middle school teacher and Teach for America alumnus.

Jentre Olsen is a doctoral student at the University of Oklahoma in the Department of Educational Leadership and Policy Studies. He is a former middle school English teacher and an alumnus of Teach for America. He earned an M.Ed. in Education, Administration, Curriculum, and Supervision from the University of Oklahoma and a B.S. in Business Management from Oral Roberts University. His current research focuses on how principal-teacher interactions support the psychological needs of teachers.

John Lepine Sr. is a doctoral student at the University of Oklahoma in the Department of Educational Leadership and Policy Studies. He is a Teach for America alumnus and currently teaches 8th and 9th grade English. He has an M.Ed. in Educational Administration, Curriculum, and Supervision from the University of Oklahoma and a B.A. in Economics from the University of Tulsa.

Laura Barnes is an Associate Professor of Educational Research and Evaluation and Senior Research Scientist in the Center for Educational Research and Evaluation (CERE) at Oklahoma State University. Her research interests include measurement and assessment issues in K–12 and higher education.

Jam Khojasteh is an Assistant Professor of Research, Evaluation, Measurement, and Statistics and Senior Research Scientist in the Center for Educational Research and Evaluation (CERE) at Oklahoma State University. His research interests include applied and methodological studies in structural equation modeling.

Mwarumba Mwavita is the Director of the Center for Educational Research and Evaluation (CERE) in the College of Education, Oklahoma State University. In addition, he is a faculty member of Research, Evaluation, Measurement and Statistics (REMS) in the School of Educational Studies (SES). His area of research includes longitudinal data analysis using generalized linear and multilevel models and applications of these models in educational and social behavioral research.

LEARNING POLICY INSTITUTE

1530 Page Mill Road, Suite 200
Palo Alto, CA 94304
p: 650.332.9797

1301 Connecticut Avenue, Suite 500
Washington, DC 20036
p: 202.830.0079
www.learningpolicyinstitute.org

The Learning Policy Institute conducts and communicates independent, high-quality research to improve education policy and practice. Working with policymakers, researchers, educators, community groups, and others, the Institute seeks to advance evidence-based policies that support empowering and equitable learning for each and every child. Nonprofit and nonpartisan, the Institute connects policymakers and stakeholders at the local, state, and federal levels with the evidence, ideas, and actions needed to strengthen the education system from preschool through college and career readiness.