



NEXT GENERATION SCHOOL ACCOUNTABILITY

A Report Commissioned by the
Oklahoma State Department of Education



“This is a very thoughtful, humane system—potentially one that other states might want to emulate.”

David Berliner, commenting on "Next Generation School Accountability"

CONTRIBUTING AUTHORS

The contributing authors are responsible for the accuracy of the information, interpretations, and claims made herein; the opinions expressed are theirs and they do not necessarily represent the views of The University of Oklahoma or Oklahoma State University.

THE OKLAHOMA CENTER FOR EDUCATION POLICY (UNIVERSITY OF OKLAHOMA)

Curt M. Adams, Senior Research Scientist, Project Lead A

Patrick B. Forsyth, Senior Research Scientist

Timothy G. Ford, Senior Research Scientist

Jordan K. Ware, Research Scientist

Jentre J. Olsen, Research Associate

John A. Lepine, Research Associate

THE CENTER FOR EDUCATIONAL RESEARCH AND EVALUATION (OKLAHOMA STATE UNIVERSITY)

Laura L.B. Barnes, Senior Research Scientist

Jam Khojasteh, Senior Research Scientist

Mwarumba Mwavita, Senior Research Scientist, Project Lead B

REVIEWERS

The following national experts were contracted to review drafts of this report and they provided extensive written and electronic feedback on the accountability design and the accuracy of analyses and claims. The Oklahoma policy team expresses gratitude to these distinguished scholars for their careful and critical recommendations.

Peter W. Cookson Jr., principal researcher and director of The Equity Project at the American Institutes of Research. Dr. Cookson teaches courses in educational policy, inequality, and social innovation in the Department of Sociology at Georgetown University. He is the author or co-author of more than 15 books on education reform and policy.

David C. Berliner, Regents' Professor of Education Emeritus at Arizona State University. He is the winner of numerous awards, most notably the Brock award and the AERA award for distinguished contributions to education, the E. L. Thorndike award from the APA for lifetime achievements, and the NEA "Friend of Education" award for his work on behalf of the education profession. Professor Berliner has authored more than 200 published articles, chapters, and books.

Linda Darling-Hammond, Stanford University Professor of Education Emeritus and Faculty Director of the Stanford Center for Opportunity Policy in Education. She is a former president of the American Educational Research Association and member of the National Academy of Education as well as the American Academy of Arts and Sciences. Her research and policy work focus on issues of educational equity, teaching quality, and school reform.

William A. Firestone, Distinguished Professor in the Rutgers Graduate School of Education. His research includes topics such as the effects of leadership, professional development, student testing, and teacher evaluation on how teachers teach. He is a member of the American Sociological Association, the Politics of Education Association, and a former Vice-President for Section L of the American Educational Research Association.

Jane Hannaway, Institute Fellow at the American Institutes of Research. Dr. Hannaway is an organizational sociologist whose work focuses on the effects of education reforms on student outcomes as well as on school policies and practices. She is a professor in the McCourt school of Public Policy at Georgetown University and previously served on the faculty of Columbia, Princeton, and Stanford Universities.

Barbara Plake, Professor Emeritus at the University of Nebraska. She was a Distinguished University Professor and Director of the Buros Center for Testing at the University of Nebraska-Lincoln. She served as co-chair of the Joint Committee for the Revision of the Standards for Educational and Psychological Testing and is former President of the National Council on Measurement in Education. She has published over 100 research articles and is the recipient of the Association of Test Publisher's Career Achievement Award in 2005 and the 2006 Career Award from the National Council on Measurement in Education.

Karen Seashore, Regents Professor in the University of Minnesota's Department of Organizational Leadership, Policy, and Development. Dr. Seashore has served as the director of the Center for Applied Research and Educational Improvement and been honored with the Roald F. Campbell Lifetime Achievement Award from the University Council for Educational Administration.

A report produced by the staff of

The Oklahoma Center for Education Policy (OCEP), University of Oklahoma and

The Center for Educational Research and Evaluation (CERE), Oklahoma State University

Cite this report as: Adams, C.M., Forsyth, P.B., Ford, T.G., Ware, J.K., Barnes, L.B., Khojasteh, J., Mwavita, M., Olsen, J.J., & Lepine, J.A. (2015). Next Generation School Accountability. A Report Commissioned by the Oklahoma State Department of Education. Oklahoma Center for Education Policy (The University of Oklahoma) and The Center for Educational Research and Evaluation (Oklahoma State University).

December, 2015

©OCEP and CERE

Corresponding author: Curt M. Adams, curt.adams-1@ou.edu

The University of Oklahoma

Jeannine Rainbolt College of Education

4502 E. 41st St.

Tulsa, OK 74135

TABLE OF CONTENTS

Contributing Authors and Reviewers.....	3
Executive Summary.....	6
Introduction.....	8

PART ONE: EVALUATION OF A-F LETTER GRADES.....9

Limitation One: A Large Conceptual Gap Between A-F Grades and College and Career Readiness.....	12
Limitation Two: Transparency in Reporting Student Achievement.....	13
Limitation Three: A-F Grades Do Not Explain School Performance.....	15
Part One Conclusion.....	15

PART TWO: NEXT GENERATION ACCOUNTABILITY.....16

A New Goal: College and Career Readiness.....	16
Principle One: Shared Accountability. Responsibility for school success is distributed.....	17
Principle Two: Adaptive Improvement. Take schools where they are and move them forward.....	17
Principle Three: Informational Significance. Put relevant and useful information in front of responsible actors.....	18
A New Paradigm.....	19
A Two-Stage Framework.....	20
Stage One Accountability.....	21
Stage Two Accountability.....	21
Next Generation Design.....	22
Defining the Concepts.....	23
Organizational Capacity.....	23
Home and Community Capacity.....	24
Instructional Capacity.....	24
Learning Capacity.....	24
Deeper Learning and College and Career Readiness.....	25
Part Two Conclusion.....	26

PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE.....28

EQUIP Measurement Guidelines.....	29
Accountability Indicators.....	30
Elementary and Middle Schools.....	31
High Schools.....	32
Improvement Indicators.....	34
Deeper Learning and Formative Indicators.....	34
Instructional Capacity Profile.....	36
Learning Capacity Profile.....	38
Organizational Capacity Profile.....	40
Home and Community Capacity Profile.....	42
EQuIP Online.....	44

PART FOUR: DESIGNING AND IMPLEMENTING NEXT GENERATION ACCOUNTABILITY.....46

Accountability Policy.....	47
Alignment of Standards, Assessments, and Accountability.....	49
School, District, and State Capacity Building.....	50
Part Four Conclusion.....	51

Technical Appendix.....	52
References.....	54

EXECUTIVE SUMMARY

The A-F grading system was not designed with the vision of college and career readiness in mind. A next generation accountability system is needed to ensure that all students are college and career ready upon graduation from an Oklahoma high school. Next generation accountability requires a more holistic assessment of school performance, including detailed and disaggregated information about student test performance and graduation rates in addition to relevant data about school resources and processes that affect student outcomes. The Educational Quality and Improvement Profile (EQiP) is proposed as an improvement to A-F School Report Cards. EQiP provides comprehensive descriptions of schools' resources, processes, and outcomes in order to facilitate school-led improvement, district decision-making, and—where necessary—state intervention.

PART ONE: EVALUATION OF A-F GRADES

A-F School Grades have at least four major limitations:

1. They are not achieving the results that were hoped for. We find declines in **math and reading scores** for a cohort of students from 2012-2014.
2. A-F Grades **do not align** with Oklahoma's goal of producing **college and career ready graduates**. A-F Report Cards do not capture a wide range of the deeper learning competencies necessary for college and career readiness.
3. A-F Grades **are not transparent**. Student performance differences, achievement gaps, and growth within schools are often missed.
4. A-F Grades do not explain school performance. With only outcome data, **schools lack vital improvement information** about their resources and processes.

PART TWO: PRINCIPLES AND FRAMEWORK FOR NEXT GENERATION ACCOUNTABILITY

Next generation accountability is governed by three principles:

1. **Shared Accountability.** Responsibility for school success is distributed.
2. **Adaptive Improvement.** Take schools where they are and move them forward.
3. **Informational Significance.** Put relevant and useful information in front of responsible actors.

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 allocation at the OSDE, putting in place a support infrastructure capable of serving the developmental and resource needs of schools across the State.



PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE (EQuIP)

The Educational Quality and Improvement Profile provides accountability information to determine how well students are meeting deeper learning and college and career ready standards, and improvement information to better understand potential reasons for observed outcomes. Six guidelines govern how data are reported:

1. Outcome indicators should report achievement differences by **student subgroups** and changes in student performance **over time**.
2. Outcome indicators should be capable of identifying **focus schools, priority schools, and reward schools** consistent with criteria set 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**.
6. Indicators and measurement methods should change over time in response to the **continuous evaluation and improvement** of Oklahoma's school accountability framework.

EQuIP examples are included in this section.

PART FOUR: DESIGNING AND IMPLEMENTING NEXT GENERATION ACCOUNTABILITY

A variety of policy reforms at the state and local level will serve to advance the goal of college and career readiness. There are three broad categories of policy change:

1. **Accountability Policy.** A-F Report Cards should be replaced by the EQuIP. EQuIP reports multiple indicators of college and career readiness and deeper learning, with evidence that clearly captures changes in performance and capacity over time.
2. **Alignment of Standards, Assessments, and Accountability.** In order to prepare students for college and careers, Oklahoma needs both curriculum and standards that are aligned with deeper learning. Measuring progress toward these standards requires a new, coordinated system of multiple assessments, both formative and summative.
3. **Capacity Building.** State, district, and school leaders must create a **system-wide culture grounded in "learning to improve."** This requires the development of **strong pedagogical data literacy skills**. Resources should be prioritized for **sustaining ongoing improvement**. **Educator labor market policy** may need adjustment to support the above elements.

INTRODUCTION

This report responds to a request from the Oklahoma State Department of Education for a thorough review of the Oklahoma A-F School Report Cards by The Oklahoma Center for Education Policy at OU and the Center for Educational Research and Evaluation at OSU pursuant to HB 1823.

The centers have been asked to propose improvements to all components of Oklahoma's A-F school grade reports. The policy team shares the legislature's desire that any proposed changes to the accountability metrics have the effect of producing clear, credible, and meaningful information, salient to achieving the goal that all children will graduate from high school college and career ready by 2020 (OSDE, 2015).

We write this report when considerable evidence shows an educational system limping and sputtering under test-based accountability. Nationally, as well as locally, student achievement is not at a level where it needs to be and achievement gaps remain entrenched problems (Mintrop & Sunderman, 2009; National Center for Educational Statistics, 2013). New challenges lurk as well. Results on the international PISA exam, an assessment measuring deeper learning competencies, show no gains during the past decade in math, reading, and science scores for US 15 year olds. Moreover, US 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 and responsibility. Oklahoma citizens want an education 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, they 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 richer spectrum of indicators, well beyond a summary of annual test performance, seems essential to account for performance transparently 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" (2014, p. 4).

What seems clear is that accountability systems reliant on regulatory compliance, such as NCLB and Oklahoma's A-F approach, are insufficient, even for achieving a myopic vision of test score improvement. That evidence is plentiful. More daunting still, the challenge for next generation accountability is aimed at an even loftier goal—universal college and career readiness—a goal the current system was not designed to achieve. So what is the "right accountability system" for Oklahoma now? The purpose of this report is to explore the new goal and imagine an accountability approach suited to the task and the education landscape of our State. To achieve at this higher level, next generation accountability must embrace a wider vision, distribute credible performance information, and build support infrastructure, all the while and most importantly, eliciting the assent, support, and enthusiasm of citizens and educators.

