Robust and Equitable Measures to Identify Quality Schools

Concept Brief

Theresa Anderson, David Blount, Constance Lindsay, Erica Blom, Semhar Gebrekristos, and Francisca Alba

April 2019

Robust and Equitable Measures to Identify Quality Schools (REMIQS), a new initiative by the Urban Institute and Jobs for the Future,[1] seeks to build a comprehensive definition of high-quality nonselective public high schools serving student populations often underserved by the public education system. REMIQS investigates the availability of meaningful, equity-centered measures to identify high-quality schools that could be comparable across states. The measures include traditional test score metrics and new indicators identified through the literature and conversations with experts in the field.

After identifying high-quality public schools based on robust and equitable results indicators, the study will seek to describe the policies, programs, practices, or other characteristics that characterize high-quality schools, as identified through more robust measures. Thus, the project intends to take a reverse-engineering approach that first captures *which* schools are producing strong outcomes and then investigates *how* those schools achieve at that level, particularly for students historically underserved by the US education system. This requires not only robust, valid, and reliable outcome measures but also good information about schools' internal processes.

This brief describes the motivations behind this study and the concepts that may characterize high-quality schools, as discerned from the literature and discussions with leaders in the field. The framework presented here could be a resource to educators, policymakers, and others interested in the components of school functioning that may influence student outcomes in meaningful ways.

---

[1] The project was conceptualized in collaboration with Jobs for the Future. In early 2019, Jobs for the Future's project leadership transitioned to KnowledgeWorks.

Why REMIQS?

The acronym REMIQS—like its homonym, *remix*, which describes musical enhancements—references the desire for this work to take a well-known classic and breathe life into the composition through new influences. In this case, the "well-known classic" is measuring achievement via test scores and investigating the school processes that associate with this narrow measure. Education policy research has played this tune on repeat for almost two decades. New influences include measures that reflect the broader purposes of public education beyond achievement *qua* achievement. New measures are emerging in the field but represent an ever-evolving genre. The REMIQS effort will mix new and original measures (i.e., test-score measures) into a harmonic symphony.

Schools and Social Justice

Philosophically, schools aspire to enact results where young people gain full access and understanding of past knowledge while innovating for present and future interests. **Equity is the central goal in developing a system where every child can develop the capacity to live a productive life.** But given American history and social and political realities, **not all students have equitable opportunities to see their aspirations come to fruition**. The country now, per Ladson-Billings (2006), has an "educational debt," meaning that society has invested in children discriminatorily, along lines of race and class, over many generations. This "debt" manifests in communities deprived of the opportunity to "share in the inherited resources of the [human] race, and use [their] own powers for social ends" (Dewey 1897), a fundamental definition of the purpose of school as a social institution. Simply, poor communities, especially poor communities of color, have been excluded from building social, political, and economic capital in their own communities through strong, foundational education. This exclusion and deprivation has compounded over generations. Without appropriate interventions, racial and class disparities will persist along a spectrum of life outcomes (Ladson-Billings 2006; Gutiérrez and Dixon-Román 2010; Milner 2013; Trujillo and Renée 2012).

Acknowledging the impact of poverty and race in education, it is necessary to examine school quality in relation to the needs developed by the history of race and class relations and exclusions. These needs require a move away from a deficit-focused discussion of achievement "gaps" and instead for conceptions to orient toward the vision and aspirations of communities that have long been marginalized in educational pursuits (Gutiérrez and Dixon-Román 2010; Ladson-Billings 1999). **This project's motivation is to conceptualize schools as venues for social justice capable of reducing social inequality through equitable treatment of students who may need compensatory resources and strategies to promote life success.** Schools may prepare students for further academic pursuits, work, citizenship, and fulfilling lives. These purposes are not mutually exclusive, and assessing school quality requires balance across various measures.

Underserved Students and the Opportunity Gap

Many stakeholders in the American education system suggest that compensatory resources and life-success strategies should be emphasized for schools that serve students who have been historically underserved by American public education. The federal government defines these "high needs" students as students "at risk of educational failure or otherwise in need of special assistance and support, such as students who are living

in poverty, who attend high-minority schools (as defined in the Race to the Top application[1]), who are far below grade level, who have left school before receiving a regular high school diploma, who are at risk of not graduating with a diploma on time, who are homeless, who are in foster care, who have been incarcerated, who have disabilities, or who are English learners."[2] How these dynamics play out at a school level can be complicated. School districts serving a substantial share of low-income students based on the federal poverty level are eligible for Title I grant resources. This does not necessarily mean that all schools serve low-income students. In the 2014–15 school year, 12,870 of 20,280 regular public US high schools (63 percent) were eligible for Title I. These schools served 61 percent of the 15.4 million 9th to 12th graders in regular public high schools in that same year (9.4 million students).[3]

To focus on schools serving student populations often marginalized by public education, **the REMIQS effort will focus on schools serving a substantial portion of students who are low income, who are non-White, or who need support learning English**. These groups have been disadvantaged by unequal distribution of economic and social resources based on race, class, language, and national origin (Ladson-Billings 2006). This disadvantage has created a persistent "achievement gap," where schools serving high-needs students have struggled to enable students to attain the advanced skills required by state standards (Haskins et al. 2012).

To move beyond the rhetoric of the achievement gap and to recognize the deep, persistent, and growing education debt that has accumulated over centuries of underserving students from nondominant racial and ethnic backgrounds (Ladson-Billings 2006), **it is more useful to frame the issue around the opportunity gap** (Carter and Welner 2013). The opportunity gap reflects the differential opportunities in schooling and society between students in the dominant cultural group and students who are minoritized because of historic and present underinvestment and structural barriers. In addition, because schools can be a signal for individuals' future opportunities (e.g., employment), the deprivation of high-quality schooling can compound not only learning deficits but important opportunity deficits later in life (Rose and Betts 2004).

