Equity in Measuring School Quality

Comparing the Robust and Equitable Measures to Inspire Quality Schools (REMIQS) Framework and State Accountability Systems in Arizona, Kentucky, Massachusetts, Texas and Virginia

by Cailyn Torpie, Noman Khanani, Bradley Quarles and Mary Rauner of WestEd

Robust and Equitable Measures to Inspire Quality Schools (REMIQS)
is a KnowledgeWorks project undertaken in partnership with WestEd.

KnowledgeWorks is a national nonprofit organization advancing a future of learning that ensures each student graduates ready for what's next. For more than 20 years, we've been partnering with states, communities and leaders across the country to imagine, build and sustain vibrant learning communities. Through evidence-based practices and a commitment to equitable outcomes, we're creating the future of learning, together. KnowledgeWorks.org

WestEd – a nonpartisan, nonprofit research, development, and service agency – works with education and other communities throughout the United States and abroad to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd.org

Suggested citation: Torpie, C., Khanani, N., Quarles, B., & Rauner, M. (2023). Equity In measuring school quality: Comparing the Robust and Equitable Measures to Inspire Quality Schools (REMIQS) framework and state accountability systems in Arizona, Kentucky, Massachusetts, Texas and Virginia. WestEd; KnowledgeWorks.

CONTENTS

INTRODUCTION

State education systems have a variety of ways to measure success and considerable flexibility in how they hold schools accountable for student performance. This report compares several frameworks to measure education outcomes and school success in order to understand the equity implications that different frameworks may have for historically resilient and marginalized students in particular. The report presents analyses comparing the accountability systems of five selected states against a measurement framework developed by the Robust and Equitable Measures for Inspiring Quality Schools (REMIQS) project.

> Findings from these analyses are intended to be useful for state-level policymakers interested in critically examining the equity implications of their accountability systems and reimagining their methods and measures.

The REMIQS project was designed to identify and understand the practices of high schools in traditional settings[1] that demonstrate strong outcomes for historically resilient and marginalized students. Through the REMIQS project, KnowledgeWorks and WestEd partnered to conduct quantitative and qualitative research to learn how schools can promote strong academic, postsecondary, career and civic outcomes for historically resilient and marginalized students and how the schools that serve these students can promote social justice and equity. KnowledgeWorks staff, the REMIQS Advisors and Stakeholder Committee members[2] provided guidance to the WestEd team on all project activities.

The REMIQS project was motivated by recognizing that test scores continue to be the primary measure for school accountability. Prior studies on test score measures have observed a strong correlation between family background and student performance.[3] Further, because families with similar demographics, household incomes and levels of formal education tend to live near each other and enroll their children in the same schools, schools serving high percentages of historically resilient and marginalized students are less likely to reach a high standardized achievement level than schools that predominantly serve advantaged students (DeLuca et al., 2016; Duncan & Murnane, 2011). Following this logic, test score measures may more closely reflect neighborhood characteristics than a school's unique impact on student learning, which can have a disproportionately negative impact on students from historically resilient and marginalized student groups. This accountability paradigm incentivizes schools to concentrate on "teaching to the test" (Cuban, 2013; Schneider, 2014).

Every Student Succeeds Act

Authorized in 2015, the federal Every Student Succeeds Act (ESSA) requires states to collect and publicly report school performance in several indicators: Academic Achievement, Academic Progress, Graduation Rate, Progress in Achieving English Language Proficiency (ELP) and School Quality or Student Success (SQSS; Every Student Succeeds Act, 2015). States must collect and report measurements on all of these indicators for all enrolled students and for economically disadvantaged students, students with disabilities, students designated as English Learners[4] and students of every major racial/ethnic group.[5] States often report on additional subgroups of interest based on their local contexts.

Importantly, ESSA requires state accountability systems to identify schools in need of improvement as follows:

» Comprehensive Support and Improvement: Schools that are either in the bottom five percent of student performance in Title I schools (CSI) or have a graduation rate of 67 percent or lower (CSII)

» Targeted Support and improvement (TSI): Schools that are "consistently underperforming" for any group of students, as defined by the state

» Additional Targeted Support and Improvement (ATSI): Schools with disproportionately low performance for one or more student groups

In addition to identifying schools requiring these levels of support, ESSA requires that states report each school's performance on all accountability measures through annual public report cards. These report cards must include both aggregate student performance on each measure and aggregate performance of each subgroup identified in state ESSA plans. Many states also calculate summative school ratings that are weighted composites of their accountability measures, though states are not required to do so. Under ESSA, states are required to weigh Academic Achievement measures more heavily than other indicator measures in their school ratings.

The flexibility offered by ESSA has resulted in considerable variation in how states hold schools accountable for student performance. Namely, states have flexibility in the metrics they choose to hold schools accountable and the weight of each indicator measure score. For example, some states may use five-year graduation rates instead of four-year rates when holding schools accountable for high school completion. Additionally, regarding math and reading performance, some states may weigh student growth higher than current achievement. These decisions impact the ratings that schools receive, which in turn may have implications for how equity is conceptualized and achieved (Schneider et al., 2017).

REMIQS Methodology

This report compares the accountability systems of five states against a methodology developed by the REMIQS project. The REMIQS methodology was developed by WestEd in partnership with KnowledgeWorks to help understand which high schools consistently promote positive outcomes for the most vulnerable students in five states: Arizona, Kentucky, Massachusetts, Texas and Virginia.[6] The REMIQS project developed a methodology for identifying schools' impact on various measures such as assessment scores, graduation rates and college-going rates. The REMIQS methodology limits its universe to students from historically resilient and marginalized backgrounds and to schools in which at least 25 percent of students are from these backgrounds.[7]

In a previous study, REMIQS project staff applied the methodology to all nonselective conventional public or charter high schools in the five selected states. Only schools that averaged 100 or more ninth grade students were examined. The project was based on the idea that schools can provide value to students in ways that achievement on standardized test scores does not measure and on the notion that conventional accountability ratings are more reflective of school demographics than of actual school quality. Results from the REMIQS project's analyses of school measures in the five selected states are incorporated into the study that is the focus of this paper.

Research Questions

This paper's investigation builds on the previous analyses by detailing how the REMIQS methodology and its school composite scores compare to the official state accountability system methodologies and school ratings of the same five states: Arizona, Kentucky, Massachusetts, Texas and Virginia. In particular, the investigation focuses on the equity implications of these different methods for assessing high school quality.

The following research questions guided this investigation:

» Research Question 1: How is the REMIQS methodology similar to and different from accountability systems in each state and what are the equity implications of these similarities and differences?

» Research Question 2: How do the REMIQS school composite scores compare to schools' rankings based on accountability methodologies employed in each state?

First, this paper compares the methodologies across each state and the REMIQS framework, organized by the five ESSA-mandated accountability indicators. Next, the paper compares summative school ratings, including a specific focus on how student groups are considered in each state and the REMIQS framework. Lastly, the paper describes the differences in the summative school ratings that result from each state's approach as they compare to schools' REMIQS composite scores.

Data

To address these questions, the research team referenced the accountability methodologies that states proposed to the U.S. Department of Education in their 2017 ESSA State Plan submissions.

To compare the REMIQS school composite scores with that of the states, the team analyzed publicly available state accountability data for high schools from the 2018–19 school year. These data reflect the policies that were in place immediately after the passing of ESSA and before the COVID-19 pandemic.[8] Although federal and state accountability policies have shifted since that time, findings from this analysis can provide policymakers with insights into how accountability policies measure educational equity and provide them with strategies to make their systems more accountable for improving historically resilient and marginalized students' schooling experiences and outcomes (Finn, 2022).

Comparing Methodologies (Methodological Comparison)

Comparing the methodology of the REMIQS framework with each state's accountability system as it relates to high school (Research Question 1) involved examining the accountability metrics used, the approaches to aggregating metrics into summative school ratings and the equity implications of how each state or REMIQS factors in the performance of individual student groups.

The research team analyzed the metrics included in each state's accountability system and in the REMIQS framework by the five ESSA indicators: Academic Achievement, Academic Progress, Graduation Rate, Progress in Achieving English Language Proficiency (generally shortened to English Language Proficiency or just ELP) and School Quality or Student Success (SQSS; Every Student Succeeds Act, 2015). For each indicator, this paper describes the student universe, academic subjects, metrics and weights applied to the summative school ratings by each state and REMIQS. Following the summary of each indicator is a section on the equity implications of the different methods that the states and REMIQS used for the indicator.