PART ONE: EVALUATION OF A-F GRADES

The implicit theory of action behind A-F grades derives from a belief that public reporting of student performance using a single, summary indicator can motivate schools to raise student test scores. By this logic, improvement originates from external pressure associated with consequences of school grades. Whether or not the use of A-F grades leads to higher achievement remains an open question that we take up in this section. The findings do not justify unequivocal claims about the effectiveness of the A-F accountability system, but the evidence does reveal consequential patterns in student achievement during the past three years.

“Any [school] grade or index hides most of the information you need to have to improve.”
Linda Darling-Hammond, commenting on “Next Generation School Accountability”

PART ONE: EVALUATION OF A-F LETTER GRADES

“...those responsible for developing or implementing test-based accountability systems often assert that these systems will lead to specific outcomes, such as increased educator motivation or improved achievement; these assertions should also be supported by evidence.”

(AERA, APA, NCME, 2014, p. 206)

Figure 1 reports three-year growth rates for OCCT math and reading exams by 2012 school grade and Free or Reduced Lunch (FRL) status). The bars are set at 0 for no growth and extend upward for average gains and downward for average declines. The blue line is for math and the orange is for reading. Given the resources, time, pressure, and costs associated with the A-F 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 400-990 with a standard deviation around 90 points.

We do not find any meaningful gains in student achievement between 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 question correct on a fifty-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.

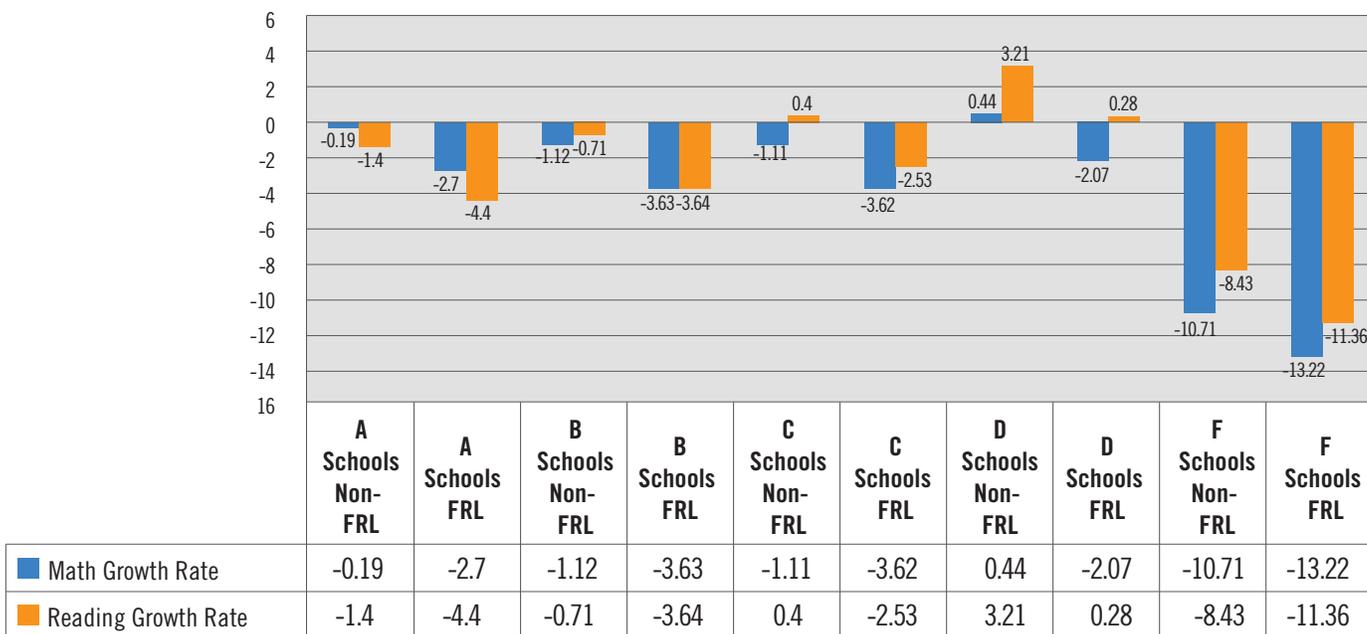


Figure 1: Three-year growth rate in math and reading by 2011-2012 school grade and free-and-reduced lunch (FRL) status.

Drops in test scores are also reflected in stagnant and declining achievement trends across the three testing periods. As seen in **Figures 2** and **3**, 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.

TO CONCLUDE

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 the goal of improving test scores, we would expect to find modest gains following the implementation of the A-F 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 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.

In contrast with the theory of action implicit in Oklahoma's A-F system, an alternative approach, derived from credible research, indicates that information used in the service of quality improvement is most effective when it fuels knowledge creation and capacity building (Deming, 2000; Langley et al., 2009). This approach addresses a pervasive problem throughout the educational system—that a lack of knowledge on what and how to improve has diminished the capacity of schools and school systems to produce better outcomes reliably at scale (Bryk et al., 2015). Responding to this problem, accountability systems need to provide clear and relevant information by which accurate knowledge about school outcomes, and the factors behind the measured results, can be generated (Rothstein, Jacobson, & Wilder, 2008). Limitations of A-F grades, however, make inferences about school and student progress toward college and career ready standards invalid.

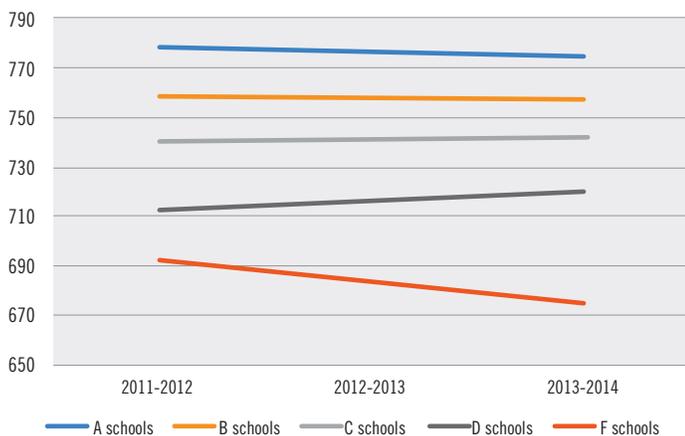


Figure 2: 2012-2014 achievement growth in reading by 2011-2012 school grade

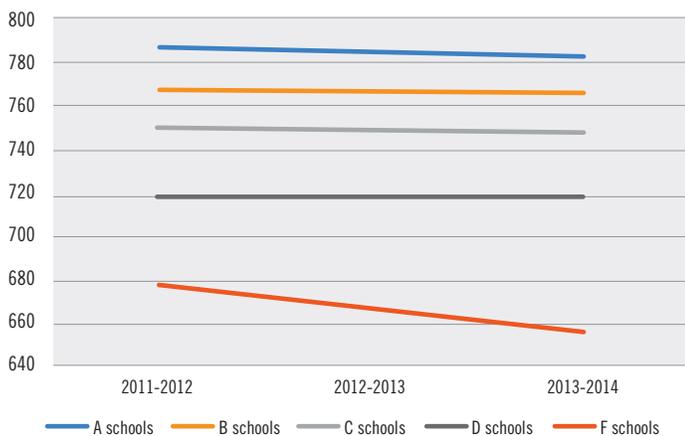


Figure 3: 2012-2014 achievement growth in math by 2011-2012 school grade

PART ONE: EVALUATION OF A-F LETTER GRADES

LIMITATION ONE: A LARGE CONCEPTUAL GAP BETWEEN A-F GRADES AND COLLEGE AND CAREER READINESS

It is in the public interest to know what conclusions can be reasonably and accurately drawn from an A-F grade. With respect to the vision of college and career readiness, A-F grades exclude most of the relevant knowledge and skills associated with the concept. College and career readiness includes traits, dispositions, skills, knowledge, and competencies not measured by the low-level cognitive tests that make up the State's curricular assessments.

A call for college and career ready graduates is a call for deeper learning (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 non-routine tasks. Deeper learning outcomes, which are distinct, yet related to processes, include diverse cognitive, intrapersonal, and interpersonal competencies that guide human thinking and adaptive behavior. College and career readiness represents deeper learning competencies that high school graduates need for effective functioning in a post-industrial society. These competencies vary some by 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 A-F grades depend on alignment between Oklahoma Core Curriculum Tests (OCCT) and End of Instruction (EOI) exams and the cognitive and non-cognitive competencies associated with deeper learning and college and career readiness. The OCCT and EOI exams test a very small subset of the competencies that businesses and colleges look for in graduates (Figure 4). At best, summative school grades may yield information about basic content knowledge and skills, but they fail to capture mastery of concepts and ideas, creativity and imagination, critical thinking and problem solving, interpersonal abilities and effective communication, and learning mindsets.

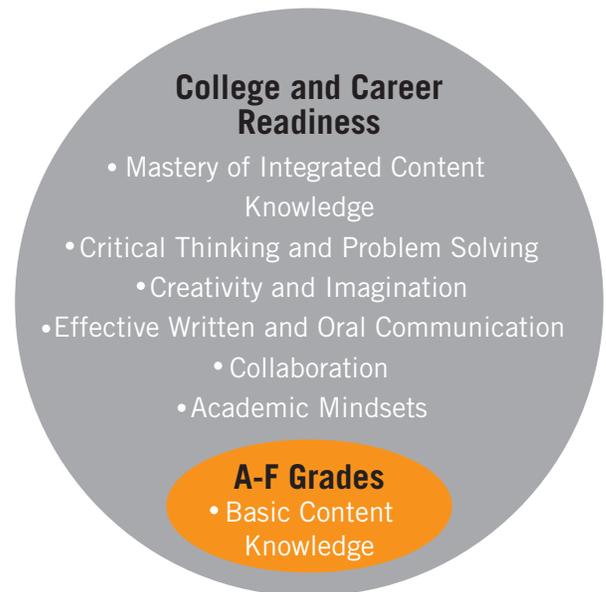


Figure 4. Alignment between A-F grades and college and career readiness.

It is possible for a set of formative and summative assessments to measure competencies of deeper learning and college and career readiness (Conley, 2014), but the assessment system in Oklahoma does not do this. State assessments test mostly low-level cognitive functioning; test questions do not assess how well students can transfer basic knowledge and skills to new situations, problems, or everyday tasks. Knowledge and skills of deeper learning are currently underrepresented. This results in an accountability indicator that does not report on student abilities to think through abstract, complex, and non-routine problems, or to perform investigations of real world tasks that require time to research, problem solve, and process information from multiple sources.

Should we be concerned about an overemphasis of low-level test questions? In short, yes! It is generally the case that what is measured gets taught. Researchers with the Standards Company found that in Oklahoma only a small percentage of assignments in English and math required strategic and extended thinking. In English an average of 2 percent of assignments in elementary schools, 10 percent in middle schools, and 13.5 percent in high schools required strategic and extended thinking. For math, the average percentage was 1 percent in elementary schools, 0.33 percent in middle schools, and 3.5 percent in high schools (Standards, 2008).

Deeper learning and college and career readiness are the right expectations for schools. But in setting this vision, the State has a responsibility to invest in systems that support progress toward higher expectations. Conceptually, A-F grades do not live up to this responsibility. Currently, no valid conclusions about school or student progress toward deeper learning and college and career readiness can be drawn confidently from A-F grades.

LIMITATION TWO: TRANSPARENCY IN REPORTING STUDENT ACHIEVEMENT

Transparency in student outcomes is a good thing. A-F grades, however, hide more than they actually reveal about the learning and growth happening in schools. If citizens know only a school's A-F grade, they are left with an incomplete and distorted picture of school quality and student achievement. We show how this occurs by describing three measurement problems.

- 1. Grades hide variation in student performance that is needed to understand how all students are progressing.** Figure 5 illustrates the problem of judging student performance by the school grade. The graphs compare test score distributions by A-F grades. The blue plot represents student scores within the specific letter grade and the plot is outlined in the orange 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 letter grades, we lose sight of the fact that the grade does not reflect the performance of many students within schools. Many students in F and D 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.