Introducing a focus on race and class consciousness in education does not mean that educators emphasize student and family deficits. Instead, it means educators recognize and are held accountable for practices that build on marginalized students' unique assets, knowledge, and positioning in social hierarchies that otherwise would be treated as deficits. It also means educators possess or are helped to develop beliefs and pedagogical skills that address status differences, inequitable allocations of resources, and procedural and structural changes so they can remove race and class identifiers as predictors of achievement.

Schooling contexts for marginalized students often present social, economic, and cultural barriers. These schools may face important safety and security concerns because of gang violence, domestic violence, and law enforcement or immigration enforcement violence. Students in these schools may be working to support their families—either their parents and siblings or their own children. Students may be homeless, either with their families or because they have been disowned by their families because of their sexual orientation or gender identity. And students may face enormous language, literacy, and skills barriers—both native English speakers and speakers of other languages. But some schools may be doing important restorative work with students and supporting needs beyond the bounds of traditional education that reductive test score measures do not capture. This project hopes to do justice to these efforts.

The Original Soundtrack: Test Scores

Over the past two decades, test scores have been the primary tool to define school success and quality. Test scores have evolved as a dominant measure for accountability because they are easy to use, store, and compare across students. They continue to be the main achievement measure in much of the research literature, though some recent studies have included indicators beyond test scores. An increased emphasis on test score measures, especially in high-poverty districts, emerged from the 2001 No Child Left Behind Act (NCLB) (Harris 2006).[4] NCLB required Title I schools to meet certain test-measured achievement levels, based on proficiency measures intended to determine if schools were bringing students to a minimally acceptable level of performance, set by each state. The federal government required that a certain share of students in Title I schools score as "proficient" on a statewide test (also known as the Percentage of Proficient Student indicator), or the school would be subject to escalating sanctions that could result in school restructuring.

The reliance on Percentage of Proficient Student indicators led to concerns about proficiency measures. One important critique is that test score measures associate strongly with student demographics, including race, parental education, and family income (Davis-Kean 2005; Hanushek 2014; Reardon 2011). Thus, students who are non-White or whose parents have lower educational attainment or income will tend to have lower test scores, on average, than students whose parents have higher educational attainment or income. Because families with similar demographics, income levels, and years of education tend to congregate in the same neighborhoods (Owens, Reardon, and Jencks 2016), school test scores often more heavily reflect the demographics of that neighborhood than student learning. Thus, a school in a low-income or non-White neighborhood may find it more difficult to reach a proficiency measure than a school in a middle- or upper-income neighborhood.

Researchers have examined how excessive emphasis on meeting test score proficiency levels can distort teacher incentives and lead to gaming behavior. Instead of improving overall learning, teachers may concentrate on "teaching the test," which reduces curricula to test-taking strategies (Koretz 2017).[5] Additionally, reliance on test score proficiency has also led teachers to disproportionately focus on "bubble kids"—students who are just below the proficiency cutoff—to increase their class Percentage of Proficient Student indicators (Ho 2008; Koretz 2017). Schools may also allocate time toward tested subjects and away from those that are not heavily tested (Koretz 2017). The risks of overreliance on a single measure was posited in Campbell's Law, which states, "The more any quantitative social indicator is used for social decisionmaking [such as test scores], the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell 2010). It seems this has come to pass, with the reporting of several school scandals resulting from test-score-centric incentives. For example, forty-four schools were implicated in a cheating scandal on a statewide standardized test in Georgia from 2001 to 2009.[6]

The drawbacks of proficiency measures are well known and have been much debated. Many have found that test score growth measures are better than proficiency benchmarks in accounting for student characteristics and the effects of schools on promoting student learning because they consider where students started out (Chingos and West 2015; Ho, Lewis, and Farris 2009). Growth measures increasingly inform teacher performance systems and other indicators of school quality (Council of Chief State School

Officers 2010). There are multiple ways to construct growth measures, and each comes with cautions and critiques (Bonk et al. 2012; Castellano and Ho 2013; Ho 2012; Lachlan-Haché and Castro 2015), especially if they are the primary basis for determining teacher effectiveness (US Department of Education 2013). A fundamental challenge of growth measures is that they are relative, so there is no guarantee that those who started out the furthest behind have achieved a minimum standard of competence desirable in society. Even if they have experienced substantial growth, they may still be substantially behind their more-advantaged peers or may not have the skills to pursue their career ambitions.

The research community has raised another problem with state standardized tests: inconsistency. NCLB required each state to create its own assessment, and therefore student achievement does not mean the same thing across states. A common approach to address this is statistical standardization of scores, but the scores still may rely on fundamentally different instruments. Recognizing this issue, other researchers have mapped state assessments to a common scale based on the National Assessment of Educational Progress (NAEP) for a set of years (Reardon, Kalogrides, and Ho 2017). The 2009 Common Core State Standards Initiative may provide increased consistency in assessments for the 42 states that participate in the effort.