INDICATOR: ACADEMIC ACHIEVEMENT

ESSA requires states to assess Academic Achievement through performance on annual state assessments and to weigh ELA and math equally. There were, however, differences in the ways that the five states and REMIQS defined the student universe, the subjects that were measured, the metrics used and the weight attributed to the Academic Achievement indicator in summative school ratings (Table 1).

Table 1. Academic Achievement Indicator: Student Universe, Metrics and Weight, by Methodology

Methodology	Student Universe	Academic Subjects	Metrics	% Weight of Summative Rating
REMIQS	Students from historically resilient and marginalized groups, in schools with at least 25 percent of the student population from these groups, who took required high school assessments, excluding ninth grade repeaters and students without results from eighth grade assessments	• English Language Arts (ELA) • Math	Value-added model	10%
Arizona	Students enrolled in end-of-course assessment subjects for the full academic year	• ELA • Math • Science (bonus points only)	Proficiency weighted at the student level: • Minimally proficient: 0 • Partially proficient: 0.6 • Proficient: 1 • Highly proficient: 1.3 SAT and/or ACT (for schools that select)	30%
Kentucky	Students enrolled in tenth grade for the full academic year	• Reading and Language Arts (RLA) • Math	Proficiency weighted at the student level by performance level: • Novice: 0 • Apprentice: 0.5 • Proficient: 1 • Distinguished: 1.25	45%
Massachusetts	Students enrolled in tenth grade for the full academic year	• ELA • Math • Science	Average scaled scores	40%
Texas	All students who took end-of-course exams	• ELA • Math • Science • U.S. History	Proficiency in end-of-course exams	28%[9]
Virginia	Students in the twelfth grade cohort	• RLA • Math	Proficiency rates	NA[10]

Overview of Academic Achievement Indicator for REMIQS and Each State

The state systems and REMIQS framework in this analysis varied in their approaches to defining the student universe included in the Academic Achievement indicator. The REMIQS framework included students from historically resilient and marginalized groups who were enrolled in schools with at least 25 percent of the student population from these groups. From this student population, the Academic Achievement measures included all students with assessment results, excluding ninth grade repeaters and students without eighth grade assessment history (Table 1).

Both Kentucky and Massachusetts included all tenth grade students who enrolled for the full academic year. Virginia included all students in the twelfth grade cohort, which includes students' performance on all statewide assessments taken at any point in their high school career. Texas and Arizona included students in all grades who were enrolled for the full academic year in the course-based assessments subjects and were thus required to participate in end-of-year course exams.[11]

The subjects included in measures of Academic Achievement also varied by state system or framework (Table 1). All states and REMIQS included math and either ELA or RLA. In addition, Massachusetts included science and Texas included science and U.S. History. Although Arizona did not include science in its Academic Achievement indicator, it did allow for up to an extra three "bonus" points to be added to schools' summative ratings (out of 100) based on how schools performed on the statewide science assessment.

Each state and REMIQS used a unique set of metrics for this indicator, highlighting the diversity in approaches to measuring Academic Achievement. REMIQS measured Academic Achievement through a value-added model to isolate each school's effect on students' standardized performance score (Table 1). Texas and Virginia used a binary proficiency metric (i.e.,

not proficient = 0; proficient = 1). Massachusetts used average scaled scores rather than a binary metric of proficiency in each subject area to better represent the distribution of scores at the district and school levels (Commonwealth of Massachusetts Revised Consolidated State Plan Under the Every Student Succeeds Act, 2017). Arizona and Kentucky used nonbinary proficiency measures. These states also gave partial weight to students who earned partially proficient scores and extra weight to students whose performance levels were above each respective state assessment's defined threshold for proficiency. In Arizona, "minimally proficient" students received a weight of 0, "partially proficient" students received a weight of 0.5, "proficient" students received a weight of 1 and "highly proficient" students received a weight of 1.3. Kentucky weighted "novice" students at 0, "partially proficient" at 0.5, "proficient" students at 1 and "distinguished" students who surpassed proficiency at 1.25.

Academic Achievement comprised 10 to 45 percent of the total summative school ratings across REMIQS and the five states (Table 1). REMIQS allocated the lowest weight at 10 percent. Texas and Arizona allocated more than twice the weight of REMIQS, at 28 percent and 30 percent, respectively. Massachusetts (40%) and Kentucky (45%) allocated the highest weight percentages for Academic Achievement. Virginia did not publicly report the weight for Academic Achievement or any other indicator.

Equity Implications (Academic Achievement)

The differences among states and REMIQS in their student universes, academic subjects, metrics and weights uncover several dimensions of equity that ultimately impact the states' composite measures of school quality.

Including All Students
Understanding the student universe, which defines the student population of a given indicator, is critical to interpreting the meaning of the metrics used for any indicator. Systematically

omitting underrepresented student groups from metric calculations can undermine the validity of the metric (Seastrom, 2017).

Among all schools and students that were a part of the REMIQS framework, REMIQS included data for all students in the Academic Achievement indicator with two exceptions.[12] This decision was guided by prior research on the negative association between student mobility and academic performance and by research documenting that students who change schools during a school year have lower levels of academic achievement (Schwartz et al., 2015). Research has also documented that, on average, historically resilient and marginalized students are more likely than advantaged students to be highly mobile, and schools with high proportions of mobile students have high proportions of historically resilient and marginalized students (Rose, 2013).

Like the REMIQS framework, some state accountability systems included all students from a grade-specific cohort (Virginia) or students who took final assessments (Texas). However, only students who were enrolled for the full academic year are included in the metrics for Arizona, Kentucky and Massachusetts. It is possible that policymakers from these states may have included the full academic year requirement to limit accountability to the performance of students for whom the school had adequate time to impact, sometimes referred to as a "dosage effect." Ultimately, there is a tradeoff between prioritizing a school's dosage effect on students and ensuring that schools are accountable for all students, particularly historically resilient and marginalized students, regardless of their length of enrollment.

Adding Academic Subjects

ESSA requires states to include ELA or reading/language arts (RLA) and math metrics in the Academic Achievement indicator. REMIQS, Arizona, Kentucky and Virginia included performance in ELA/RLA and math. Other states expanded the Academic Achievement indicator beyond ELA/RLA and math.

Specifically, Massachusetts and Texas added student performance in science; Texas also added students' course exams in history. More expansive definitions of Academic Achievement such as these can capture a broader and more diverse set of concepts, competencies, understandings and skills, which could be considered more equitable (York et al., 2019). The REMIQS framework was limited to achievement data in ELA and math because those data were available for all states in the study.

Measuring Academic Achievement Beyond Proficiency

Virginia and Texas used raw proficiency rates as their metrics for Academic Achievement, measuring the percentage of students who meet the cut point for proficiency on their statewide assessments. Arizona, Kentucky and Massachusetts opted for methods beyond these raw measures. Arizona and Kentucky weighted scores at the student level. Both states reported that they applied student-level weights to incentivize schools to move students from below proficiency closer to proficiency and to give schools credit for helping get students to the highest achievement level beyond proficiency (Arizona, 2019; Commonwealth of Kentucky Revised Consolidated State Plan Under the Every Student Succeeds Act, 2017). Massachusetts used students' average scaled scores to measure achievement because proficiency rates do not capture the full range of performance in schools and districts (Commonwealth of Massachusetts Revised Consolidated State Plan under the Every Student Succeeds Act, 2017). For example, there may be a great concentration of students who fall just below the cut point for proficiency in one school, whereas most students in another school may fall well below the threshold for proficiency. Proficiency rates for these two schools may be the same, whereas their average scaled scores would be quite different. Ultimately, raw measures of proficiency can obscure important differences and therefore may not capture the most valid and valued measures of students' academic performance.

Controlling for Prior Performance

The REMIQS framework used a value-added model to measure the impact of schools on students' academic performance. To isolate the effect of the school on performance, this model controlled for prior performance on assessments as well as student and school-level characteristics such as student demographics and schools' demographic composition. However, the measures of achievement employed by all five states in this analysis do not factor in students' prior achievement or student characteristics. Therefore, the influences of those factors on achievement were effectively included in those states' metric scores (Institute of Education Sciences Statewide Longitudinal Data Systems Grant Program, 2012). By not controlling for prior performance, it is difficult to assess the degree to which schools contributed to students' current performance. Consequently, the Academic Achievement measures used in the state systems may not accurately reflect the quality of the schools.