- 2. Grades mask student growth within schools.** Because growth represents 50 percent of the composite grade, it is reasonable to assume that a high grade reflects strong student growth and a low grade poor growth. Such assumptions are untenable. When examining the distribution of test scores based on the prior year proficiency category, we do not find much difference in the reading scores across A-F grades.

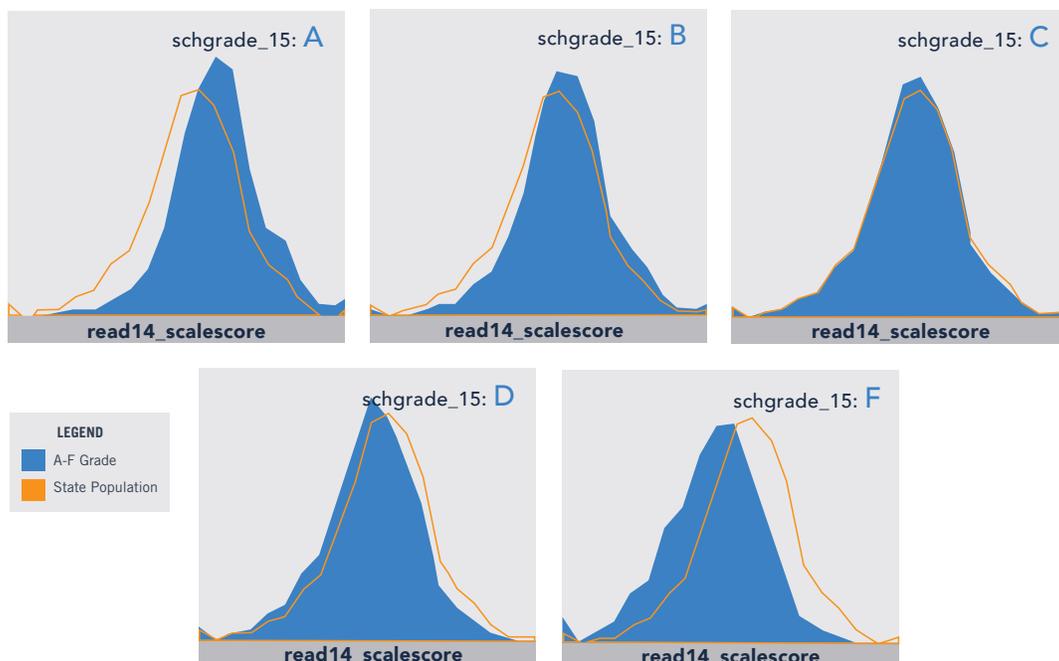


Figure 5. Distribution of 2014 reading scores For elementary and middle school students.

PART ONE: EVALUATION OF A-F LETTER GRADES

Figure 6 is a “box and whiskers plot” of 2014 reading scores by school grade and student proficiency category from the 2013 reading test. The colored boxes represent the interquartile range (i.e., middle 50 percent of scores) for the distribution. That is, the top end of the box is the 75th percentile and the bottom end is the 25th percentile. The line represents the median score. Whiskers account for scores that are beyond the 75th and 25th percentiles, and the symbols are outliers. The purpose of the box and whiskers plot is to show the variability (i.e., spread) of scores for students in A, B, C, D, and F schools. What is remarkable is the lack of differences by 2013 proficiency category in the distributions by A-F school grades.

As evident in the figure, student performance as defined by proficiency categories (i.e., unsatisfactory, limited knowledge, proficient, and advanced) is essentially the same across schools, regardless of school letter grade. That is, the distributions of scores in A schools are nearly the same as distributions in B, C, D, or even F schools.

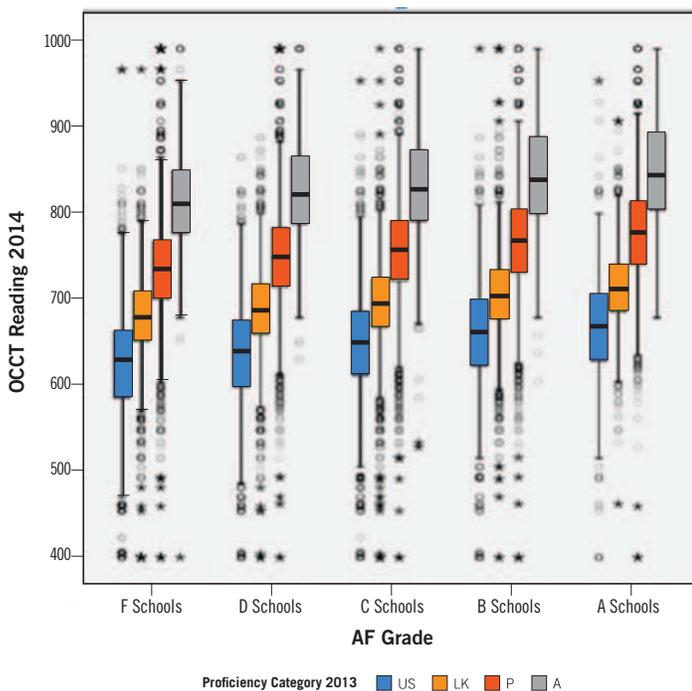


Figure 6. Box plot of 2014 reading scores by 2014 school A-F Grade and 2013 proficiency category

In other words, students in A and B schools who scored limited knowledge or unsatisfactory were not any more likely to have better 2014 reading scores than limited knowledge and unsatisfactory students in C, D, or F schools. Moreover, students in F and D schools scoring proficient and advanced in 2013 were just as likely to stay above proficiency in 2014 as students in A and B schools. Stated simply, letter grades are incapable of revealing changes in student achievement from one year to the next.

3. A-F Grades lack transparency in reporting achievement gaps

(**Figure 7**). Federal regulations require states to account for achievement equity, but achievement gaps are not given any weight in the calculation of Oklahoma’s A-F letter grades. This decision is not without consequence. Notice the large difference in reading scores between FRL and Non-FRL students in A and B schools. For 2012 and 2013, FRL students scored approximately 0.47 standard deviation units lower than non-FRL students. In 2014, this gap increased to 0.54 standard deviation units. This amounts to a difference of approximately 45 scale score units, a sizable difference between these groups.

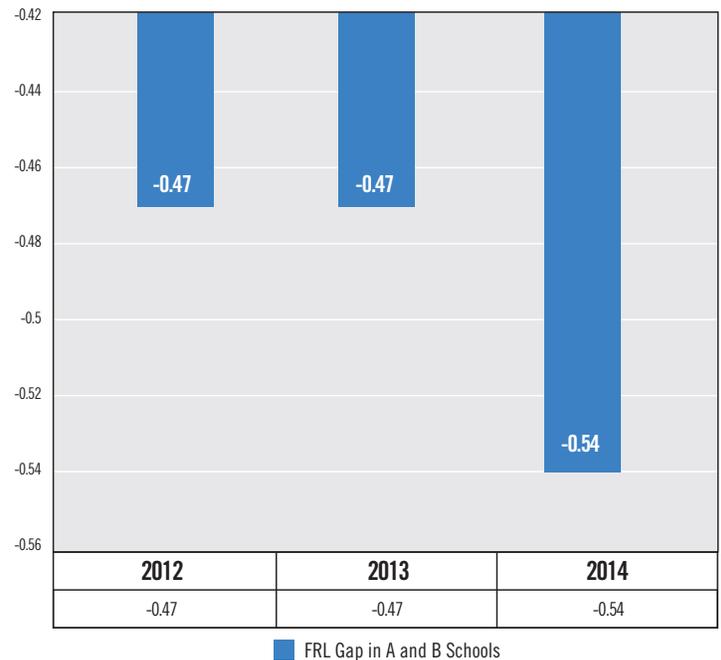


Figure 7. Average Achievement gap in A and B schools between 2012-2014

LIMITATION THREE: A-F GRADES DO NOT EXPLAIN SCHOOL PERFORMANCE

An appeal of accountability systems like A-F 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, A-F grades 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, policy makers, 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

We set out to determine if A-F grades are capable of supporting efforts to raise the quality of teaching and learning in schools. The evaluation shows that student achievement has not improved under the A-F grading system. Not only have test scores stagnated or declined generally, performance drops have been most severe among low-income students. With a vision of college and career readiness firmly established, we are cynical about the A-F indicator's capacity to transform a stagnant educational system. The most serious design flaws that need to be addressed are summarized below.

1. *There is a large conceptual gap between A-F grades and deeper learning and college and career readiness.* The A-F accountability system purports 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. Moreover, deeper learning and college and career readiness are multi-dimensional concepts that include a broad set of competencies not currently measured by State assessments. Current curricular standards, assessments, and the A-F accountability system were not designed with deeper learning and college and career ready graduates in mind. 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 post-industrial workforce.

2. *Letter grades hide achievement differences within schools. Achievement variation is a natural product 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. A-F grades 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. A-F grades do not provide this information.
3. *A related problem involves the false assumption that group indicators reflect the performance of individuals within the group.* Social scientists refer to this as an ecological fallacy. So 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 refute this assumption. Of course, some students in F schools have low test scores, but assuming that a particular student in an F school is failing does not square with evidence. 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, but this variation remains hidden to educators and the public.
4. *Letter grades obscure achievement growth. Points awarded for moving above proficiency, or remaining in the proficiency categories, mask actual changes in achievement.* 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.
5. *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 is the product of numerous factors that are not reflected in aggregated test scores. In no other sector—banking, healthcare, manufacturing—would it be acceptable to make strategic decisions on outcome-only indicators. Yet A-F grades rest on the assumption that policymakers and school leaders can diagnose the effectiveness of schools and the health of the educational system based on a single summative outcome indicator.

PART TWO: NEXT GENERATION ACCOUNTABILITY

A NEW GOAL: COLLEGE AND CAREER READINESS

The first generation of accountability policies for the State of Oklahoma exposed vast, inequitable differences in student test scores across and within schools. As we have demonstrated, accountability under this policy was not effective at building capacity within school systems to raise achievement and equalize achievement distributions. Greater capacity lies at the heart of quality improvement (Darling-Hammond et al., 2005; Harris, 2011), and with Oklahoma adopting the expectation that by 2020 every student will graduate college and career ready, the next generation of 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 both the cognitive and non-cognitive competencies expected for success in a world economy. What kind of accountability framework will facilitate the State's progress toward this newly identified and specified vision? Three principles derived from the education policy and accountability literature guide 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 (Louis et al., 2010; Darling-Hammond et al., 2014). 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 Oklahoma schools can attract and retain their share of the most talented, appropriately trained, and competent teachers. The State Department of Education is responsible for, among other things, identifying schools that have failed their communities catastrophically and providing support so that children can realize their full potential. The department must also provide the resources, expertise, and know-how to assist schools that need help to achieve the readiness goal. School boards and superintendents share responsibility for maintaining a qualified, competent, and stable teaching corps in all schools, but 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 are performing relative to the vision of college and career readiness. As Darling-Hammond et al., 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 which only need to "redirect efforts" in order 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, the State's approach to working with schools is contingent on their current assessed conditions and their need for support, expert assistance, and other resources.

PART TWO: NEXT GENERATION ACCOUNTABILITY

PRINCIPLE THREE: INFORMATIONAL SIGNIFICANCE. PUT RELEVANT AND USEFUL INFORMATION IN FRONT OF RESPONSIBLE ACTORS.

What is sought in next generation accountability systems is ambitious and includes collecting and distributing significant information salient to the work and interests of all stakeholders. A broad palate 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, 2002). 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 Oklahoma's schools. It should also address the different information needs of State, district, school site leadership, teachers, and parents. Information about resources, processes, and a variety of outcomes can enable sense-making 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 OSDE 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 movement toward the State's 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 Oklahoma's 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 Oklahoma’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. **Table 1** summarizes differences between these two accountability frameworks.