Despite some cross-state inconsistency, test scores can be reliable (i.e., consistent) measures of student knowledge on *discrete concepts*. The problem is that they are not necessarily valid (i.e., accurate) in that they do not necessarily measure what students know and understand (Conley 2014). Relatedly, each concept is considered in isolation without a holistic picture of a student's knowledge (Conley 2014). New tests spurred by Common Core, such as the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC), include more complex questions, such as short answers with prompts for students to explain their responses to multiple-choice questions. But even open-ended questions on these improved tests are often presented without context and do not offer the opportunity for students to demonstrate their familiarity with holistic concepts that characterize deeper understanding.[7] Many have argued that these exams do not measure the range of skills and knowledge that are necessary for students to succeed in the 21st-century world (Conley 2014; Heller and Wolfe 2015; Mehta and Fine 2015). These have been called "deeper learning outcomes," "soft skills," "21st-century skills," and various other names. By one definition, deeper learning reflects the intersection of mastery, identity, and creativity (Mehta and Fine 2015). Others identify the key elements as the mastery of core academic content, critical thinking and problem solving, collaboration, effective communication, learning how to learn, and academic mind-set (Heller and Wolfe 2015). These important learning outcomes have gained renewed traction in recent education policy discussions (Conley 2014; Mehta and Fine 2015).

Lastly and importantly, critics of standardized tests argue that the content and format of tests reflect the dominant White culture and ignore the concerns and values of students of color: "Those determining what is to be mastered by students and how this mastery is to be reinforced through standardized tests represent the values and assumptions common to a once-dominant culture" (Gergen and Dixon-Román 2014). Thus, standardized tests ignore the pluralist makeup of modern society. Critical race theorists argue that the relationship between assessment outcomes and race protects White people as the dominant social group (Ladson-Billings 1999), exaggerating the achievement gap to perpetuate racial power dynamics. Gutiérrez and Dixon-Román (2010) note that, "Achievement gap studies often fail to question the validity of measurement tools or the choice to focus on measurement." Responsible use of score information from new

tests and consideration of school quality measures beyond traditional test-score-determined achievement are important steps in the right direction.
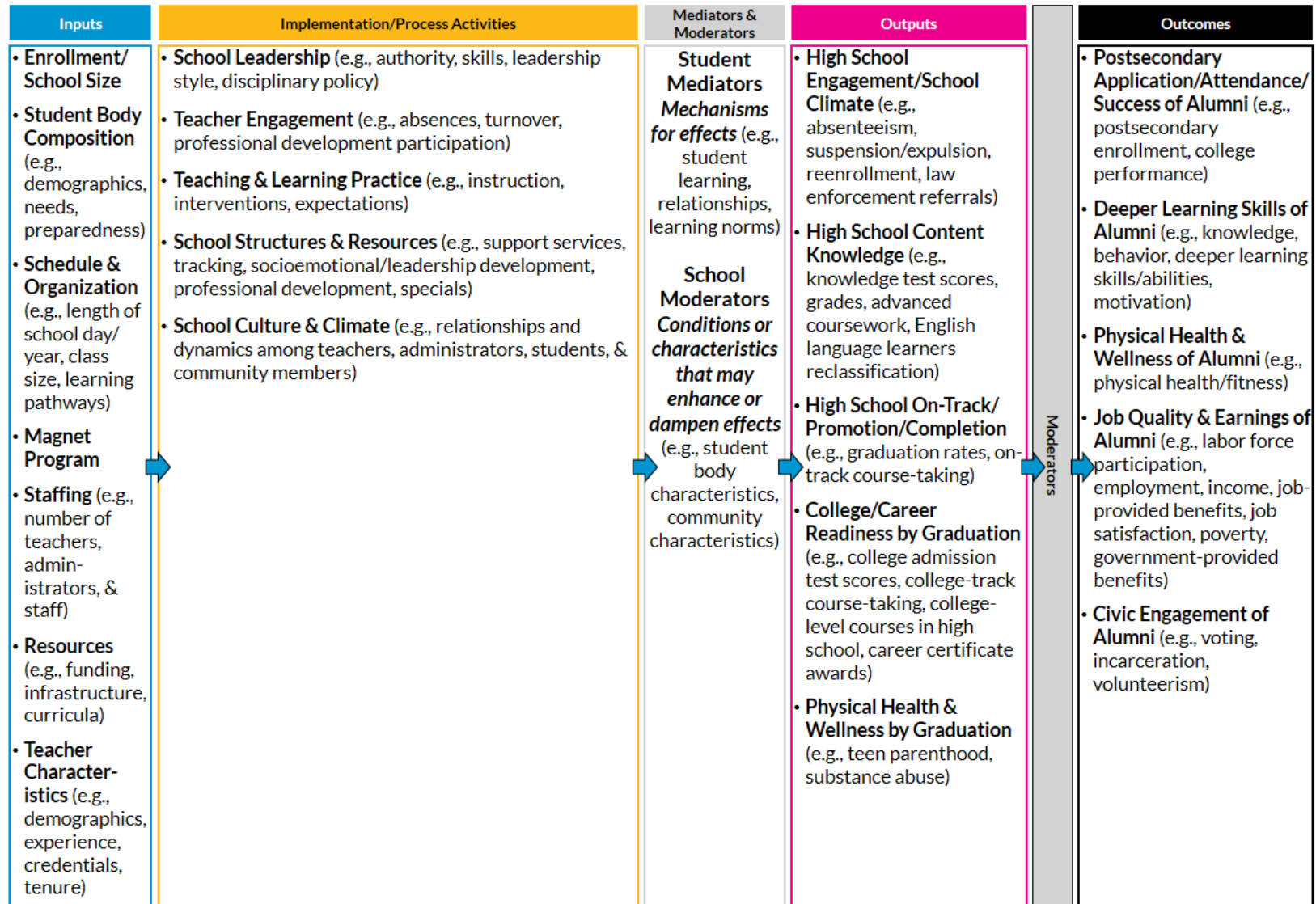
The REMIQS: Robust and Equitable Measures

Given these considerations, it is time to expand the range of measures used to judge school quality. The research team recognizes that "quality" can be situated within the social and historical context in which the discussion occurs (Ladson-Billings 1999). We have tried to be as inclusive as possible in creating robust definitions. If the nature of available data expands in the future, it may be possible to take more nuanced approaches to school identification that fully appreciate the complex construction of education and quality among diverse communities.