Weighing Academic Achievement in Summative School Ratings

All states in the analysis weighted Academic Achievement at a higher percentage of the summative school ratings than the REMIQS school composite score (10%). The decision to minimize the weight of Academic Achievement for the REMIQS school composite score was grounded in the literature that shows a strong association between test scores and student demographics—students from historically marginalized backgrounds, on average, have lower scores on standardized assessments (Hanushek, 2014; Reardon, 2011). This association is partly due to bias in the content and format of assessments, leading to legitimate questions of validity (Anderson et al., 2019). When Academic Achievement is weighted more heavily than other indicators, districts and schools can be incentivized to focus on reductive measures of student achievement such as standardized tests rather than more holistic measures of students' knowledge and understanding. This tendency can promote a constraining "teach to the test" culture in schools and systems that has been widely critiqued in the No Child Left Behind era in which such high-stakes testing was (and is) widespread (Jennings & Bearak, 2014). Given the equity concerns with Academic Achievement metrics, the higher the weight attributed to this indicator, the greater the equity concern for the summative school rating overall.

INDICATOR: ACADEMIC PROGRESS

Academic Progress includes growth metrics for individual students or improvements in outcomes for student groups within a school. Notably, while measures for this indicator leverage the same student assessments as Academic Achievement, they foreground different dimensions of school quality. Academic Achievement measures whether a student's achievement is "good enough" per state standards at a specific time. By contrast, Academic Progress focuses on learning over time, emphasizing the knowledge and competencies students learn rather than what they can demonstrate by the end of the year.

ESSA affords states considerable discretion for measuring Academic Progress and does not require states to include progress measures as part of their high school ratings. Despite this weak mandate, and perhaps spurred on by policy debates questioning whether proficiency rates are effective school performance measures, many states' high school ratings in this study included Academic Progress measures.

The REMIQS framework focused on school effects by controlling for prior performance across measures and controlling for student and school demographics. However, REMIQS did not explicitly include Academic Progress measures in its school composite scores. Kentucky and Virginia also did not incorporate any measures of Academic Progress in their school ratings. Arizona, Massachusetts and Texas measured Academic Progress using different metrics, student universes and weights (Table 2).

Table 2. Academic Progress Indicator: Student Universe, Metrics and Weight, by Methodology

Methodology	Student Universe for Growth	Metrics	% Weight of Summative Rating
REMIQS[13]	NA	NA	NA
Arizona	Student groups with at least 10 students enrolled for the full academic year (FAY) in the current year and prior year The number of eligible student groups is the denominator; the number of student groups that showed improvement is the numerator	• Subgroup Improvement: proficiency in ELA and math • Subgroup Improvement: graduation rate • Subgroup Improvement: dropout rate	30%
Kentucky	NA	NA	NA
Massachusetts	Students enrolled in tenth grade	Student growth percentiles in ELA and math	20%
Texas	All students who took English II and Algebra I end-of-course exams	Percent of students who maintained proficiency or met growth expectations in English II and Algebra I end-of-course exams	70%[14]
Virginia[15]	NA	NA	NA

Overview of Academic Progress Indicator for REMIQS and Each State

The states that included measures of Academic Progress took a year-over-year approach to operationalizing growth. In doing so, each state restricted its student universe to students with consecutive years of achievement data. Massachusetts and Texas measured Academic Progress via student growth on statewide assessments. Therefore, all students who were included in the Academic Achievement student universe for those assessments were also included in the Academic Progress indicator. But as it focused on year-over-year student growth, the Academic Progress indicator also required students to have belonged to the Academic Achievement student universe in the prior year as well. Arizona's Academic Progress metrics included all student groups with two years of data (the current school year and the preceding school year's Academic Progress reporting universe).

Arizona's Academic Progress indicator was comprised of several Subgroup Improvement measures. In contrast to the focus on individual student growth in Massachusetts and Texas, Subgroup Improvement calculated a school's percentage of student groups performing at a higher level than the prior school year for ELA and math assessments, graduation rates and dropout rates. Importantly, the students whose scores contributed to the current year performance of a given student group may not have been the same students for whom performance was measured for the prior year. Because graduation rates were calculated for individual cohorts of students year-over-year, the student groups were comprised of different individual students for the graduation Subgroup Improvement metric.

Massachusetts measured Academic Progress through student growth percentiles (SGPs), which were relative measures of how students performed year-over-year. SGPs did not measure growth toward a particular standard. In contrast, Texas' Academic Progress was based on whether students met an explicit target for growth in ELA and math, indicating a student was on track for gaining proficiency.

Both Arizona and Massachusetts weighed Academic Progress at 20 percent of their overall systems. In Texas, Academic Progress could contribute up to 70 percent of the weight of a school's rating if the school's Academic Progress score was higher than its Academic Achievement score.

Equity Implications (Academic Progress)

How frameworks and systems do or do not measure Academic Progress, or growth, can carry significant implications for the equity of the system overall.

Controlling for Prior Performance

Controlling for prior performance, as the REMIQS framework did across indicators, can do a better job of capturing schools' impact than not controlling for prior performance. SGPs also controlled for prior performance, but unlike the REMIQS framework, they did not control for student and school demographics. Controlling for both prior performance and student demographics allows REMIQS to better isolate and measure a school's impact on the students it serves, which is particularly important for schools that serve a high proportion of students from historically resilient and marginalized backgrounds.

Student-level Progress Versus Group-Level Improvement

Although measuring improvement of individual student groups could be considered an equitable aim, relying on aggregate improvement produces a less precise measure of school impact than student-level progress. If metrics compare the performance of different groups of students year-over-year, there is a risk of a higher degree of "noise," or external factors, influencing changes in performance that may not be a direct result of changes in school quality.

Relative Versus Criterion-referenced Growth

Value-added measures in REMIQS were not classified as Academic Performance measures within its framework, but, as with SGPs, value-added models produce relative measures of performance. Both sets of metrics assessed an individual student against other students included in the model, not against an external, objective standard. Relative measures of growth communicate individual student's growth and, in the aggregate, schools' growth in academic performance compared to other students and schools within the same system. And since these models controlled for prior performance, they could more effectively capture school impact. However, these measures did not capture students' progress toward attaining proficiency, nor did they explicitly reward progress among historically marginalized student groups that resulted from narrowing opportunity gaps.

INDICATOR: GRADUATION RATE

ESSA requires that accountability systems include high school graduation measures. The U.S. Department of Education established guidelines for calculating graduation to ensure these measures are "uniform and accurate...and comparable across states and consistently reported over time" (U.S. Department of Education, 2017). Given this prescription, REMIQS and the states in this analysis took similar approaches to operationalizing graduation (Table 3).

Table 3. Graduation Rate Indicator: Student Universe, Metrics and Weight, by Methodology

Methodology	Student Universe for Graduation Rate	Metrics	% Weight of Summative Rating
REMIQS	Students from historically resilient and marginalized groups, in schools with at least 25 percent of the student population from these groups who were in the first ninth grade year cohort	4-year ACGR (Adjusted Cohort Graduation Rate)	15%
Arizona	First ninth grade year cohorts[16]	• 4-year ACGR • 5-year ACGR • 6-year ACGR • 7-year ACGR	20%
Kentucky	First ninth grade year cohorts	• 4-year ACGR • 5-year ACGR	6%
Massachusetts	First ninth grade year cohorts	• 4-year ACGR • Annual dropout rate[17] • Extended engagement rate[18]	20%
Texas	First ninth grade year cohorts	Highest rate among the most recent four-year, five-year and six-year ACGR	14%[19]
Virginia	First ninth grade year cohorts	4-year ACGR	NA[20]

Overview of Graduation Rate Indicator for REMIQS and Each State

Under ESSA, the federal government requires each state to calculate and report an annual four-year Adjusted Cohort Graduation Rate (ACGR) at the school, district and state levels and by student group. Specifically, the four-year ACGR was the number of students who graduate from high school in four years with a regular high school diploma divided by the number of students who form the four-year cohort for the graduating class. An adjusted first ninth grade year cohort includes the number of students entering ninth grade for the first time plus any students entering at a later point during the ninth grade year or at any point up until the expected year of graduation, minus any students exiting the state public school system with a validated reason during the same time (U.S. Department of Education, 2017).