FIRST GENERATION ACCOUNTABILITY FRAMEWORK

- No Child Left Behind
- Focus on improving test scores
- Compliance to meet proficiency standards
- Schools accountable to State
- Summative emphasis
- Single composite indicator of student performance
- One-size-fits-all interventions
- Replace teachers and leaders

NEXT GENERATION ACCOUNTABILITY FRAMEWORK

- All children healthy and thriving
- Focus on fostering deeper learning
- Capacity building
- Shared accountability
- Formative emphasis
- Multiple indicators of whole system performance
- Adaptive interventions
- Retain and support quality teachers and leaders

Table 1. *Contrasting first and next generation accountability*

PART TWO: NEXT GENERATION ACCOUNTABILITY

A TWO-STAGE FRAMEWORK

Consistent with the statutory directives of the Oklahoma legislature and policies adopted by the Oklahoma State Department of Education, next generation accountability has two primary stages and objectives. The second objective is embedded within the first but 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, 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.



STAGE ONE ACCOUNTABILITY

The central goals of deeper learning and college and career readiness are the foundation of this formative stage and constitute the Oklahoma State Department of Education's (OSDE) premier work focus. This stage has the State providing every school and school district with timely, longitudinal information about the quality of school resources, processes, and outcomes. Newly conceptualized and designed school profiles will enable education stakeholders to see clearly how student groups are performing over time on measures of deeper learning and college and career readiness. But equally important, the profiles gauge key school and community resources and processes, enabling explanations for school outcomes. System-wide indicators spotlight weaknesses as well as suggest responsibility for action. Stage One will undoubtedly require OSDE to broaden the scope of the existing school support office and create an infrastructure that expands and enhances the expert support team 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 the State's information gathering process, 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 particular challenges and managing their unique improvement journey. The OSDE is on call to provide prompt and expert assistance and support when it is requested by any school in the State.

STAGE TWO ACCOUNTABILITY

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 federal guidelines calling for the identification of focus and priority schools so that urgent interventions can address unacceptable outcomes. It also is the basis of 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 the OSDE to report school progress toward Annual Measurable Objectives (AMOs).

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. The OSDE will oversee an investigatory process that includes participation of community and school representatives. The rich profile of longitudinal system indicators reported to schools for Stage One simultaneously provides significant information needed by the OSDE to make a determination 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. The objective of the intervention, in cases of serious malfunction, 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).

PART TWO: NEXT GENERATION ACCOUNTABILITY

NEXT GENERATION DESIGN

How will next generation accountability change the mindset and approach used in Oklahoma to address school failure and stimulate unprecedented commitment to educational excellence across the State? The shift from a preoccupation with failure to success is not merely a semantic one; instead, it directs the system and its momentum 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 8 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 judged 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. Policy makers, local school professionals, and 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.

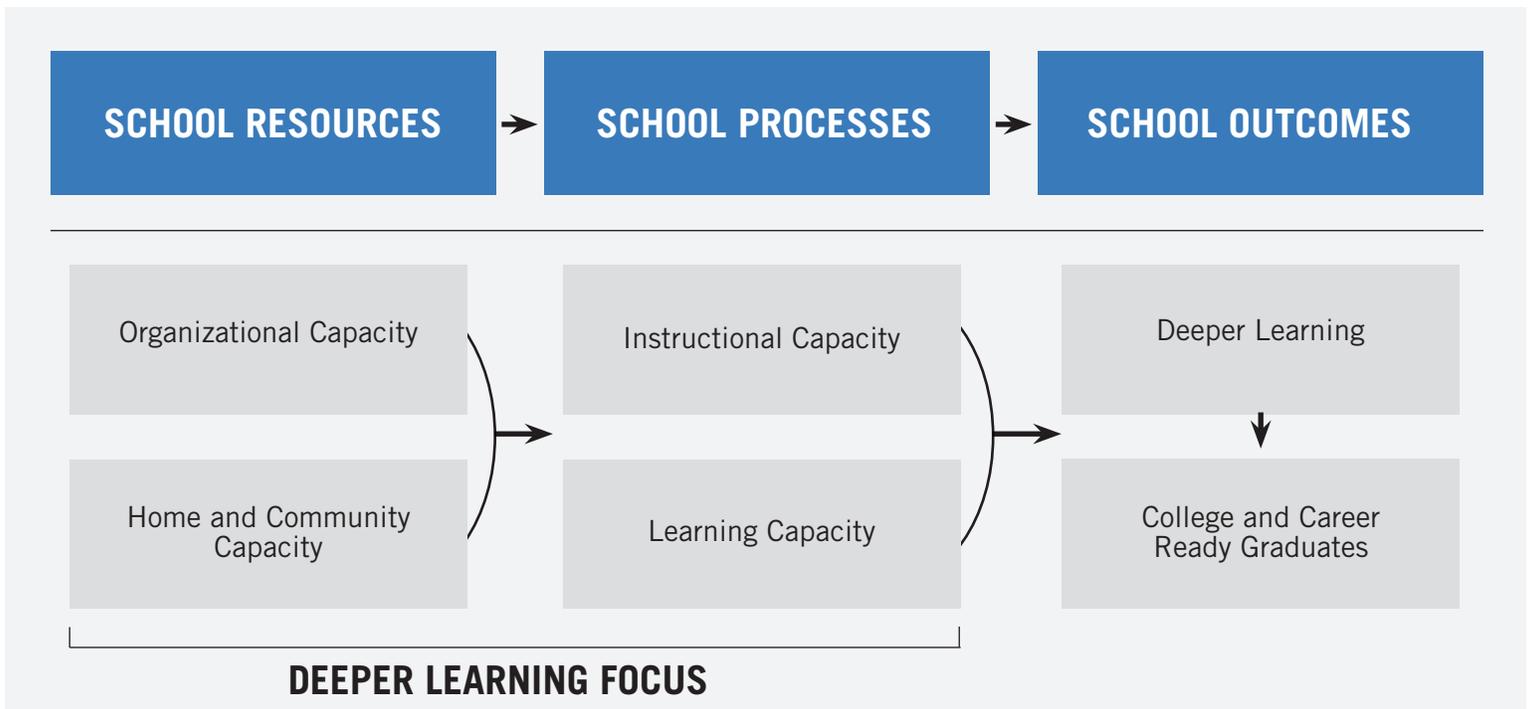


Figure 8. Conceptual framework for next generation accountability.

DEFINING THE CONCEPTS

In this section, we define the concepts included in the accountability framework before discussing in detail its proposed function. An abbreviated description of each concept may be found in [Table 2](#).

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 & CAREER READINESS	The preparation of high school graduates to enter a career, extended training, or specialized education without need for remediation.

Table 2. Key concepts in next generation accountability.

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 capacity 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 resides not 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.

PART TWO: NEXT GENERATION ACCOUNTABILITY

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, p. 951). Generally, schools have not excelled at developing this capacity, especially in high poverty communities. Nevertheless, these environmental factors and resources, while 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 quite directly relevant to instruction, such as its 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 in place 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 Oklahoma 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 capacity 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 is key, as is knowledge about the extent to which the school promotes a self-regulatory climate (Adams, Forsyth, Dollarhide, Miskell, & Ware, 2015; Schunk & Zimmerman, 2008). Armed with these understandings, the school's staff can work to enhance school learning capacity while adjusting the instructional system to meet them where they are.

DEEPER LEARNING AND COLLEGE AND CAREER READINESS

Deeper learning has both a process and outcome component. Deeper learning processes reflect the act of transferring knowledge and skills developed in one setting to new situations, contexts, and problems (National Research Council, 2012). Outcomes consist of cognitive, intrapersonal, and interpersonal competencies that individuals draw on in a particular way depending on the circumstance. The three competency domains are foundational, and interconnected, 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 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 ability to communicate and collaborate, trust, empathy, conflict management, leadership, social awareness, and leadership.

The State of Oklahoma has determined that the goal of its public school system is effectively to ready its graduates for advanced education or careers. 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 high school, but are graduating with knowledge and skills for an effective transition to post-secondary opportunities. The exit indicators of graduation rates, ACT scores, and post-secondary enrollment are the final evidence warranting claims of college and career readiness.

PART TWO: NEXT GENERATION ACCOUNTABILITY

PART TWO CONCLUSION

Building on conceptual inadequacies and technical difficulties found in Oklahoma's current school accountability approach, Part Two seeks 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 the State and its public education system to instruction that produces "deeper learning" and high school graduation standards that are aligned with the emerging world economy and readying Oklahoma's graduates for advanced training or careers.

Three principles are derived from the accountability, organizational, and school reform literatures 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 the State's institutions, professionals, and citizens. (2) The principle of Adaptive Improvement recognizes that any accountability system should account for differences in the capacities of Oklahoma's schools to pursue standards reflective of 21st century skills and competencies. (3) The principle of Informational Significance points to 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 Oklahoma's school accountability commitment. 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 in catastrophic failure. Next generation accountability requires a significant shift in resource allocation at the OSDE, putting in place a support infrastructure capable of serving the developmental and resource needs of schools across the State.

Both stages depend on a view of schools as systems whose outcomes are results of 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 the State to identify with confidence those schools in need of expert support and additional resources.

“At the end of the day, the argument of the Next Generation School Accountability system is that students will graduate from high school college and career ready, and, most importantly, they will be deep learners capable of critical and creative thinking. This is the foundation of an economically vibrant Oklahoma because in the new technology-driven, demand economy, higher order skills and the disposition to innovate will separate those who succeed from those who are likely to flounder.”

-Peter W. Cookson, Jr., Georgetown University



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 student outcomes far exceeding A-F or any 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 conditions and resources in schools that need to change for observed outcomes to improve.

EQUIP IS COMMITTED TO THE FOLLOWING MEASUREMENT GUIDELINES:

- 1) Outcome indicators should report achievement differences by student subgroups and changes in student performance over time.
- 2) Outcome indicators should enable the identification of focus schools, priority schools, and reward schools consistent with criteria set for federal waiver requirements.
- 3) Process and resource indicators should be scientifically defensible and measure conditions, attitudes, structures, and behaviors.
- 4) Indicators should be assessed with appropriate frequency and minimal disruption.
- 5) Credible scientific evidence substantiating the validity and reliability of measures should be clearly reported.
- 6) Indicators and measurement methods should change over time in response to the continuous evaluation and needs of Oklahoma's school accountability framework.

The EQuiP profile begins with accountability indicators so that educators and the public can judge how well schools are moving students toward proficiency of deeper learning and college and career ready standards. Formative indicators of student non-cognitive skills and attributes follow accountability data. Process and resource indicators round out the profile, providing additional evidence to help state, community, and school leaders make meaning of the observed outcomes.



PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE

ACCOUNTABILITY INDICATORS: ELEMENTARY AND MIDDLE SCHOOLS

Accountability indicators for elementary and middle schools align with the method used by the State to report school progress toward Annual Measureable Objectives (AMOs). Indicators come from student test scores on state curricular exams. Use of student test scores for accountability purposes requires an assessment system of sufficient rigor to measure student mastery of concepts and ideas, critical thinking and problem solving, creativity and imagination, and effective communication. The Oklahoma State Department of Education should consult with leading experts in this area to identify and adopt an assessment system that measures the competencies of deeper learning and college and career readiness (e.g., OECD; RAND; the National Center for Research on Evaluation, Standards, and Student Testing; and Stanford Center for Opportunity Policy in Education).

Assuming cognitively rigorous curricular assessments, EQuIP lays out a comprehensive profile of school progress toward deeper learning standards by reporting student performance, student growth, and achievement gaps in concise and easy to understand 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 school profile enables educators, policy makers, and the public to determine definitively if a school's students are achieving proficiency standards.