The REMIQS feasibility study has identified measures associated with school functioning and quality based on a review of popular frameworks and discussions with leading thinkers. **Figure 1 summarizes the inputs, implementation and process activities in the school, mediators and moderators, outputs, outcomes, and contextual factors that emerged from this review.**

These are constructed at the *school level*, treating the school as the unit of analysis, hence the focus on percentages, rates, and averages in the outputs and outcomes columns. We focused on the school level in appreciation of data source limitations.

FIGURE 1

Quality Diagram for Schools (Overview)

Inputs	Implementation/Process Activities	Mediators & Moderators	Outputs		Outcomes
• **Enrollment/School Size** • **Student Body Composition** (e.g., demographics, needs, preparedness) • **Schedule & Organization** (e.g., length of school day/year, class size, learning pathways) • **Magnet Program** • **Staffing** (e.g., number of teachers, administrators, & staff) • **Resources** (e.g., funding, infrastructure, curricula) • **Teacher Characteristics** (e.g., demographics, experience, credentials, tenure)	• **School Leadership** (e.g., authority, skills, leadership style, disciplinary policy) • **Teacher Engagement** (e.g., absences, turnover, professional development participation) • **Teaching & Learning Practice** (e.g., instruction, interventions, expectations) • **School Structures & Resources** (e.g., support services, tracking, socioemotional/leadership development, professional development, specials) • **School Culture & Climate** (e.g., relationships and dynamics among teachers, administrators, students, & community members)	**Student Mediators** *Mechanisms for effects* (e.g., student learning, relationships, learning norms) **School Moderators** *Conditions or characteristics that may enhance or dampen effects* (e.g., student body characteristics, community characteristics)	• **High School Engagement/School Climate** (e.g., absenteeism, suspension/expulsion, reenrollment, law enforcement referrals) • **High School Content Knowledge** (e.g., knowledge test scores, grades, advanced coursework, English language learners reclassification) • **High School On-Track/Promotion/Completion** (e.g., graduation rates, on-track course-taking) • **College/Career Readiness by Graduation** (e.g., college admission test scores, college-track course-taking, college-level courses in high school, career certificate awards) • **Physical Health & Wellness by Graduation** (e.g., teen parenthood, substance abuse)	Moderators	• **Postsecondary Application/Attendance/Success of Alumni** (e.g., postsecondary enrollment, college performance) • **Deeper Learning Skills of Alumni** (e.g., knowledge, behavior, deeper learning skills/abilities, motivation) • **Physical Health & Wellness of Alumni** (e.g., physical health/fitness) • **Job Quality & Earnings of Alumni** (e.g., labor force participation, employment, income, job-provided benefits, job satisfaction, poverty, government-provided benefits) • **Civic Engagement of Alumni** (e.g., voting, incarceration, volunteerism)

Contextual factors (e.g., neighborhood conditions, funding formulas, teacher pools, pre-K–8th grade education pipeline, state/district policies)

Indicators of Quality

We identified traditional measures through these efforts, as well as new indicators. The measures are organized into four broad categories: high school, further education, employment, and civic engagement and wellness. As noted, all measures treat the school as the unit of analysis.

High School

Outputs and outcomes that focus on indicators that occur while students are still in high school are often easiest to measure and track. These indicators may reflect the quality of the high school environment as it is contemporarily experienced by students and families, but they may not relate to long-term, more meaningful outcomes. These measures are often used as leading indicators for performance management (Foley et al. 2008).

Key high school results measures include the following categories and all related submeasures:

- High School Engagement/School Climate (outputs)
- High School Content Knowledge (outputs)
- High School On-Track/Promotion/Completion (outputs)

Further Education

Secondary education is increasingly important. Per the Bureau of Labor Statistics, workers with only a high school diploma earned 44 percent less than workers with some college or four-year degrees in 2017. Also, about three-quarters of the fastest-growing jobs in the next decade will require a high school credential and some postsecondary education or advanced training (Anderson et al. 2017).

Moving beyond high school toward postsecondary education—or a student's preparation for it—provides a more robust perspective on the quality of secondary schooling. We recognize that college attendance is susceptible to influences beyond the control of schools themselves, such as a student's or a family's financial resources, a family's orientation toward postsecondary education, the demands of the local labor market, and the availability of high-quality postsecondary options in the area. Nonetheless, students should be more able to complete their college careers if they receive good high school preparation.

Key further education results measures include the following categories and all related submeasures:

- College/Career Readiness by Graduation (outputs)
- Postsecondary Application/Attendance/Success of Alumni (outcomes)

Employment

One purpose of education is to prepare students to compete in the labor force. Measures of job quality and earnings and income are critical to measuring schools' success in meeting this goal. Measures of long-term employment, earnings, and family economic stability are particularly popular among economists (Betts 2001; Betts and Grogger 2003; Card 1995; Jackson, Johnson, and Persico 2016; Long 2016; Rose and Betts 2004), though these take longer to measure.

Key employment results measures include the following category and all related submeasures:

- Job Quality and Earnings of Alumni (outcomes)

Deeper Learning, Civic Engagement, and Wellness

A child is valued not only for her future economic potential but also her social and political contributions, which have broader societal implications and enrich the student's life trajectories. The literature increasingly focuses on the importance of holistic results beyond education and employment outcomes—particularly socioemotional and deeper learning skills, civic engagement, and physical health. These appear in state and district data collection efforts, such as the CORE School Quality Improvement Index, Boston Public Schools School Quality Framework, and the Oakland School Performance Framework. The Every Student Succeeds Act calls for all states to identify at least one measure that extends beyond traditional achievement, such as "student engagement; educator engagement; student access to and completion of advanced coursework; postsecondary readiness; school climate and safety; and any other indicator the State chooses that meets the requirements of this clause."