REMIQS and Virginia measured graduation with four-year ACGR; the other four states measured both four- and five-year ACGR. These four states also varied in the way that they accounted for students graduating on longer timelines. Arizona measured the four-, five-, six- and seven-year ACGR. Kentucky used the four- and five-year ACGR. Massachusetts included the four-year ACGR, an annual dropout rate measure and an extended engagement rate. Texas measured the four-, five- and six-year ACGR and, using a best-of approach, included the highest rate of the three in summative ratings. While these alternative measures afford additional flexibility in a student's expected time until graduation, their calculations and methodologies are similar to that of the four-year ACGR.

Graduation Rate accounted for between six percent (Kentucky) and 20 percent (Massachusetts) of each of the summative school ratings.

Equity Implications (Graduation Rate)

Among all indicators, Graduation Rate is the one for which ESSA provided states with the least flexibility in terms of measurement. Still, states exhibited discretion in determining which cohorts to assess. Considering cohorts beyond the four-year ACGR incentivizes schools to continue supporting students requiring more time and additional support to meet graduation requirements. Since learners needing more time are often highly concentrated in certain schools (e.g., alternative schools, credit recovery), they usually have relatively low four-year ACGR rates. Still, including alternative graduation measures can help to contextualize and highlight the work these schools do to meet a broad spectrum of student circumstances and ensure that their students eventually fulfill graduation requirements.

INDICATOR: PROGRESS IN ACHIEVING ENGLISH LANGUAGE PROFICIENCY

Federal nonregulatory guidance suggests ESSA "provides a valuable opportunity for states to reshape their accountability systems to support improved outcomes for ELs."[21] Specifically, ESSA Title I requires states to more fully include EL students in traditional state accountability systems instead of through separate district-level accountability systems as required under Title III of the No Child Left Behind Act of 2001 (NCLB). As NCLB did, ESSA requires that state accountability systems include ELA and math performance goals for EL students. The new ESSA requirements specify that states must have accountability goals that are related to EL progress in attaining English Language Proficiency (ELP) over time per the state's ELP assessment. In addition, ESSA requires each state's accountability system to measure progress in achieving an ELP indicator. ESSA provides flexibility in ensuring that the ELP goals and targets are attentive to each state's unique EL population. However, the states in this study developed very similar ELP indicator measures rather than innovative and context-specific ones (Table 4).

Table 4. English Language Proficiency Indicator: Student Universe, Metrics and Weight, by Methodology

Methodology	Student Universe for ELP	Metrics	% Weight of Summative Rating[22]
REMIQS[23]	NA	NA	NA
Arizona	Current EL students	Proficiency and growth for EL students on ELP assessment	10%
Kentucky	Current EL students	Proportion of EL students who attained proficiency on ELP assessment	NA
Massachusetts	Current EL students	Percentage of students demonstrating growth on ELP assessment	10%
Texas	Current EL students	Percentage of students demonstrating growth or scoring Advanced on ELP assessment	3%[24]
Virginia	Current EL students	Percentage of students demonstrating growth on ELP assessment	NA

Overview of English Language Proficiency Indicator for REMIQS and Each State

Each state reported progress for EL students during the reporting year on their respective ELP assessment. Arizona and Texas measured proficiency and growth measures, Massachusetts and Virginia used only growth measures and Kentucky included only proficiency measures (Table 4). REMIQS did not measure or account for this indicator in the school composite score because relevant data were unavailable.

Arizona, Massachusetts and Texas reported the weight of ELP measures separately in their school ratings. Kentucky's ELP measures were included in the SQSS indicator, and Virginia did not report a weight.

Arizona and Massachusetts each assigned 10 percent weight in their state summative ratings to ELP. ELP was not always included in Texas' school ratings. ELP made up three percent of Texas's Academic Performance indicator score but was part of the school ratings only if it was higher than the Academic Progress indicator score. Kentucky included ELP in its SQSS indicator but did not report the portion of the SQSS indicator scores that was allocated to ELP. Virginia did not publicly report how ELP or any other indicator scores were weighted in its summative ratings.

Equity Implications (English Language Proficiency)

There were some differences in the ways that states accounted for and measured ELP.

Achievement Versus Growth

The main source of variation in the ELP indicator pertains to using current language proficiency, growth in language proficiency or both as the metric. Proficiency—even in the context of ELP—may not accurately capture school contributions to student performance. For example, EL students in schools with larger populations of recent immigrants typically have lower levels of ELP than EL students in schools with smaller immigrant populations. Schools with large immigrant populations may therefore have lower school ratings when proficiency is used. On the other hand, using growth measures and target score achievement may offer a more equitable measure, as EL students with low ELP may still experience strong growth from one year to the next.

Measuring Academic Achievement Beyond Proficiency

The use of proficiency rates may also be sensitive to cut scores. Scores based on these thresholds can obscure important differences and therefore may not capture the most valid and valued measures of students' academic performance.

Weighing English Language Proficiency in Summative School Ratings

Because federal law did not require states to report on ELP before ESSA, it is not surprising that states allocated low weights to the ELP indicator compared to the other indicators. In general, states allocate weights equal to the proportion of EL students served. For example, a state where only five percent of students are ELL would therefore allocate five percent of the weight for school summative ratings to ELP.

Massachusetts is the only state in this analysis in which the weight of the ELP indicator matched its percentage of EL students (10 percent). Arizona also allocated 10 percent of its state score to the ELP indicator, but the percentage of EL students in Arizona is greater (14 percent). Texas serves the most EL students nationwide[25] but allocated only three percent of its school rating to ELP. These mismatches between proportions of EL students served and summative school rating weights raise questions about how well these states hold schools accountable for serving this significant and growing student group.

INDICATOR: SCHOOL QUALITY OR STUDENT SUCCESS

ESSA allows states to use a measure outside of traditional accountability measures in their school ratings as part of the SQSS indicator. The SQSS indicator measures must meaningfully differentiate school performance, be valid and reliable, be used within each grade span, be comparable across schools statewide and be reported annually for all students and student groups. Table 5 and the following sections depict and discuss the student universe, metrics and weight allocated to SQSS indicator measures for each state and REMIQS.

Table 5. School Quality or Student Success Indicator: Student Universe, Metrics and Weight, by Methodology

Methodology	Student Universe for SQSS	Metrics	% Weight of Summative Rating
REMIQS	Students from historically resilient and marginalized groups, in schools in which at least 25 percent of the student population is from these groups	Advanced coursework completion	10%
		Chronic absenteeism	5%
		School suspension	5%
		College enrollment and persistence	55%
Arizona	Students in the 2019 cohort	Self-reported college and career readiness indicator	20%
Kentucky	• Graduating class • All high school students • Students enrolled in end-of-year assessments • Current EL students	• College and career readiness • School climate and safety survey • Science, social studies, writing • ELP transition rate	34%
Massachusetts	• All high school students • Eleventh grade and twelfth grade students	• Chronic absenteeism • Advanced coursework completion	10%
Texas	Graduating students (regardless of cohort)	College, career and military readiness	28%[26]
Virginia	All high school students	• Chronic absenteeism • Standards of accreditation rating	NA%[27]

Overview of School Quality or Student Success Indicator for REMIQS and Each State

Given states' flexibility in developing SQSS metrics, there were large differences across the states and REMIQS in the student universe, metrics and weight allocated for this indicator (Table 5). Further complicating the analyses, some states included multiple measures that may have had different student universes. REMIQS used data for all high school students from historically marginalized groups. Arizona and Texas included students from the most recent graduating cohort, whereas Virginia included all students. Kentucky and Massachusetts included all students for some measures but included only certain grades or subgroups for others.

Measures that are similar to those included in the SQSS indicator accounted for most (75 percent) of the REMIQS framework's summative school rating but accounted for significantly less of the school ratings in the state systems (Table 5). The weight allocated to SQSS in Arizona was 20 percent, in Kentucky was 34 percent, in Massachusetts was 10 percent and in Texas was 28 percent for schools with Academic Achievement indicator scores that were higher than Academic Progress. Virginia did not publicly report the weight for the SQSS or any other indicator.