Figure 9 presents an example accountability page 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. A-F grades hide this information in a formula that masks the actual proficiency distribution of students. With EQuIP, educators and the public know the actual percentage of students scoring advanced, proficient, limited knowledge, or unsatisfactory. EQuIP also provides the AMO target for 2020.

Pie graphs are used to display data on **student growth**. For elementary schools, graphs report the change 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 grades, it is not clear what proportion of students moved, in either direction, across proficiency bands. The formula hides this information from the public. In contrast, EQuIP makes known achievement changes, enabling discernments about the progress of lower and higher performing students alike.

Line graphs are used to track five-year trends in **achievement gaps**. EQuIP displays changes in average achievement for subgroups with 30 or more students, as well as changes in achievement gaps. This allows schools and the public to track, longitudinally, 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. A-F grades do not calculate nor report information about achievement gaps, leaving educators and the public unaware of achievement patterns 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.

DEEPER LEARNING

SAMPLE ELEMENTARY SCHOOL



Deeper learning refers to the expectation that schools develop in students cognitive and non-cognitive 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 deeper learning.

COMPOSITION OF SCHOOL

Asian: 3% Black: 29% Hispanic: 63% White: 1%
 American Indian: 3% Multiple Races: 1%

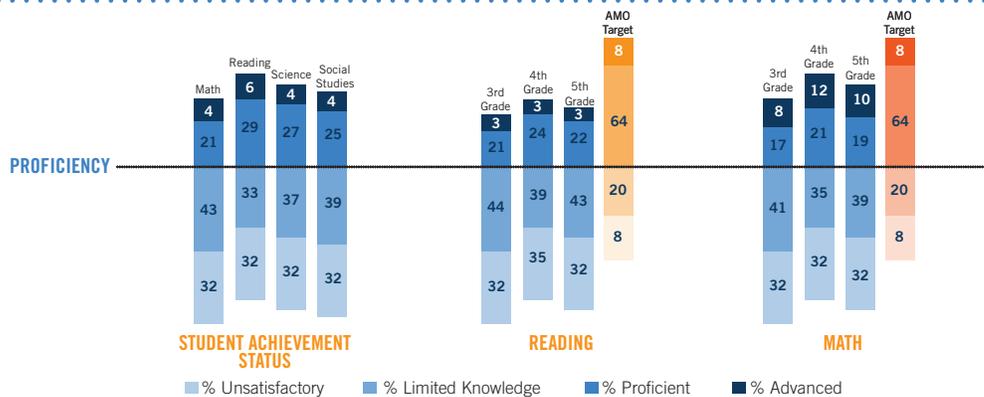
English Language Learners.....10%
 Students with Special Needs.....24%
 Free and Reduced Lunch.....23%
 School Size.....900

Attendance
 Student Attendance.....92%
 Teacher Attendance.....97%

Staff Experience
 % of Teachers with 3 or More Years of Experience.....75%
 Years of Principal Experience at This School.....8.9

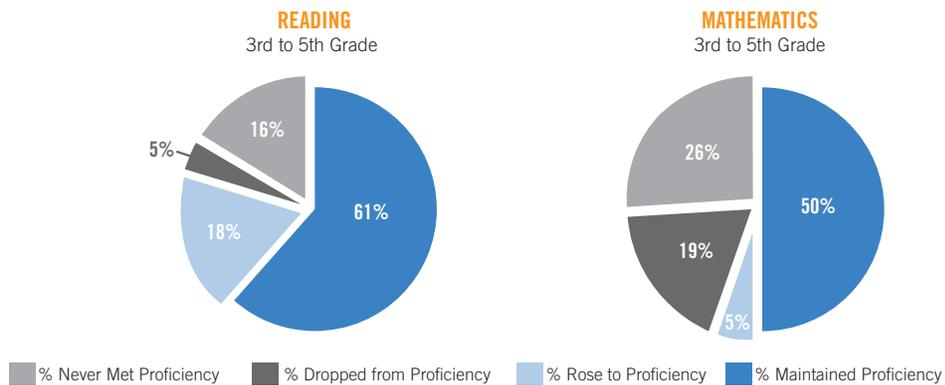
STUDENT PERFORMANCE

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 percent 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 2013 to 5th grade in 2015.



ACHIEVEMENT GAPS

Line graphs report changes in reading and math achievement gaps for student subgroups with 30 or more students.

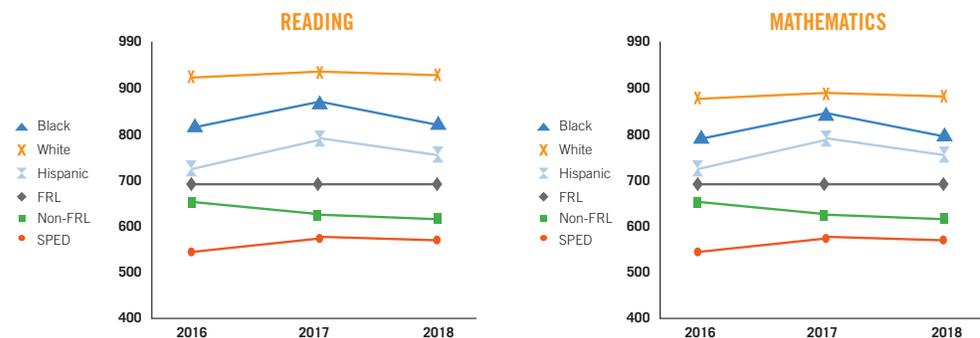


Figure 9. An example of an accountability page for an elementary/middle school.

PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE

ACCOUNTABILITY INDICATORS: HIGH SCHOOLS

Accountability indicators for high schools align with college and career ready expectations and are consistent with Oklahoma's criteria for identifying schools making progress toward AMOs. The set of accountability indicators includes the percentage of students graduating high school with knowledge and skills compatible 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 10 is an example accountability profile for high schools. A line graph is used to display **graduation rates** for the last four years. The average graduation rate is based on the four-year adjusted cohort formula; it is reported for all students and student subgroups with thirty or more students. With A-F 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. EQuIP does 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 **ACT scores** as a “readiness” indicator. For each subject test, a bar graph shows the percentage of students scoring at or above benchmark scores, within 2 points of the benchmark, or 3 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 two-year or four-year colleges and career/technical school (ACT, 2015). In addition to benchmark scores, EQuIP graphs composite scores over four years by student subgroups with thirty or more students. Contrast this with A-F grades, which do not include a “readiness” indicator in their calculation, nor information about student preparedness for the rigor of college and career training.

Graduation rates and ACT scores are useful indicators but they leave open the question of access to and enrollment in post-secondary education. EQuIP reports **post-secondary** as the percentage of students enrolling in college (by four-year and two-year), career training, or the military. In accounting for post-secondary access, 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.

EQuIP also provides an early warning signal to high schools by accounting for student progress toward graduating college and career ready. **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.

COLLEGE & 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 the college and career readiness.

COMPOSITION OF SCHOOL

Asian: 3% Black: 29% Hispanic: 63% White: 1%
 American Indian: 3% Multiple Races: 1%

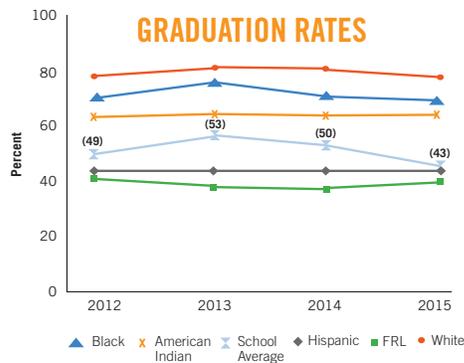
English Language Learners.....10%
 Students with Special Needs.....24%
 Free and Reduced Lunch.....23%
 School Size.....900

Attendance
 Student Attendance.....92%
 Teacher Attendance.....97%

Staff Experience
 % of Teachers with 3 or More Years of Experience.....75%
 Years of Principal Experience at This School.....8.9

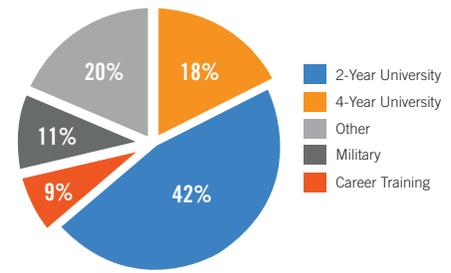
GRADUATION RATES & POST-SECONDARY

The line graph tracks graduation rate by student subgroup over 4 years. The pie graph indicates post-secondary options for the class of 2015.



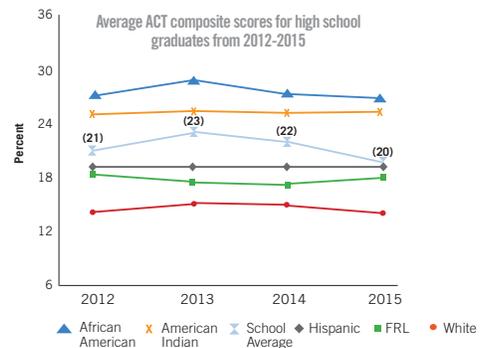
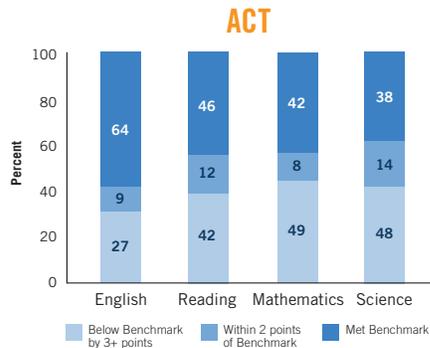
POST-SECONDARY OPTIONS

% of graduating class of 2015 enrolled in:



ACT PERFORMANCE

This bar graph indicates the percentage of students achieving the benchmark score for each subject test. The line graph tracks changes in composite scores by student 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

92%

earned enough credits in 9th grade to be on-track for graduation

83%

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

ACT ASPIRE

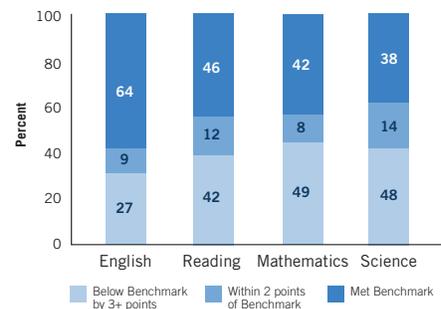


Figure 10. An example of an accountability page for high schools.

PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE

IMPROVEMENT INDICATORS: DEEPER LEARNING FORMATIVE INDICATORS

EQuIP reports improvement indicators separate from accountability. Improvement indicators are designed to inform improvement decisions by capturing elements of deeper learning not measured with State curricular tests, as well as to provide information that can account for plausible reasons for observed outcomes. Proposed indicators 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 11 is an example page of formative, high school indicators for deeper learning and college and career readiness. Deeper learning process indicators focus on the degree to which students have opportunities to apply basic knowledge and skills to real-world situations. Deeper learning processes can be demonstrated in different ways. One way is to follow the lead of the Organization for Economic and Cooperative 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. A third way would be to use authentic and performance-based assessments used by districts and schools.

Intrapersonal indicators report on social-emotional characteristics related to cognitive competencies and successful school and workplace performance. EQuIP relies on a self-report student measure of self-regulated learning and the percent 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 is easily calculated from administrative records and reports on a behavioral pattern that is detrimental to student learning and development (Balfanz & Byrnes, 2012).