Various scholars have found that noncognitive skills endowed by schools are critical to long-term outcomes, including higher education and workforce success (Anderson et al. 2016; Bowles, Gintis, and Osborne 2001; Farrington et al. 2012; Heckman, Stixrud, and Urzua 2006; Levin 2012). In this brief, we refer to these skills as "deeper learning skills" (also called "socioemotional skills," "21st-century skills," "soft skills," or "employability skills"). The literature contains extensive and in-depth measures of these skills. Some scholars have expressed concern about using the available measures, such as the grit scale, for high-stakes assessment.[8] But that would not be the intention of this work, which is for research purposes. In addition, the research team would need to select or create a measure with fewer methodological drawbacks. This may be informed by a forthcoming repository of social and emotional learning assessments from the RAND Corporation (Schweig et al. 2017).

Key deeper learning, civic engagement, and wellness results measures include the following categories and all related submeasures:

- Deeper Learning Skills of Alumni (outcomes)

- Physical Health and Wellness of Alumni (outputs and outcomes)

- Civic Engagement of Alumni (outcomes)

School Phenomena

The second planned stage of the REMIQS study will consist of a mixed-methods investigation of school phenomena, the characteristics of and practices in schools that relate to quality. The literature offers varied perspectives on school-based processes, suggesting measures and data to inquire which school processes most effectively support the holistic growth and learning of the most vulnerable students (Preston et al. 2016; Rutledge et al. 2015; Schneider et al. 2017; Trujillo and Renée 2012). School phenomena may relate to *elements* or *components*, such as learner-centered leadership and quality instruction, that correlate with student success, broadly defined (Rutledge et al. 2015; Tichnor-Wagner, Harrison, and Cohen-Vogel 2016). Phenomena may also relate to specific *actors*, such as teachers and principals, examining the impact of strategies within their sphere of influence on student success (Preston et al. 2016; Stiggins 2005; Tichnor-Wagner, Harrison, and Cohen-Vogel 2016; West et al. 2016). Finally, phenomena may relate to *activities*, such as trauma-informed pedagogies and math-focused interventions to understand learning within a certain discipline or related to certain social problems (Lewis et al. 2014; Milner 2013).

An investigation of in-school processes should seek to understand *how do* and *how should* schools orient themselves to the specific needs of the most vulnerable and historically disadvantaged students. Research on race and class inequity in education, combined with macroeconomic shifts in societal conditions and skill demands, have meant that schools have needed to address a broader spectrum of needs in education beyond traditional academic fundamentals (Kyllonen 2012; Milner 2013; Preston et al. 2016; West et al. 2016). For example, culturally responsive teaching, trauma-informed counseling, socioemotional learning programming, and alternative historical narrative courses may address specific issues resulting from social inequity that may emerge only with certain student demographics.

With new approaches comes the need for new measures and new methodologies that can identify and describe effective activities with valid, reliable, and meaningful measurement. This project has identified phenomena along the spectrum of needs and growth areas that may strongly align with higher student academic and social-emotional growth. The identified phenomena fall into the following three categories:

- **School management and organizational indicators** inform how schools are managed. For example, principals may have strong collaborative decisionmaking processes with teachers and students. Such processes indicate the presence of participatory leadership.

- **Academic indicators** inform how schools enact foundational cognitive skill development for academic tasks, such as reading comprehension and math computation. For example, teachers may undertake differentiated instruction techniques per past student performance.

- **School culture and climate indicators** inform how schools reinforce beliefs and perceptions of learning, education, and schooling, which motivate behaviors and support meaning and purpose of structures within schools. For example, teachers may do student affirmations for specific tasks that show hard work.

Ideally, a study of school phenomena would investigate indicators in all three categories. In addition, understanding the interconnectivity between the three sets of indicators provides a more holistic view of internal processes influencing student achievement and sheds light on how to holistically meet student needs in the face of external stressors.

Systemic and Environmental Indicators

The context *around* schools influences their internal functions and their ability to meet student needs and produce desirable results. Greater understanding of schools' external factors helps researchers better account for causes of underperformance linked to sociohistorical pathologies that limit student progress.

Often, policies and other social and political conditions influence schools' capacity to serve the most vulnerable students (Ladson-Billings 2006; Schneider et al. 2017; Trujillo and Renée 2012). Most schools consistently work to implement best practices and provide the best possible education for their school community within their resource and policy constraints. But district and state educational leadership often hold responsibility for distribution of resources and policymaking based on educational priorities for the community. Decisions on curricula, teacher and staff development support, and funding at these higher levels can structure the landscape in a way to better serve all students. Unfortunately, schools often are not resourced to be most effective in serving the most vulnerable students. Commonly noted issues are disruptive top-down accountability mandates, limited teacher workforce, limited support for teachers and staff, and lack of funding and access to resources based on need (Betts, Zhau, and Rice 2003; Ladson-Billings 2006; Lewis et al. 2014; Milner 2013; Noguera, Darling-Hammond, and Friedlander 2015; Preston et al. 2016; Tichnor-Wagner, Harrison, and Cohen-Vogel 2016).[9]

In addition, this effort should evaluate what happens in schools in relation to what corresponding social safety nets provide in meeting needs outside school. Poverty and associated deprivation continue to be strong determinants of achievement. Schools may display systemic reactions to students who introduce specific chronic needs, such as hunger or homelessness. For example, schools in a high-poverty area or area with weak supports for affordable housing must confront high student mobility because of housing instability. Such conditions have important implications for student growth and learning trajectories, and they create systemic pressure for school districts to adapt policies and programming in response. Schools may offer supports to offset these conditions. These efforts, though likely critical for student success, may not be tracked in existing data or appropriately valued in the dominant lens of what constitutes "quality education."