All frameworks except that of Virginia included at least one college and career readiness (CCR) measure in their SQSS indicator scores, though these measures and weights differed considerably (Table 5). The REMIQS CCR measure was advanced coursework completion and made up 10 percent of the school composite score. Massachusetts also included advanced coursework completion, but it accounted for a smaller percentage of the state school rating than the REMIQS school composite score because the 10 percent SQSS weight was shared with a chronic absenteeism measure. Twenty percent of Arizona's school rating was a self-reported measure of college and career readiness.

College and career readiness was one of several SQSS metrics in Kentucky, and the state did not report the weight of the individual CCR measure. Texas included a college, career and military readiness metric in its Academic Achievement indicator score. Texas's CCR metric was 28 percent of the state school rating when the Academic Achievement score was greater than the Academic Progress score.

Equity Implications (School Quality or Student Success)

The following sections discuss equity implications concerning the metrics, methodologies and weights that states employed for measuring SQSS.

Multiple Measures

SQSS metrics provide states an opportunity to highlight alternative dimensions of school performance and functioning that student achievement metrics cannot assess. The flexibility allotted to states through ESSA in choosing metrics for SQSS allows for the inclusion of nontraditional performance measures that may provide a more well-rounded understanding of school performance. College and career readiness, chronic absenteeism, discipline, alternative academic assessments and school climate and safety are all examples of nontraditional metrics featured in these state systems. Those that use several metrics to assess SQSS may offer more equitable outcomes for schools serving high proportions of historically resilient and marginalized students by placing less emphasis on standardized assessments.

Other nontraditional performance measures can contribute to the calculation of a more robust and equitable accounting of school quality, including:

- » Knowledge-based test scores
- » College admission test scores
- » College-track course taking
- » College-level courses taken during high school
- » Career certificate awards
- » Law enforcement referrals
- » Postsecondary enrollment and performance
- » Deeper learning skills
- » Abilities
- » Motivation
- » Physical health and well-being
- » Labor force participation
- » Poverty
- » Voting
- » Incarceration
- » Substance abuse
- » Teen parenthood

Though some of these measures, along with others, were included as part of the REMIQS logic model (Blom et al., 2020), the REMIQS school composite scores did not include these measures because sufficient data was either not available for the states in this analysis or there were concerns about consistent measurement and reporting.

Measuring College and Career Readiness

How states measure college and career readiness has equity implications. REMIQS used measures of actual college enrollment and persistence while also controlling for student characteristics and prior performance. This value-added approach reflects how schools contribute to postsecondary outcomes by controlling for some factors that may impact student ability to enroll in a postsecondary institution (such as free and reduced-price lunch as a proxy for financial means). In contrast, the states in this analysis did not include any postsecondary metrics. Rather, the state systems measured CCR through high school–level measures such as advanced coursework, college entrance exams, completing college-level classes in high school and qualifying industry credentials.

Though these measures may be highly correlated, on average, with postsecondary degree or credential attainment, there are concerns surrounding equitable access to advanced coursework and resources to prepare students for exams like the ACT and SAT (Xu et al., 2021).

Weighing School Quality or Student Success in Summative School Ratings

The weight allocated to SQSS ranged from 10 percent (Massachusetts) to 75 percent (REMIQS). The developers of the REMIQS framework justified the large weight on postsecondary success based on established evidence that individuals who attain no more than a high school diploma earn far less in the labor market than those with some postsecondary credit. Accordingly, the REMIQS developers noted that postsecondary attainment and a student's preparation for it provides the most robust perspective on the quality of secondary schooling. On the contrary, most state systems still reflected ESSA's emphasis on math and reading performance as measured by standardized assessments, which meant less weight for SQSS.

SUMMATIVE SCHOOL RATINGS

Under ESSA, each state is required to identify the lowest-performing schools in order to provide them with CSI and to identify schools with underperforming student groups to provide TSI.[28] States identify schools for CSI based on the schools' performance across metrics from each indicator, either through a summative school rating that is often shared publicly or an alternative method.

The performance of individual student groups is central to designating schools for TSI, but outside of ELP for ELs, ESSA does not require states to factor student groups into overall ratings, despite the potential impact on equity (Ho, 2008). This section describes the way in which the REMIQS framework and each state system calculated and communicated their summative school ratings and it discusses the ways in which the methodologies can lead to inequitable outcomes.

Table 6. Summative School Rating Methodology and Rating Scale, by Framework or System

Framework or System	Rating Methodology	Summative School Rating Scale
REMIQS	Highest overall score within three strata	Not shared publicly
Arizona	Totaling points earned across indicators	A–F
Kentucky	Criteria-based individual performance level by indicator, weighted, then aggregated	1–5 Stars
Massachusetts	Normative and criteria-referenced summative school rating	0–100
Texas	"Best of" between Academic Achievement and Academic Progress	A–F
Virginia	Three-step process to identify schools for Comprehensive and Targeted Support based on school performance within each indicator	No summative school rating

Measuring and Communicating Summative School Ratings

REMIQS calculated metric scores using models that capture the effect of each individual school on metrics from each indicator (Table 6).

School performance was weighted according to the assigned weights by indicator:

» 10 percent for Academic Achievement

» 15 percent for Graduation Rate

» 75 percent for metrics that fall under the category of SQSS

– Advanced coursework: 10 percent
– Attendance: five percent
– Discipline: five percent
– College enrollment and persistence: 55 percent

After all schools were assigned a composite score, they were stratified into three groups based on the percentage of students from historically marginalized backgrounds: small (25 to 49 percent), medium (50 to 74 percent) and large (75 to 100 percent). Based on the weighted composite score, the highest performing schools were identified in each of the three strata.[29]

Arizona

The Arizona Department of Education reported summative school ratings on an A–F scale (Table 6). To move from metric scores to summative school ratings, the state calculated points earned for each indicator based on the product of the indicator's weight against schools' performance (as a percentage) on metrics in the indicator.[30] Arizona then totaled the points earned in each indicator and translated the resulting score into an A–F rating based on preestablished cut-points.[31] The state also allowed for possible bonus points to be added to the summative school rating: up to two points for enrollment of students with disabilities that is greater than or equal to 80 percent of the statewide average, and up to three additional points for proficiency rates on the statewide science assessment.

Kentucky

The Kentucky Department of Education published an overall rating for each school ranging from one star (the lowest) to five stars (the highest; Table 6). A school's performance on each metric was translated into a performance level (very low, low, medium, high, very high) and weighted according to the weights assigned to each indicator. Performance levels for metrics were then aggregated into a summative school rating.[32] Kentucky's methodology for generating

summative school rating was criteria-based, meaning that schools' performance was not determined by their relative performance against other schools.

Massachusetts

The accountability system at the Massachusetts Department of Elementary and Secondary Education (DESE) calculated both a normative and criteria-referenced rating score for each school (Table 6). To calculate the normative score, schools' overall scores based on their weighted performance in each indicator were ranked and ordered by percentiles. The normative percentile score, ranging from one to 99, represented a school's aggregate performance compared to other schools serving students of the same grade span. The criteria-referenced score that each school received quantified the extent to which the school met targets for all students as well as targets for the school's lowest performing student group. Every year, DESE sets improvement targets for all students in each of the indicators. For the lowest performing subgroup at each school, DESE set improvement targets for three indicators: Academic Achievement, Academic Growth and Chronic Absenteeism. Based on each target and actual performance, DESE assigned points for each indicator as declined, no change, improved, met target or exceeded target.[33] The number of points that each school received for the all-student group and the lowest performing subgroup were averaged to compute the criteria-referenced score.

Texas

The Texas Education Agency (TEA) graded schools on an A–F scale. Texas employed a "best-of" methodology in calculating its summative school ratings, meaning a school's performance in either Academic Achievement or School Progress was used in its overall ratings, depending on which was higher. And for within-school progress, TEA made an additional "best-of" calculation: selecting whichever score was the higher between academic growth (extent

to which the State of Texas Assessments of Academic Readiness were achieved in math and reading) and relative performance scores (STAAR performance and College Career and Military Preparation score compared to schools with similar economically disadvantaged populations). The higher score of either Academic Achievement or School Progress made up 70 percent of the summative school rating. The remaining 30 percent was based on whether the school met achievement, graduation or growth, ELP and SQSS performance targets for eligible student groups.[34]

Virginia

Although the Virginia Department of Education did not publish summative school ratings, a detailed report on each school was made public. The state followed three steps to identify schools for CSI and TSI. In the first step, they identified schools that did not meet the interim measure of progress for two consecutive years in one or more subgroups for English (reading), math and graduation rate and were in the lowest two quartiles for academic growth in English (reading) or math. For the second step, of those schools identified in step one, the state identified schools that did not meet the interim measure for EL progress and were in the lowest two quartiles for EL progress. Step three was to take the schools identified in steps one and two and identify those that did not meet the interim measure of progress for Chronic Absenteeism for two consecutive years and had a Standards of Accreditation rating of "Accredited with Conditions" or "Accreditation Denied." The schools that were identified through all three steps were those that needed comprehensive support.