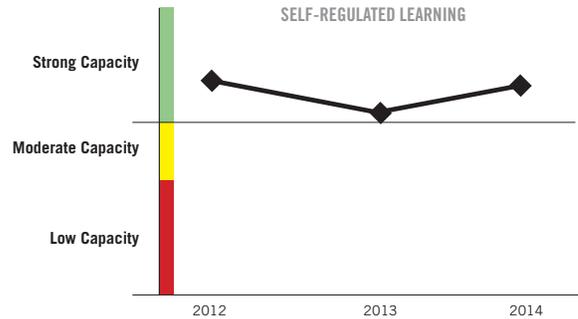
Interpersonal indicators reflect competencies of collaborative problem solving, communication, and teamwork. These skills can be measured through deeper learning processes, authentic district assessments, or a direct measure of students' interpersonal skills (Stecher & Hamilton, 2014). Wang, MacCann,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

DEEPER LEARNING FORMATIVE INDICATORS

SAMPLE HIGH SCHOOL NAME

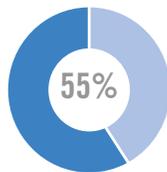


Formative indicators inform improvement decisions by capturing elements of deeper learning not measured with State curricular tests. Initially, formative indicators include deeper learning processes, intrapersonal dispositions, and interpersonal characteristics.

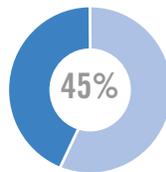


DEEPER LEARNING PROCESSES

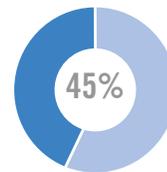
The degree to which students have opportunities to apply basic knowledge and skills to real-world situations. Indicators include students' self-report of instructional tasks and objective indicators on the access to deeper learning opportunities.



Of students believe they regularly work on real-world problems in school.



Of students believe they regularly work in teams to find solutions to problems.



Of students believe they have regular opportunities to present their work to peers and adults.

STUDENTS ENROLLED IN 1 AP/IB COURSE

60%

STUDENTS ENROLLED IN CONCURRENT ENROLLMENT

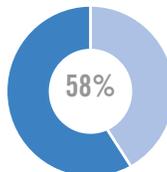
25%

STUDENTS ENROLLED IN CAREER/TECHNICAL TRAINING

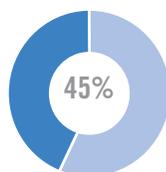
15%

INTRAPERSONAL

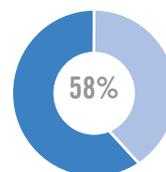
The degree to which the regulation of learning processes and outcomes derives from the internal control of students. Indicators report self-regulated learning and the percentage of students chronically absent.



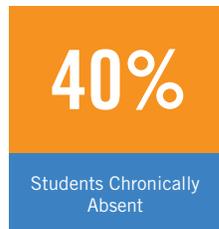
Of students believe they do their classwork because they want to learn new things.



Of students believe that they keep trying even after they fail.

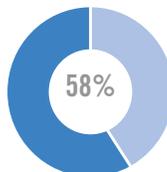


Of students believe that they get school work done even when they don't feel like doing it.

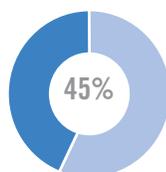


INTERPERSONAL

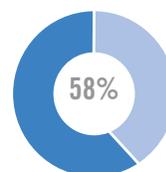
The degree to which students can work collaboratively and as an active contributor of a team.



Of teachers believe students listen and respect other students ideas.



Of teachers believe students work well with each other.



Of students believe students can work through their differences.

Figure 11. An example of a capacity page for EQUIP.

PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE

IMPROVEMENT INDICATORS: PROCESS AND RESOURCE CAPACITY

Following the lead of improvement science in healthcare and education, the objective for using process and resource 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 EQUiP 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 economic advantage and disadvantaged communities (Putnam, 2015).

IMPROVEMENT INDICATORS: INSTRUCTIONAL CAPACITY

Instructional capacity exists in resources and social processes that enable teachers both individually and collectively to reflect on and learn from their instructional practices. Useful indicators capture the degree to which a school is organized to help teachers continue 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 indicators report 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 indicators account for the quality of relationships among a teaching faculty and between teachers and school leaders. Trust is 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.

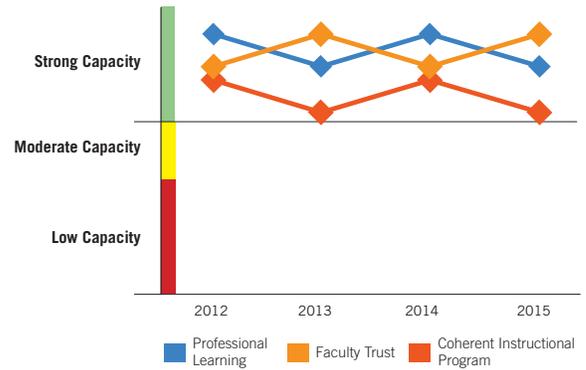
Coherent instructional program indicators reflect a school that is organized around a shared and coherent instructional approach, has a vertically and horizontally aligned curriculum, and takes action to make teaching and learning measurably better. We propose the instructional program coherence scale from The University of Chicago Consortium of Chicago School Research.

INSTRUCTIONAL CAPACITY

SCHOOL NAME GOES HERE

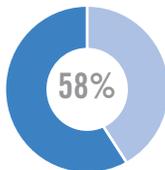


Instructional capacity refers to the aggregate ability and readiness of the school's teaching corps to design and deliver appropriate, challenging, and goal-related instruction to all students. Indicators for this capacity include measures of teacher professional learning, quality of professional relationships, and the coherence of the instructional program.

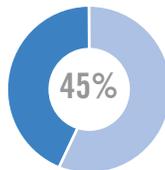


PROFESSIONAL LEARNING

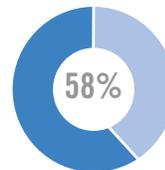
The degree to which teachers experience the school environment as supporting them as professionals and enriching their development as expert practitioners.



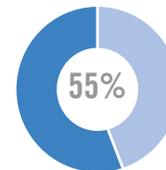
Of teachers believe learning opportunities provided useful knowledge.



Of teachers believe their learning opportunities made them pay closer attention to their instruction.



Of teachers believe their learning opportunities provided useful feedback on their teaching.



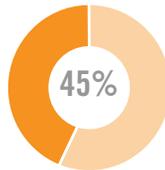
Of teachers believe their learning opportunities led them to try new things in the classroom.

FACULTY TRUST

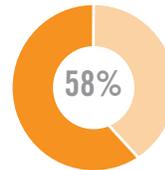
The quality of relationships among a teaching faculty and between teachers and school leaders.



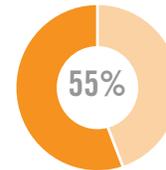
Of teachers believe that teachers in this school typically look out for each other.



Of teachers believe that teachers in this school trust the principal.



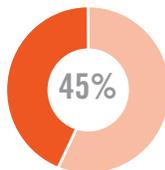
Of teachers believe that teachers in this school are open with each other.



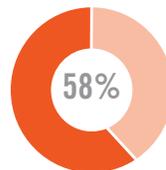
Of teachers believe that teachers in this school have faith in the integrity of the principal.

COHERENT INSTRUCTIONAL PROGRAM

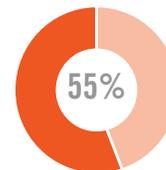
A school that is organized around a shared and coherent instructional approach has a vertically and horizontally aligned curriculum and takes continual action to make teaching and learning measurably better.



Of teachers believe that curriculum, instruction, and learning materials are well coordinated across the different grade levels at this school.



Of teachers believe that there is consistency in curriculum, instruction, and learning materials among teachers in the same grade level at this school.



Of teachers believe that once the school starts a new program, we follow up to make sure that it is working.

PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE

IMPROVEMENT INDICATORS: LEARNING CAPACITY

Learning capacity exists in an instructional environment defined by high levels of student trust, motivation, and engagement. These conditions activate student interest, curiosity, and determination to persist in academic pursuits (Adams, Forsyth, Dollarhide, Miskell, & Ware, 2015). High levels of learning capacity do not guarantee positive educational outcomes, but it is hard to imagine deeper learning flourishing without conditions of trust, motivation, and engagement.

Trust indicators account for the quality of relationships among and between students and teachers. Trusting relationships reflect a learning climate that brings out the best attitudes and behaviors in students, enabling them to persist toward the challenging expectations of deeper learning (Adams, 2014). Student trust should be measured with a de-identified survey, and results should be aggregated to the school level. We propose items from the student trust in teachers and student trust in students surveys for students in grades 5th through 12th grades (Forsyth, Adams, & Hoy, 2011).

Motivation indicators 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, Miskell, & Ware, 2015; Niemic & Ryan, 2009; Kusurkar, Cate, Vos, Westers, & Croiset, 2012). 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.

Engagement indicators reflect a school environment where students are engaged in deep learning, are excited about school, and have positive attitudes toward their future. Many engagement surveys and items exist and can be used. We propose the engagement items from the Quaglia National Student Voice Survey.

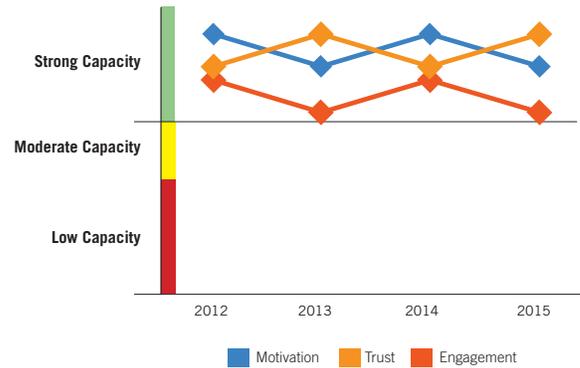


LEARNING CAPACITY

SAMPLE ELEMENTARY SCHOOL

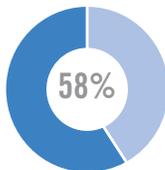


Learning capacity refers to the quality of the learning environment and its ability to activate student interest, curiosity, and determination to persist. Indicators of this capacity include aggregate measures of student motivation, student-teacher trust, and engagement.

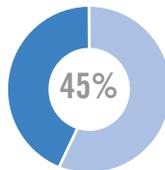


MOTIVATION

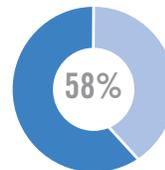
The degree to which students experience instructional practices as supporting their autonomous motivation for learning and achieving high academic standards.



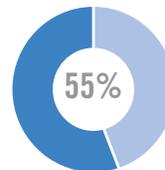
Of students believe teachers show students how to solve problems themselves.



Of students believe teachers talk about the connection between what is studied in school and what happens in real life.



Of students believe teachers in this school expect students to work hard.



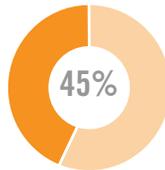
Of students believe teachers in this school challenge students to achieve academic goals.

TRUST

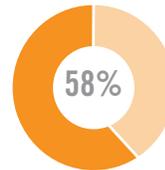
The quality of the relational connection between students and teachers as well as among students.



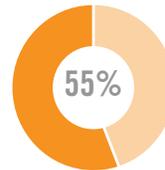
Of students believe teachers really listen to students.



Of students believe they can depend on teachers for help.



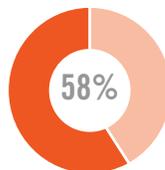
Of students believe students are honest.



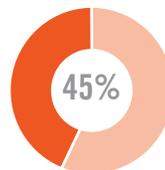
Of students believe students help each other.

ENGAGEMENT

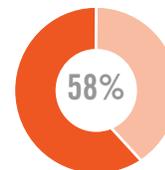
The extent to which students are emotionally involved in their schoolwork and feel that it activates their curiosity, creativity, and excitement.



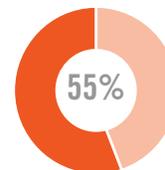
Of students believe teachers make school an exciting place to learn.



Of students feel proud of being part of the school.