Key measures of systems and environment include the following categories and all related submeasures:

- School characteristics
- Student body composition

- Resources
- Teacher characteristics
- Contextual factors

Study Description and Conclusion

For this study, the research team will gather existing data from up to seven state longitudinal data systems and national school-level data from public sources. The team will use this information to model the quality measures discussed in this brief to the extent feasible within the constraints of existing measurement. The research team will identify schools that seem to be promoting stellar outcomes among their students along several measures. Follow-up research may include in-depth qualitative research through case studies of exemplar schools to understand what is happening within the buildings that seems to lead to these results.

We are aware that the measures identified through the REMIQS framework will not fully escape the measurement and interpretation challenges inherent to test scores. Many of the school quality measures come from existing data sources that tabulate student and school characteristics and outcomes with questionable validity and reliability. But by triangulating across several measures and through multiple modes of inquiry (i.e., quantitative and qualitative approaches), this study aims to make a meaningful and novel contribution to the discussion of school quality and to identify best practices that seem to set students up for success, along various trajectories. This study will also reveal strengths and drawbacks of currently available data and measurement to give recommendations to data collection entities (states and beyond) about how to measure quality better.

Ultimately, we would like to affect how researchers, administrators, policymakers, and the public think about school quality and to reveal what schools are doing to serve students well in ways that are underappreciated by current measures. This could inform education practice for students historically underserved by public education, particularly for state- and district-level stakeholders. Also, though this project does *not* seek to develop an alternative accountability system, the findings could help states consider how they want to structure their tracking systems in the current post-NCLB era.

Notes

1 Race to the Top guidelines provide special consideration for high-minority and high-poverty schools, but the regulations leave substantial discretion for these definitions to states (US Department of Education 2009).

2 "Race to the Top District Competition Draft: Definitions," US Department of Education, accessed February 18, 2019, https://www.ed.gov/race-top/district-competition/definitions.

3 These figures come from original analysis of the Common Core of Data.

4 For a useful summary of educational measurement since the early 20th century, see Conley (2014).

5 Sarah D. Sparks, "Panel Finds Few Learning Gains from Testing Movement," *Education Week*, May 26, 2011, https://www.edweek.org/ew/articles/2011/05/26/33academy.h30.html.

6 Alan Blinder, "Atlanta Educators Convicted in School Cheating Scandal," *New York Times*, April 1, 2015, https://www.nytimes.com/2015/04/02/us/verdict-reached-in-atlanta-school-testing-trial.html.

7 Patrick O'Donnell, "How Common Core Tests Are Scored: PARCC and Pearson Graders Can Shoot for 60 Answers per Hour," *Cleveland Plain Dealer*, May 4, 2015, https://www.cleveland.com/metro/index.ssf/2015/05/how_common_core_tests_are_scored_parcc_and_pearson _graders_can_shoot_for_60_answers_per_hour.html.

8 Evie Blad, "Accountability Measures for Traits Like 'Grit' Questions: Researchers Warn of Unreliable Results," *Education Week*, May 19, 2015, https://www.edweek.org/ew/articles/2015/05/20/accountability-measures-for- traits-like-grit-questioned.html; Evie Blad, "Moving beyond Just Academics in Assessing Effectiveness," *Education Week*, December 30, 2015, https://www.edweek.org/ew/articles/2016/01/07/moving-beyond-just- academics-in-assessing-effectiveness.html; Evie Blad, "No State Will Measure Social-Emotional Learning under ESSA. Will That Slow Its Momentum?" *Education Week*, October 4, 2017, https://www.edweek.org/ew/articles/2017/10/04/no-state-will-measure-social-emotional-learning- under.html.

9 See also Matt Chingos and Kristin Blagg, "School Funding: Do Poor Kids Get Their Fair Share?" Urban Institute, accessed February 15, 2019, http://apps.urban.org/features/school-funding-do-poor-kids-get-fair-share/.

References

Anderson, Charity, Ashley Cureton Turner, Ryan D. Heath, and Charles M. Payne. 2016. "On the Meaning of Grit…and Hope…and Fate Control…and Alienation…and Locus of Control…and…Self-Efficacy…and…Effort Optimism…and…." *Urban Review: Issues and Ideas in Public Education* 48 (2): 198–219.

Anderson, Theresa, Daniel Kuehn, Lauren Eyster, Burt S. Barnow, and Robert I. Lerman. 2017. *New Evidence on Integrated Career Pathways: Final Impact Report for Accelerating Opportunity*. Washington, DC: Urban Institute.

Betts, Julian R. 2001. "The Impact of School Resources on Women's Earnings and Educational Attainment: Findings from the National Longitudinal Survey of Young Women." *Journal of Labor Economics* 19 (3): 635–57.

Betts, Julian R., and Jeff Grogger. 2003. "The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings." *Economics of Education Review* 22 (4): 343–52.

Betts, Julian R., Andrew C. Zhau, and Lorien A. Rice. 2003. *Determinants of Student Achievement: New Evidence from San Diego*. San Francisco: Public Policy Institute of California.