Communicating School Ratings

Many states communicated school ratings to parents and policymakers through letter grades or stars during the NCLB era (2002–15), and some continue to use this system under ESSA (Howe & Murray, 2015). The relative ease with which school quality can be communicated to parents and policymakers made this system

appealing for many states (Schneider & Hutt, 2014). Though an A–F rating system is intuitive for parents and policymakers to understand, it can overlook details such as subgroup performance and performance on individual indicators that may be of particular importance to parents. ESSA requires that states report these details through school report cards, and these report cards do not provide any specifics beyond the single letter or star grade.

Based on research suggesting that letter and star rating systems are not valid measures of school accountability or policy development, several states recently stopped communicating in these ways (Howe & Murray, 2015). Massachusetts, for example, moved from a Level 1–5 rating system to a normative and criteria-referenced rating scale after the passing of ESSA. A 0–100 scale can still be intuitive for parents while also adding more differentiation between schools. Still, a single score or rating of school quality can dissuade parents and policymakers from examining more detailed school information.

Virginia does not report school ratings, so parents and policymakers must view each school report card individually, which can lead to more in-depth review of each school. However, it can also be a cumbersome process, so audiences might seek out alternative sources to understand comparative school quality. Organizations like GreatSchools use public accountability data to create publicly available and easy-to-access school ratings. Researchers have questioned the methods used by GreatSchools and similar entities for relying on student achievement and giving short shrift to other measures of school quality (Noonan & Schneider, 2022).

Equity Implications of Measuring and Messaging Summative School Ratings

The ways that summative school ratings are constructed and communicated can be consequential for the way (in)equity is conceptualized and revealed in schools and within systems of education. This section discusses equity implications related to the relative weight of each accountability

indicator and the methods used to account for subgroups in overall ratings, and it concludes with a discussion on how these decisions risk perpetuating bias for schools serving historically resilient and marginalized students.

Relative Indicator Weights

To calculate school ratings, all states except Texas applied a higher weight to Academic Performance (as measured by state standardized assessments) than to the other indicators (Table 7). In Texas, if the Academic Progress indicator score was greater than the Academic Achievement indicator score, the majority (70 percent) of the school rating was Academic Progress (Table 7, Texas option A). If the Academic Achievement score was greater than the Academic Progress score, 28 percent of the school rating was Academic Achievement (Table 7, Texas option B). For every school, Closing the Gaps (a metric that was not associated with an ESSA indicator) was weighted at 30 percent. For REMIQS, indicators that would fall within SQSS were 75 percent of the REMIQS school composite score while only 10 percent was allocated to Academic Achievement. Except in Texas, where Academic Progress was substituted for Academic Achievement if a school had a higher Academic Progress score, the three other states weighted Academic Progress considerably lower than the other indicators. When Academic Achievement—as measured by performance on standardized tests—is weighted heavily in school ratings, the school ratings may be less reflective of school contributions to student performance and more reflective of the demographic composition of the students they serve, potentially obscuring school-level effects on student outcomes. Despite their strengths and resilience, students from these backgrounds often perform lower on standardized assessments due to factors outside of school, such as lack of access to high-quality early childhood education environments. In contrast, Academic Progress may better measure success for historically resilient and marginalized students (Chingos, 2017; Schneider, 2017).

All state systems in the analysis allocated similar weight to the Academic Achievement, Graduation Rate and ELP indicators. SQSS

Table 7. Percentage Weight of Each Indicator in School Ratings, by Framework or System

	REMIQS	Arizona	Kentucky[37]	Massachusetts	Texas Option A[38]	Texas Option B	Virginia
Academic Achievement	10%	30%	45%	40%	28%	NA	NA
Academic Progress	NA	20%	–	20%	–	70%	NA
Graduation Rate	15%	20%	6%	20%	14%	NA	NA
English Language Proficiency	NA	10%	NA	10%	3%	NA	NA
School Quality or Student Success	75%	20%	34%	10%	28%	NA	NA
Nonmandated ESSA indicator (Closing the Gaps)[36]	NA	NA	NA	NA	30%	30%	NA

had the greatest range of weight allocation, with REMIQS allocating 75 percent and Massachusetts allocating 10 percent for this indicator. Kentucky (34 percent), Texas (28 percent) and Arizona (20 percent) also had lower allocations than REMIQS did for SQSS in their school ratings, possibly due to ESSA's emphasis on academic performance.

Measuring and Including Student Groups

When school ratings include data from all student groups, the ratings can hide inequities that otherwise might be visible if data are disaggregated by student group. ESSA requires states to publicly report on student group performance across all indicators when there is a "sufficient" number of students within a given subgroup.[39] Although ESSA does not require states to account for the performance of subgroups in their summative school ratings, some choose to do so. This section describes whether, and how, student subgroup performance contributed to the summative school ratings in each state and to the REMIQS composite scores and explores the equity implications of these decisions. (See Appendix, Table A1, for the subgroups that were included in overall ratings for the REMIQS framework and state accountability systems.) The REMIQS methodology, with its explicit focus on equity, is used as a point of comparison to illustrate the equity considerations for each state system.

REMIQS

The REMIQS framework focused on the performance of historically resilient and marginalized students (Appendix, Table A1). Its school performance ratings were thus only based on the performance of students who were economically disadvantaged, Black or African American, Indigenous, Hispanic or Latino and belonging to two or more races, along with students designated as ELs and those in special education. Accordingly, REMIQS scores were only for schools in which at least 25 percent of enrolled students were historically marginalized.

Arizona

Student subgroup performance as it relates to ELA and math achievement, graduation rates and dropout rates accounted for 20 percent of Arizona's high school summative ratings. Specifically, academic proficiency was worth 10 percent, and graduation and dropout rates were each worth 5 percent. Arizona's student subgroups included the ESSA mandated racial/ethnic groups, EL students, students in special education, economically disadvantaged students, those who were military connected and students experiencing homelessness, if there were at least 10 students in a given group (see Appendix). Arizona is unique among the states in this analysis because it included students experiencing homelessness and those with a military-connected family in its Academic

Progress indicator (Subgroup Improvement) metrics. Each student group was weighted equally in Arizona's methodology.

Kentucky

In determining summative school ratings, the Kentucky Department of Education accounted for the performance gaps among students from all student groups required for public reporting by ESSA (see Appendix). Summative school ratings in Kentucky were impacted if there was a significant achievement gap between the performance of two student groups within each category (e.g., economically disadvantaged students compared to non–economically disadvantaged students). The reference group for comparison was the highest performing student group within the category. If one or more achievement gaps existed, the school's overall rating decreased by one star (e.g., a four-star school would become a three-star school). Student groups of 10 or more students were included in these analyses.

Massachusetts

DESE did not account for student subgroup performance for the state's normative measure for summative school ratings. School criteria-referenced scores (used for CSI and TSI designation), however, accounted for each school's lowest performing subgroup. The lowest performing subgroup was identified by calculating the normative rating for all of 11 possible subgroups, which encompassed all subgroups required for reporting by ESSA (see Appendix). A school had to have a minimum of 20 students in the subgroup for that subgroup to be included in the school's evaluation. The score for the lowest performing group of students made up half the total criteria-referenced score.

Texas

Student subgroup performance accounted for 30 percent of total summative school ratings in Texas through its Closing the Gaps indicator. The TEA annually sets performance targets for 14 subgroups for achievement; graduation; and college, career and military readiness. It also sets a specific EL proficiency target (see Appendix).

Subgroups that were unique to Texas, compared with the other states, include former students with disabilities and continuous/noncontinuous enrolled students. Schools were accountable for subgroups that had at least 25 students enrolled.

Virginia

Virginia did not calculate summative school ratings. The state still required schools to report subgroup performance for the metrics within each accountability indicator. Virginia reported on the least number of subgroups compared to the other states reviewed in this report and required the highest minimum group size (30) for reporting (see Appendix, Table A1).