Of students enjoy learning new things.



Of students feel engaged in their learning.

PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE

IMPROVEMENT INDICATORS: ORGANIZATIONAL CAPACITY

Organizational capacity calls attention to the essential function of human resources in 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.

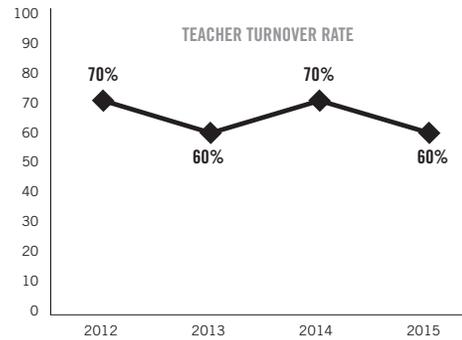
Human resource indicators account for the stability of a high quality teaching corps and stable principal leadership. Teacher core stability is measured by the percentage of teachers returning to the school (tracked over five years), the percentage of teachers who remain in a school for a five-year period, 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, years of assistant principals in the school, and number of principals in a ten-year period. Information is also reported on number of instructional coaches, student-teacher ratio, counselor-to-student ratio, SPED student to SPED teacher ratio, average class size, and largest and smallest class size.

ORGANIZATIONAL CAPACITY

SCHOOL NAME GOES HERE



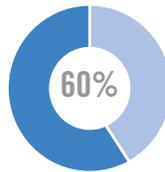
Organizational Capacity refers to school resources that are foundational to the development and maintenance of high quality teaching and learning. Indicators of this capacity include measures of teacher and leader corps stability, teacher substitution rate, class size, and student-teacher ratio.



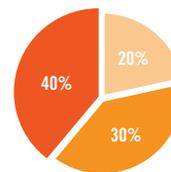
TEACHER CORPS STABILITY

The percentage of teachers who remain in a school over a five year period, the distribution of teachers by experience, and a Teacher Substitution Ratio (total teacher days covered by subs / (total school FTE x total days)).

% OF TEACHERS REMAINING IN THE SCHOOL

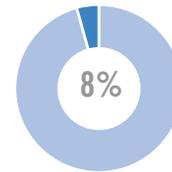


DISTRIBUTION OF TEACHING EXPERIENCE



0-5 years 5-10 years 10 and more years

SUBSTITUTION RATE



LEADERSHIP STABILITY

The years of principal experience at the current school, the number of principals at the school over a ten year period, and the average years of assistant principals' experience at the current school.

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

SUPPORT STAFF

Information on number of instructional coaches, counselor-to-student ratios, and ratio of SPED students and SPED teachers.

FULL TIME INSTRUCTIONAL COACHES	2
COUNSELOR TO STUDENT RATIO	1/150
SPED STUDENT TO SPED TEACHER RATIO	1/3

CLASS SIZES

Average class size, largest class size, smallest class size, and student-teacher ratio.

AVERAGE CLASS SIZE	22
LARGEST CLASS SIZE	44
SMALLEST CLASS SIZE	8
STUDENT-TEACHER RATIO	1/20

PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE

IMPROVEMENT INDICATORS: HOME AND COMMUNITY CAPACITY

Home and community capacity addresses the relational connections between schools and families as well as family resources available to children. Indicators report on the quality of parent-school interactions based on parent trust and parent perceived school outreach. Household resources for children are indicated by food insecurity and the number of parents in the home.

Parent Trust in School measures the quality of relationships between teachers and parents. Questions ask parents 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. Questions ask parents about how well the school communicates information, about parent opportunities to provide feedback to school authorities, and about 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 percent of children in the school from households with an inadequate food supply as a result of insufficient funds or resources. Questions from the CPS Food Security Survey ask parents 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.

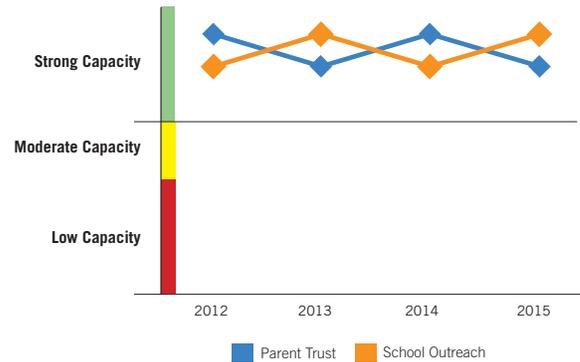
Number of Parents in the Home measures the percent of children in the school who do not live with both parents in the home. Growing up in a single-parent family is linked to a range of negative outcomes in school and later in life (Sigle-Rushton & McLanahan, 2004). Higher percentages of children living with both parents tend to suggest increased supervision for children and increased access to material, human, and social capital.

HOME AND COMMUNITY CAPACITY

SCHOOL NAME GOES HERE

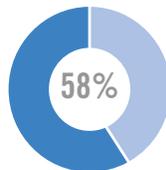


Home and community capacity refers to the relational connections between schools and families, as well as the family resources available to children. Indicators report on the quality of parent-school interactions based on parent trust and parent perceived school outreach. Household resources for children are indicated by food insecurity and the number of parents in the home.

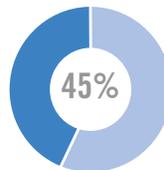


PARENT TRUST

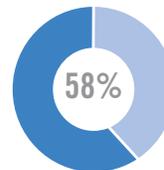
The degree to which parents perceive teachers and the school as being open, honest, reliable, competent, and benevolent. Questions ask parents about teacher academic standards for all students, teacher concern for students, teacher communication with parents, teacher competence, teacher honesty, and teacher reliability in their actions and commitments.



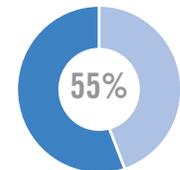
Of parents believe the school has high standards for all students.



Of parents believe the school is always ready to help.



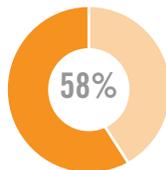
Of parents believe students are well cared for.



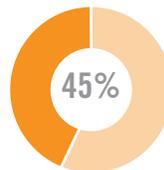
Of parents believe the school does a terrific job.

SCHOOL OUTREACH

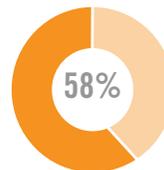
The degree to which parents perceive school-parent communication and interactions as open and supportive. Questions ask parents about how well the school communicates information, about parent opportunities to provide feedback to school authorities, and about parent feelings of belonging in the school community.



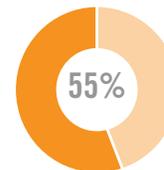
Of parents believe teachers communicate regularly with parents.



Of parents believe the school works closely with parents to meet student needs.



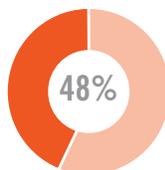
Of parents believe the school regularly communicates with parents about how they can help their children.



Of parents believe that parents are encouraged to give feedback to the school.

FOOD INSECURITY

The percent of children in the school from households with an inadequate food supply as a result of insufficient funds or resources. Questions from the CPS Food Security Survey ask parents 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.



NUMBER OF PARENTS IN THE HOME

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



PART THREE: AN EDUCATIONAL QUALITY AND IMPROVEMENT PROFILE

EQuIP ONLINE

EQuIP comes with electronic accessibility through an online platform that is compatible with personal computers and mobile devices. The online site guides the user in accessing profile pages, graphs, and descriptions of the concepts captured in the reports. When fully activated, the online feature enables the user to drill down to additional information on school outcomes, processes, and resources.

Figure 12 is a screen shot of a sample "landing page" and "profile page." Users looking for information about deeper learning, college and career readiness, or different capacities simply click on the components of EQuIP they desire to see. Information graphs then display descriptive data and information in a way that is clear and easy to understand.

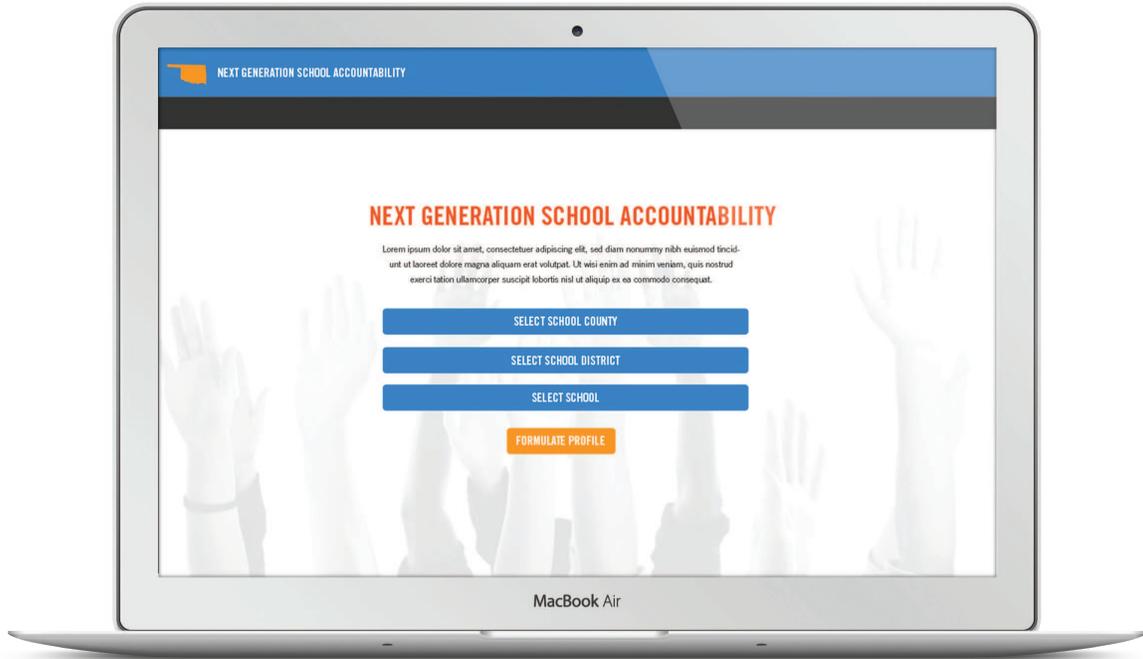


Figure 12. Electronic landing site for EQUIP.

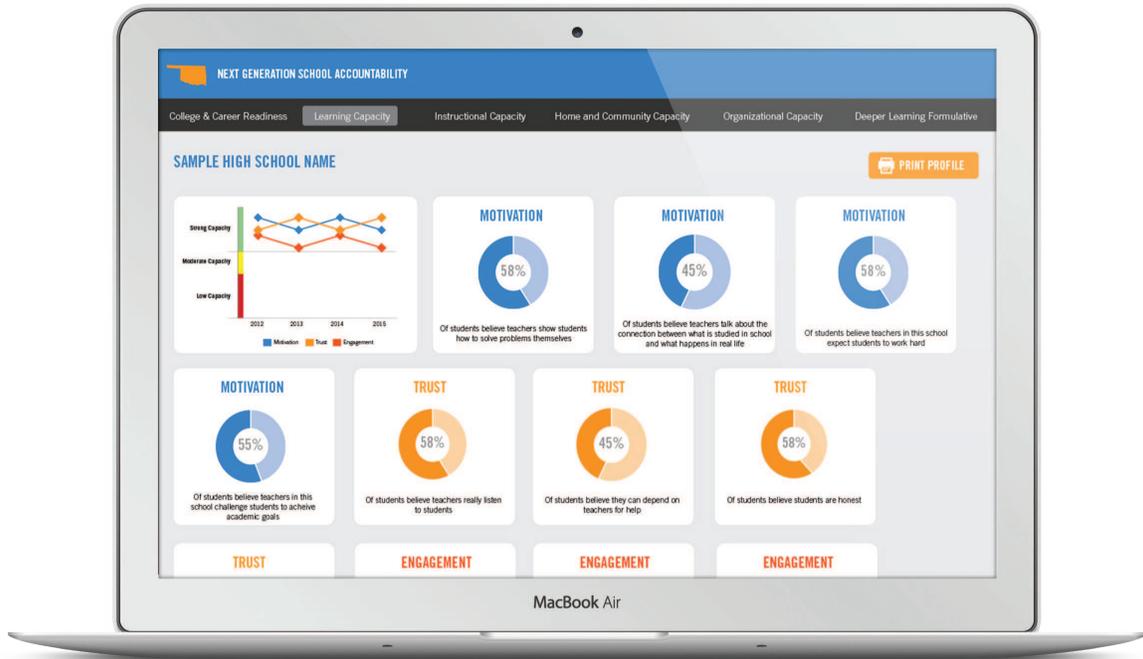


Figure 12. Electronic landing site for EQUIP.