Bonk, William, Juan Copa, Neal Gison, Thomas Gillin, John Nau, Adrian L. Peoples, Xiaoping Wang, et al. 2012. *Growth Models: Issues and Advice from the States*. Washington, DC: US Department of Education, National Center for Education Statistics.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne. 2001. "The Determinants of Earnings: A Behavioral Approach." *Journal of Economic Literature* 39 (4): 1137–76.

Campbell, Donald T. 2010. "Assessing the Impact of Planned Social Change." *Journal of Multidisciplinary Evaluation* 7 (15): 3–43.

Card, David. 1995. "Earnings, Schooling, and Ability Revisited." *Research in Labor Economics* 14:23–48.

Carter, Prudence L., and Kevin G. Welner, eds. 2013. *Closing the Opportunity Gap: What America Must Do to Give Every Child an Even Chance.* New York: Oxford University Press.

Castellano, Katherine E., and Andrew D. Ho. 2013. *A Practitioner's Guide to Growth Models.* Washington, DC: Council of Chief State School Officers.

Chingos, Matthew M., and Martin R. West. 2015. "Why Annual Statewide Testing Is Critical to Judging School Quality." Washington, DC: Brookings Institution.

Conley, David T. 2014. *A New Era for Educational Assessment.* Boston: Jobs for the Future.

Council of Chief State School Officers. 2010. "State Growth Models for School Accountability: Progress on Development and Reporting Measures of Student Growth." Washington DC: Council of Chief State School Officers.

Davis-Kean, Pamela E. 2005. "The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment." *Journal of Family Psychology* 19 (2): 294–304.

Dewey, John. 1897. "My Pedagogic Creed." *The School Journal* 54 (3): 77–80.

Dixon-Román, Ezekiel. 2016. "Diffractive Possibilities: Cultural Studies and Quantification." *Transforming Anthropology* 24 (2): 157–67.

Farrington, Camille A., Melissa Roderick, Elaine Allensworth, Jenny Nagaoka, Tasha Seneca Keyes, David W. Johnson, and Nicole O. Beechum. 2012. *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance: A Critical Literature Review.* Chicago: University of Chicago Consortium on Chicago School Research.

Foley, Ellen, Jacob Mishook, Joanne Thompson, Michael Kubiak, Jonathan Supovitz, and Mary Kaye Rhude-Faust. 2008. *Beyond Test Scores: Leading Indicators for Education.* Providence, RI: Brown University, Annenberg Institute for School Reform.

Gergen, Kenneth J., and Ezekiel Dixon-Román. 2014. "Social Epistemology and the Pragmatics of Assessment." *Teachers College Record* 116 (11): 1–22.

Gutiérrez, Rochelle, and Ezekiel Dixon-Román. 2010. "Beyond Gap Gazing: How Can Thinking about Education Comprehensively Help Us (Re)envision Mathematics Education?" In *Mapping Equity and Quality in Mathematics Education*, edited by Bill Atweh, Mellony Graven, Walter Secada, and Paola Valero, 21–34. New York: Springer.

Hanushek, Eric. 2014. "Achievement Gap." In *Encyclopedia of Educational Theory and Philosophy*, edited by D. C. Phillips, 4–7. Los Angeles: SAGE.

Harris, Douglas N. 2006. *Ending the Blame Game on Educational Inequity: A Study of "High Flying" Schools and NCLB.* Tempe: Arizona State University, Education Policy Studies Laboratory, Education Policy Research Unit.

Haskins, Ron, Richard Murnane, Isabel V. Sawhill, and Catherine Snow. 2012. "Can Academic Standards Boost Literacy and Close the Achievement Gap?" Princeton, NJ: Trustees of Princeton University.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3): 411–82.

Heller, Rafael, and Rebecca E. Wolfe. 2015. "Effective Schools for Deeper Learning: An Exploratory Study." Boston: Jobs for the Future.

Ho, Andrew. 2008. "The Problem with 'Proficiency': Limitations of Statistics and Policy under No Child Left Behind." *Educational Researcher* 37 (6): 351–60.

———. 2012. "Accuracy, Transparency, and Incentives: Contrasting Criteria for Evaluating Growth Models." College Park: Maryland Assessment Research Center for Education Success.

Ho, Andrew, Daniel M. Lewis, and Jason L. MacGregor Farris. 2009. "The Dependence of Growth-Model Results on Proficiency Cut Scores." *Educational Measurement: Issues and Practice* 28 (4): 15–26.

Jackson, Clement Kirabo, Rucker C. Johnson, and Claudia Persico. 2016. "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms." *Quarterly Journal of Economics* 131 (1): 157–218.

Koretz, Daniel. 2017. *The Testing Charade: Pretending to Make Schools Better.* Chicago: University of Chicago Press.

Kyllonen, Patrick C. 2012. "Measurement of 21st-Century Skills within the Common Core State Standards." Paper presented at the Invitational Research Symposium on Technology Enhanced Assessments, Washington, DC, May 7–8.

Lachlan-Haché, Lisa, and Marina Castro. 2015. "Proficiency or Growth? An Exploration of Two Approaches for Writing Student Learning Targets." Washington, DC: American Institutes for Research.

Ladson-Billings, Gloria. 1999. "Just What Is Critical Race Theory and What's It Doing in a Nice Field Like Education?" In *Race Is...Race Isn't: Critical Race Theory and Qualitative Studies in Education*, edited by Lawrence Parker, Donna Deyhle, and Sofia Villenas, 7–30. Boulder, CO: Westview Press.

———. 2006. "From the Achievement Gap to the Education Debt: Understanding Achievement in US Schools." *Educational Researcher* 35 (7): 3–12.

Levin, Henry M. 2012. "More Than Just Test Scores." *Prospects: Quarterly Review of Comparative Education* 42 (3): 269–84.