Equity Implications of Accounting for Subgroup Performance

The REMIQS framework emphasizes the performance of historically resilient and marginalized students by including only those students in its framework and creating summative ratings for schools with a student body of least 25 percent historically resilient and marginalized students. Under ESSA, states are not able to fully mirror the REMIQS methodology, but the decisions that states make regarding how they account for student group performance in summative school ratings can have important equity implications for the conceptualization and achievement of equity.

Arizona and Texas account for the performance of each student group as a percentage of their summative school ratings, but all student groups are weighted equally. Using this methodology, low performing student groups that make up only a small proportion of a school could have an outsize impact on the school's summative school rating which could incentivize schools to focus on improving those student groups' performance.

In Kentucky, a school will receive a reduced star rating when there is a significant achievement gap between student groups (for example, EL students compared to non-EL students). This approach can incentivize schools to address gaps in opportunity, access and enrichment that can result in disproportional outcomes in Academic Achievement.

> The REMIQS framework emphasizes the performance of historically resilient and marginalized students by including only those students in its framework and creating summative ratings for schools with a student body of least 25 percent historically resilient and marginalized students.

In Massachusetts, student groups are not accounted for in the summative school ratings. Schools serving only small numbers of historically resilient and marginalized students may thus have less incentive to focus on improving those groups' performance because the groups are unlikely to impact the schools' overall ratings.

Risk of Perpetuating Bias

Assigning summative ratings to schools inherently carries equity implications, as the scores and associated school designations can impact perceptions of school quality (Barrows et al., 2016). Labeling schools as low or high performing can perpetuate and entrench bias and stereotypes that bear real impacts on the schools (Chingos et al., 2012). Negative discourse surrounding schools in communities may deter parents and families from enrolling their students in those schools and dissuade high-quality teachers from taking jobs in these schools, which could create a vicious cycle in which the ratings become endogenous to the measure of school quality itself. In other words, negative perceptions about such schools may in fact decrease school quality due to decreasing resources stemming from low enrollment as well as challenges with recruiting and retaining high-quality teachers (Feng et al., 2018).

Comparing Summative School Ratings

After exploring the methodological differences between the REMIQS framework and the state accountability systems, the research team examined the differences in the outcomes of those methods – namely, the REMIQS school composite scores and the state school ratings – to answer research question two (i.e., How do the REMIQS school composite scores compare to schools' rankings based on the state accountability methodologies employed in each state?).

Data

Three of the five REMIQS states—Kentucky, Massachusetts and Texas—are included in these analyses. Virginia was excluded because it does not publish state school rating scores and therefore could not be compared to the REMIQS school composite scores. Arizona was excluded because the data sharing agreement expired, so the team no longer had access to the data that was needed to calculate the REMIQS school composite scores.

Indicator scores and state school ratings were gathered from publicly available accountability data from each state's department of education for the 2018–19 school year. The research team used z-scores to standardize these data.

There were substantive differences across states in the way indicator measures were calculated. ESSA does not require states to align their school ratings with the requirements for public reporting on school report cards, nor does it require states to measure indicators in the same way. This report therefore refers to indicator measures using each state's sometimes unique terminology and the associated ESSA indicator name.

The specific high schools from each state that were included in these analyses are those for which REMIQS school composite scores were calculated when WestEd, in partnership with KnowledgeWorks, developed the REMIQS methodology.[40] These included all nonselective conventional public or charter high schools, each school having an average of 100 or more ninth grade students and at least 25 percent of its ninth grade student cohort from historically resilient and marginalized groups. The percentage of historically resilient and marginalized students was calculated as the percentage of students in every school who fall into one or more of the aforementioned groups. Based on these parameters, 179 high schools in

Kentucky, 209 high schools in Massachusetts and 549 high schools in Texas were included in the analysis.

Data on historically resilient and marginalized student groups were gathered from student-level data provided by the states and included the following groups:

- » EL students
- » Students in special education
- » Students who were eligible for free and reduced-price lunch
- » Black students
- » Indigenous students
- » Hispanic students
- » Students who identified as multiracial

The REMIQS population before indicator-specific exclusions consisted of students from these historically resilient and marginalized groups, and the states' populations before indicator-specific exclusions consisted of all students in the included high schools.

Methods
First, the research team used pairwise correlation tests to investigate the association between the REMIQS school composite scores and school ratings in Kentucky, Massachusetts and Texas.

Second, the team used pairwise correlation tests to better understand the association between the REMIQS school composite scores and the indicator scores for each state. This second round of testing differs from the first because it focused on the indicator scores rather than the state school ratings. These analyses shed light on how the differences in indicator-level associations may be driving the overall differences observed at the school ratings level.

Third, the research team computed the correlations between the percentage of students from historically resilient and marginalized backgrounds, the REMIQS framework and state systems. These analyses also relied on pairwise correlation tests. Though not causal, these analyses show the extent to which school composite ratings are related to the proportion of historically resilient and marginalized backgrounds served in schools.

Hypotheses
The research team did not expect to find high correlations between the REMIQS school composite scores and school ratings for several reasons. First, REMIQS controlled for past performance in all measures. Although some states included measures of progress in the Academic Progress indicator, these measures did not make up most of the states' school ratings.

Second, before making indicator-specific exclusions in the student universe of interest, the states included all students in their analysis. REMIQS, on the other hand, limited the student universe to those in historically resilient and marginalized groups.

Third, REMIQS school composite scores included postsecondary outcome measures. Because of limited data availability, workforce outcome measures were included in REMIQS school composite scores for Virginia but not for the other states. The combination of workforce and postsecondary outcome measures comprised 55 percent of the REMIQS school composite scores, with variation in measures used based on data availability. The states in this analysis did not include either postsecondary or workforce outcome measures in their ratings of schools.

The research team expected varying correlation levels between the REMIQS school composite scores and each of the states' indicator scores based on the metrics included in each state's school ratings, how similar the metrics were to the REMIQS metrics and the weight that was allotted to them. The team expected the correlation between the REMIQS composite score and each indicator score to differ based on how similar the indicators were to the measures and weight of the REMIQS school composite score metrics. For example, the research team anticipated that states that included Academic Progress measures with higher weights than Academic Achievement would have stronger correlations with REMIQS school composite

scores because controlling for prior performance is central to the REMIQS framework (Table 7).

The REMIQS framework was designed to measure the extent to which schools support success for historically resilient and marginalized students by using and highly weighting multiple measures that control for prior performance and by including students from these groups in the calculations. The researchers therefore expected that the association between the REMIQS composite score and the percentage of historically resilient and marginalized students in a given school would be weaker than the association between any state system's scores and the percentage of historically resilient and marginalized students.

Findings

The research team calculated coefficients based on pairwise correlation to understand the extent to which the school ratings for each state correlated with the REMIQS school composite scores. The analysis found a weak to moderate association between the REMIQS school composite scores and the state summative school ratings in Kentucky, Massachusetts and Texas, suggesting that the REMIQS model is a departure from the states' school rating methodologies (see Table 8) and may therefore offer insights into how such states may elect to expand and/or deepen their analyses to make equity more visible. These findings and the results comparing REMIQS school composite scores with indicator scores in Kentucky, Massachusetts and Texas are discussed in the following sections.

Kentucky

There was a positive but weak correlation (r = 0.25) between REMIQS school composite scores and Kentucky's school ratings (Table 8). This finding means that higher scores in one measure are associated with higher scores in the other; however, the association is not causal—that is, one measure does not predict a change in the other. The results were statistically significant to the 0.01 level, suggesting that the finding is not likely to be due to chance.

The weak correlation coefficient may be partially due to the differences between the measures included in the REMIQS school composite scores and those included in Kentucky's school ratings. Achievement on standardized tests makes up 45 percent of Kentucky's school ratings, whereas 75 percent of REMIQS school composite scores include measures of school contributions to postsecondary success (Table 7).

INDICATOR SCORES

To understand how specific indicator scores compare to the REMIQS school composite scores, researchers calculated correlations between the REMIQS school composite score and the three Kentucky indicator measures that were included in the state's school scores: mean proficiency,[41] transition rate[42] and high school graduation (Table 9).