PART FOUR: DESIGNING AND IMPLEMENTING NEXT GENERATION ACCOUNTABILITY

The previous three parts outlined deficiencies in Oklahoma’s A-F grading system and proposed a design for a next generation accountability that addresses these deficiencies and establishes alignment with the State’s explicit goal of achieving college and career readiness for all graduates. In this part, we translate next generation accountability into a set of recommendations for enacting this framework. These recommendations have three distinct targets: 1) accountability policy, 2) alignment of standards, assessments, and accountability, and 3) school, district, and state capacity-building in support of the accountability framework.

ACCOUNTABILITY POLICY

An accountability policy that grades schools based on aggregate test scores of low-level knowledge and skills is certain to leave Oklahoma students behind. We envision effective accountability policy as establishing high expectations for schools and students and supporting meaningful and worthwhile investments in resources and processes that can move the entire educational system toward a desired future state (Darling-Hammond et al., 2014). Oklahoma has set a vision of deeper learning and college and career readiness for all students; now it needs to invest in an accountability system that supports innovation, transformation, and ongoing improvement.

The design of next-generation accountability requires revisions to state statute prescribing the calculation of A-F grades. As demonstrated earlier, the current law prescribes a system for grading schools that fails to deliver useful information about school progress toward the goals of deeper learning and college and career readiness. Moreover, the system does not support the intelligent investment in strategies that can build school capacity by targeting sources of underperformance. Going forward, 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 the State Department of Education with the managerial flexibility to design a system that fulfills these expectations.

“...the political climate—including both bills to revise ESEA and (as far as I can tell) the presidential election—indicate that more authority over data systems will shift to the states. By moving this way now, Oklahoma could become a leader in the new round of state educational improvement policies.”

-William A. Firestone, Rutgers University



PART FOUR: DESIGNING AND IMPLEMENTING NEXT GENERATION ACCOUNTABILITY

We propose the following recommendations and rationale for accountability policy.

1. Eliminate the A-F grading system.

- Grades do not align with deeper learning and college and career readiness.
- Grades cannot be used to make valid and reliable judgments of school quality.
- Grades do not reflect the performance of many students in the school.
- Grades do not provide useful information for improvement.
- Grades do not identify schools in catastrophic failure.
- Grades do not meet the intent of federal expectations for achievement equity and improvement.
- Grades have not led to any meaningful improvements in student achievement.

2. Do not use a single, summative index to report accountability information. Outcome evidence should be reported in ways that clearly portray student progress toward deeper learning and college and career ready standards, changes in student performance over time, and achievement gaps.

- High 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 with the instability of time point estimates.
- Evidence on achievement equity and performance trends allow for more reliable identification of schools in need of State intervention.

3. 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.
 - The State can better identify schools in need of State intervention by understanding capacity differences among schools.
- ### 4. Accountability policy should adhere to the Standards for Educational and Psychological Testing, which note, “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).
- The 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 Oklahoma's 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 of the State of Oklahoma. Toward this end, the following recommendations in the areas of curricular standards, assessment and evaluation are advanced:

- 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):**
 - Assessment of higher order cognitive skills.
 - High fidelity assessment of critical skills.
 - Benchmarked to international standards.
 - Instructionally sensitive and educationally valuable.
 - 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.**
- 6. The accountability assessment indicators 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.**

PART FOUR: DESIGNING AND IMPLEMENTING NEXT GENERATION ACCOUNTABILITY

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; professional learning opportunities for all school professionals are important, but also necessary is the provision of time and space for sense-making about the intent of the new policy, how it differs from past policy, and the significance these differences hold for changes to practice (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 system-wide components of a support infrastructure for building capacity across the educational system:

A system-wide culture grounded in “learning to improve.”

Change under next-generation accountability is dynamic and context-specific. The new support system needs to communicate its importance for enhancing local and state-wide 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. Science of improvement methodology and the idea of 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 Networked Improvement Communities (NICs). The idea of NICs reflects a recognition of social networks as a valuable resource for change which facilitates the sharing of knowledge across improvement contexts. By connecting like schools or districts, NICs institutionalize channels through which information on learning can be shared.

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 the technical ability related to one or more of the following areas: numeracy and statistical knowledge/ability, 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 has far outstripped our attention to the need for better training on how to go about this process (Datnow & Park, 2014; Mandinach, 2012). The next generation accountability system maintains the current focus on data-driven decision making, but also asks school professionals to become expert on the use of data to both 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 and other improvement/data.

Prioritization of resources for sustaining on-going improvement.

System-wide availability of resources such as time, access to various forms of expertise, and collaborative opportunities are critical. 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 adults will be able to see a wide-range of possibilities in addressing issues of teaching, leading, and learning. Allocating time and improving access to expertise and collaborative opportunities may require some increased or reallocation of school funding. The intentional allocation of these key resources sends a signal about the importance of such endeavors and creates conditions by which critical learning can be achieved by schools and school personnel.

A coherent structure of State-level support for learning to improve.

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 aligned with next-generation accountability. This support structure could build on existing support resources such as the Office of School Support in the Oklahoma State Department of Education, but it would require significant expansion to accommodate learning needs across the State. The current resources provided to the Office of School Support are insufficient.

Educator labor market policy which supports the above elements.

Little progress in the above elements will be made without addressing key Oklahoma educator labor market challenges, and this includes a re-examination of current policy tied to the supply of experienced educators in Oklahoma. A recent study of Oklahoma educator supply and demand highlights some disturbing current and future trends with respect to teacher supply and demand in Oklahoma. Currently, teacher salaries fall well below those of adjacent states but also as compared to their non-educator counterparts within the same Oklahoma labor market.

Further, the gap between the influx of newly trained teachers and those who are leaving the profession is widening quickly (Berg-Jacobson & Levin, 2015). 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 one.

Furthermore, our proposed system of supports will require an influx of professional educators who have had considerable past experience working in schools. 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 for highly qualified former educators to participate in these needed support positions will be necessary. These policy changes are not without precedent: states such as Michigan have responded to labor market shortages by relaxing these regulations.

PART FOUR CONCLUSION

Just as the over-reliance 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 that comprise them can make meaningful improvements with little-to-no support. Changing existing patterns of behavior is difficult work, even if one is highly motivated to do so. As Spillane (2004), professor at Northwestern University, once noted in his study of Michigan school districts undertaking standards-based reform, “...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). Attention to both the design of a better framework for improvement as well as attention to the support it needs to thrive at the State and local level are critical for the ultimate success of this next-generation accountability framework.

TECHNICAL APPENDIX

APPENDIX A

TECHNICAL APPENDIX—HIERARCHICAL LINEAR MODELS (HLM) OF OKLAHOMA STUDENT ACHIEVEMENT GROWTH FROM 2011-2012 TO 2013-2014 IN MATH AND READING

In Part One, we report the results of two HLM growth models of Oklahoma student achievement. These models estimated the relationship of reading and math achievement growth over the academic years of 2011-2012 to 2013-2014 to important characteristics of students in grades 3-8. This brief section summarizes the technical nature of the analyses which were reported in aggregate in Part One.

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 Lunch status of student, ELL status, race/ethnicity, gender, and A-F grade of the school in which the student resided. Only 1 percent of students in the final sample were missing the A-F letter grade recorded for the 2011-2012 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 three 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):

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

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_{it} 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 2011-2012 school year.

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

$Time_{it}$ was measured in academic year and centered on Year 1 (the 2011-2012 school year).

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

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

β_{00} represents the average reading or math score for a white, non-ELL, non-Free-or-Reduced Lunch student in an A school during the 2011-2012 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 2011-2012 school year. As mentioned before, these were: Free-and-reduced lunch status of student, ELL status, race/ethnicity, gender, and A-F 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 Lunch students in an A schools.

$\beta_{1q}X_q$ signals that q number of student-level covariates were entered into the achievement models predicting linear growth rate in achievement during the three years under study. These were the same variables entered to predict student achievement initial status: Free-and-Reduced Lunch status of student, ELL status, race/ethnicity, gender, and A-F 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

- ACT. (2015). The Condition of college and career readiness 2015. ACT. Inc. Available at: <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., 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., A. Duckworth, J. J. Heckman, and T. Kautz (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: 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://annenberginstitute.org/pdf/Results.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). *The importance of being in school: A report on absenteeism in the nation's public schools*. Baltimore: 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.
- *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: 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: University of Chicago Press.
- *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.

REFERENCES

- *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.
- Contey, 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: ScarecrowEducation.
- 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., 567-579.
- Forsyth, P. B., Adams, C. M., & Hoy, W. K. (2011). *Collective trust: Why schools can't improve without it*. New York: Teachers College Press.
- Forsyth, P.B. & Talerico, M. (1998). Accountability and city school leadership. *Education and Urban Society*, 30(4), 546-555.
- Fullan, M. (2010). *All systems go: The change imperative for whole system reform*. Newbury Park, CA: Corwin Press.
- *Glaser, R. (1984). The role of knowledge. *American Psychologist*, 39(2), 93-104.
- *Grubb, W. N. (2006). When money might matter: Using NELS88 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*. Russell Sage Foundation.
- Hall, D. (2013). *A step forward or a step back? State accountability in the waiver era*. Author: The Education Trust. Retrieved at <http://files.eric.ed.gov/fulltext/ED543222.pdf>
- Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (No. w11154). 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: 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. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- *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.
- 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.

REFERENCES

- *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,
- Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success* (No. w20749). 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*. 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: University of Chicago Press.
- Louis, K.S., Leithwood, K., Wahistrom, K., Anderson, S.A., & Michlin, M. (2010). Learning from leadership: Investigating the links to improved student learning: Final report of research findings. New York: 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
- *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*. University of Chicago Press.

REFERENCES

- 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: McKinsey.
- National Center for Education Statistics. (2013). *NAEP 2012: Trends in academic progress, reading 1971-2012, Math 1973-2012*. US 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. Committee on Defining Deeper learning and 21st Century Skills. J. W. Pellegrina and M. L. Hilton, (Eds). Washington, DC: National Academics Press.
- OECD (2014). Program for international student assessment: Results from 2012. Available on-line at <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/eseaflex/approved-requests/okapp11212014.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*.
- 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>
- *Reeve, J. (2006). Extrinsic rewards and inner motivation. *Handbook of classroom management: Research, practice, and contemporary issues*, (pp. 645-664.) Location: Publisher
- 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: 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: 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. *Theory 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 (Eds.), (2008). *Motivation and self-regulated learning: theory, research, and applications*. New York: Lawrence Erlbaum Associates.

REFERENCES

- *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). Oklahoma State University, Stillwater, OK.
- Shepard, L., Hannaway, J., & Baker, E. (2009). *Standards, assessments, and accountability*. Education Policy White Paper. 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 (Feds.), *The future of the family* (116-155). New York: 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.
- *Wertheimer, M. (1959). *Productive thinking*. New York: 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.
- *NOTE:** References marked with an asterisk indicate resources consulted but not cited.