Lewis, Mattnew W., Rick Eden, Chandra Garber, Mollie Rudnick, Lucrecia Santibañez, and Tiffany Tsai. 2014. *Equity in Competency Education: Realizing the Potential, Overcoming the Obstacles*. Boston: Jobs for the Future.

Long, Mark C. 2016. "Generating Hypotheses and Upper-Bound Effect Sizes Using a Large Number of School Characteristics and Student Outcomes." *Journal of Research on Educational Effectiveness* 9 (1): 128–47.

Mehta, Jal, and Sarah Fine. 2015. *The Why, What, Where, and How of Deeper Learning in American Secondary Schools*. Boston: Jobs for the Future.

Milner, H. Richard, IV. 2013. "Analyzing Poverty, Learning, and Teaching through a Critical Race Theory Lens." *Review of Research in Education* 37 (1): 1–53.

Noguera, Pedro, Linda Darling-Hammond, and Diane Friedlander. 2015. *Equal Opportunity for Deeper Learning*. Boston: Jobs for the Future.

Owens, Ann, Sean F. Reardon, and Christopher Jencks. 2016. "Income Segregation between Schools and School Districts." *American Educational Research Journal* 53 (4): 1159–97.

Preston, Courtney, Ellen Goldring, J. Edward Guthrie, Russell Ramsey, and Jason Huff. 2016. "Conceptualizing Essential Components of Effective High Schools." *Leadership and Policy in Schools* 16 (4): 525–62.

Reardon, Sean F. 2011. "The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations." In *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murname, 91–116. New York: Russell Sage Foundation.

Reardon, Sean F., Demetra Kalogrides, and Andrew D. Ho. 2017. *Linking US School District Test Score Distributions to a Common Scale*. Working Paper 16-09. Palo Alto, CA: Stanford Center for Education Policy Analysis.

Rose, Heather, and Julian R. Betts. 2004. "The Effect of High School Courses on Earnings." *Review of Economics and Statistics* 86 (2): 497–513.

Rutledge, Stacey A., Lora Cohen-Vogel, La'Tara Osborne-Lampkin, and Ronnie Roberts. 2015. "Understanding Effective High Schools: Evidence for Personalization for Academic and Social Emotional." *American Educational Research Journal* 52 (6): 1060–92.

Schneider, Jack, Rebecca Jacobsen, Rachel S. White, and Hunter Gehlbach. 2017. "Building a Better Measure of School Quality." *Phi Delta Kappan* 98 (7): 43–48.

Schweig, Jonathan, Laura S. Hamilton, Brian M. Stecher, and Garrett Baker. 2017. "Building a Repository of Social and Emotional Learning Assessments." Paper presented at the 39th Association for Public Policy Analysis and Management Fall Research Conference, Chicago, IL, November 2.

Stiggins, Rick. 2005. "From Formative Assessment to Assessment for Learning: A Path to Success in Standards-Based Schools." *Phi Delta Kappan* 87 (4): 324–28.

Tichnor-Wagner, Ariel, Christopher Harrison, and Lora Cohen-Vogel. 2016. "Cultures of Learning in Effective High Schools." *Educational Administration Quarterly* 52 (4): 602–42.

Trujillo, Tina, and Michelle Renée. 2012. "Democratic School Turnarounds: Pursuing Equity and Learning from Evidence." *Voices in Urban Education* 36:18–26.

US Department of Education. 2009. "Race to the Top Program: Executive Summary." Washington, DC: US Department of Education.

———. 2013. "Measuring Teaching Matters: What Different Ways of Looking at Student Results Tell Us about Teacher Effectiveness." Washington, DC: US Department of Education.

West, Martin R., Matthew A. Kraft, Amy S. Finn, Rebecca E. Martin, Angela L. Duckworth, Christopher F. O. Gabrieli, and John D. E. Gabrieli. 2016. "Promise and Paradox: Measuring Students' Noncognitive Skills and the Impact of Schooling." *Educational Evaluation and Policy Analysis* 38 (1): 148–70.

About the Authors

Theresa Anderson is a research associate in the Income and Benefits Policy Center at the Urban Institute, where she works primarily on completing evaluations of workforce development programs. Anderson is pursuing a PhD in public policy and public administration at the George Washington University, where she earned her MPP. She also received a BA from Hampshire College in Massachusetts.

David C. Blount is a research analyst in the Income and Benefits Policy Center. His areas of work include education and youth development, youth employment, race and gender equity, and systemic barriers to mobility out of poverty.

Constance A. Lindsay is a research associate in the Center on Education Data and Policy at the Urban Institute, where she studies K–12 education policies. Lindsay earned a doctorate in human development and social policy from Northwestern University,

Erica Blom is a research associate in the Center on Education Data and Policy, where she studies higher education policy. Blom received a bachelor's degree in mathematics and political science from Queen's University and a master's degree in economics from Western University. She also earned a doctoral degree in economics from Yale University

Semhar Gebrekristos is a research analyst in the Income and Benefits Policy Center. She conducts quantitative and qualitative research on such topics as career pathways and the intersection of workforce development and child care. She holds a bachelor's degree in economics from Mount Holyoke College.

Francisca Alba, formerly of the Urban Institute, is a senior research assistant in the Economic Studies program at the Brookings Institution. Previously, she interned at the Center on Budget and Policy Priorities and the Urban Institute. She earned a master's degree in public policy with a specialization in social policy from the Trachtenberg School of Public Policy and Public Administration at the George Washington University.

Acknowledgments

URBAN
INSTITUTE

2100 M Street NW
Washington, DC 20037

www.urban.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.