The results indicate a moderate and positive association between the REMIQS school composite score and two of the indicator scores: mean proficiency (r = 0.33) and graduation (r = 0.21; Table 9). Both correlations were found to have a highly statistically significant association (p = .01), suggesting that the observed correlation is unlikely due to chance. A very weak positive and insignificant association was found between REMIQS school composite score and Kentucky's transition rate score (r = .02), indicating minimal association.

Separately testing the association between the REMIQS school composite score and each measure that was included in Kentucky's school rating system helps with understanding why there was a modest correlation between the REMIQS school composite score and Kentucky's school rating (Table 8). The association between transition rate (Kentucky's SQSS indicator measure) and the REMIQS school composite score was weak and accounted for about one third of Kentucky's school rating. This finding may be partially explained by the comparatively small overlap in students who are included in REMIQS's analyses and Kentucky's analyses. Kentucky's schools have, on average, low proportions of students from historically resilient and marginalized groups.

Table 8. Correlations Between REMIQS School Composite Scores and the State School Ratings for Kentucky, Massachusetts and Texas

	Kentucky School Rating	Massachusetts School Rating	Texas School Rating
REMIQS School Composite Scores	0.25***	0.51***	0.54***

***p < 0.01.

Table 9. Correlations Between Kentucky's Indicator Scores and REMIQS's School Composite Scores

	Academic Achievement (Mean Proficiency)[43]	School Quality or Student Success (Transition Rate)[44]	Graduation Rate (High School Graduation)
REMIQS School Composite Scores	0.33***	0.02***	0.21***

***p < 0.01.

HISTORICALLY RESILIENT AND MARGINALIZED GROUPS

The research team also conducted paired correlation tests to examine the correlation between the percentage of historically resilient and marginalized student groups in schools and the ratings or scores assigned to those schools by the states and REMIQS (Table 10).

The findings show negative correlations, meaning that a higher percentage of historically resilient and marginalized students in schools was associated with lower REMIQS school composite scores and lower Kentucky school ratings. This association should not be interpreted as evidence that higher percentages of resilient and historically marginalized students cause lower school ratings. Rather, it may reflect the impact of methodologies that measure the demographics of student populations rather than the impact of the school on student performance, particularly for historically resilient and marginalized students.

These findings were highly statistically significant (p < 0.01), meaning that the variables in the analyses had a consistent inverse association. The strength of the correlation differed, as expected. The correlation between the percentage of historically resilient and marginalized students and Kentucky school ratings was very strong (r = -0.74), whereas the

correlation with REMIQS composite scores was weak (r = -0.23). Thus, an important finding is that the percentage of historically resilient and marginalized students in a school was strongly associated with a low rating in the Kentucky system, whereas the association between the Kentucky school rating and the REMIQS school composite score was weak. This finding suggests that school ratings in Kentucky could be related to the percentage of historically resilient and marginalized students enrolled, whereas this relationship is weaker in the REMIQS framework.

Massachusetts

REMIQS school composite scores were moderately positively correlated (r = 0.51) with Massachusetts school ratings, as predicted. This finding is statistically significant at the 0.01 level. Schools in Massachusetts with high REMIQS composite scores were associated with high ratings from the state's system, and vice versa. A notable difference in the school-level scores is that Academic Achievement (Mean Proficiency), and growth scores made up 60 percent of the school ratings in the state's system, compared to 10 percent for similar measures in REMIQS school composite scores (Table 7).

INDICATOR SCORES

The research team calculated correlations between REMIQS school composite scores and the five indicator scores in the Massachusetts

Table 10. Correlations Between Percentages of School Population from Historically Resilient and Marginalized Student Groups and School Scores or Ratings for REMIQS and Kentucky

	School-level Percentages of Historically Resilient and Marginalized Student Groups
REMIQS School Composite Scores (Kentucky)	-0.23***
Kentucky School Ratings	-0.74***

***p < 0.01.

Note: The student groups included in this calculation are EL students, students in special education, students who were eligible for free and reduced-price lunch, Black students, Indigenous students, Hispanic or Latino students and students identified as multiracial.

Table 11. Correlations Between Massachusetts's Indicator Scores and REMIQS School Composite Scores

	Academic Achievement (Mean Proficiency)	Academic Progress (SGP)	Graduation Rate	School Quality or Student Success	
				Chronic Absenteeism	Advanced Coursework Completion
REMIQS School Composite Scores	0.46***	0.33***	0.52***	-0.53***	0.38***

***p < 0.01.

system to understand how specific indicator scores compared to the REMIQS school composite scores. The indicator scores included in the state's system were Academic Achievement (Mean Proficiency),[45] Academic Progress (SGP),[46] Graduation Rate and School Quality or Student Success (consisting of Chronic Absenteeism and Advanced Coursework Completion measures).

All indicator measures included in the Massachusetts school composite scores were positively correlated with the REMIQS composite scores, except Chronic Absenteeism (Table 11). This finding means that schools with high REMIQS composite scores also had relatively higher indicator scores for Mean Proficiency (r = 0.46), SGP (r = 0.33), Graduation Rate (r = 0.52), and Advanced Coursework Completion (r = 0.38). Schools with higher REMIQS scores had better attendance as measured by lower rates of chronic absenteeism (r = -0.53).

These indicator score analyses help with understanding the positive, moderately correlated and statistically significant association between Massachusetts school ratings and

REMIQS school composite scores (Table 8). The relationship between Academic Achievement indicator scores and the REMIQS composite scores were positive, moderately correlated (r = 0.46) and statistically significant to the .01 level. Further, the indicator most heavily weighted under Massachusetts's school ratings methodology was Academic Achievement (40 percent), whereas 75 percent of REMIQS school composite scores consisted of measures of school contributions to postsecondary success (Table 7).

HISTORICALLY RESILIENT AND MARGINALIZED GROUPS
Researchers also examined the relationship between REMIQS school composite scores and state school ratings in Massachusetts by testing the association between the percentages of historically resilient and marginalized student groups in the schools and school-level scores or ratings for Massachusetts and REMIQS.

The research team expected to find a strong negative association between the Massachusetts school ratings and the percentages of historically resilient and marginalized students in schools

Table 12. Correlations Between Percentages of School Population From Historically Resilient and Marginalized Student Groups and School Scores or Ratings for REMIQS and Massachusetts

	School-level Percentages of Historically Resilient and Marginalized Student Groups
REMIQS School Composite Scores (Massachusetts)	-0.53***
Massachusetts School Ratings	-0.69***

***p < 0.01.

Note: The student groups included in this measure are EL students, students in special education, students who were eligible for free and reduced-price lunch, Black students, Indigenous students, Hispanic students and students identified as multiracial.

because Academic Achievement, as measured by standardized assessments, made up 40 percent of the school ratings (Table 7).

Lower Massachusetts school ratings (Table 12) were also associated with higher percentages of historically resilient and marginalized students. Both tests were highly statistically significant (p < 0.01). As anticipated, the correlation was weaker for the REMIQS school composite scores (r = -0.53) than for the Massachusetts school ratings (r = -0.69).

Texas

There was a moderate correlation (r = 0.54) between REMIQS school composite scores and Texas school ratings (Table 8). This finding, statistically significant to the 0.01 level, suggests that schools in Texas with high REMIQS composite scores are associated with having high ratings in the Texas school accountability system, and vice versa.

INDICATOR SCORES

To understand how specific indicator scores in Texas compared to the REMIQS school composite scores, the research team calculated correlations between the REMIQS school composite scores and the three indicator scores in the state's school ratings: Academic Achievement (Academic Performance), Academic Progress (School Progress, which is the score for either Academic Growth or Relative Performance, whichever is higher) and Closing the Gaps. Each school in Texas received either an Academic Achievement or Academic Progress score, whichever was

higher. The third measure, Closing the Gaps, is not associated with an ESSA indicator but was included in Texas's school ratings.

All measures included in Texas's school ratings were positively correlated with REMIQS composite scores and statistically significant at the < 0.01 level (Table 13). There were moderate correlations between REMIQS composite scores and Academic Performance (r = 0.59) and Closing the Gaps (r = 0.59), but a weak correlation between REMIQS and Academic Progress (r = 0.30).

Because the REMIQS framework used a value-added approach, Academic Progress (either Academic Growth or Relative Performance in the Texas system) was expected to be more highly correlated with REMIQS than Academic Achievement (Academic Performance in the Texas system). Instead, Academic Achievement was more highly correlated (r = 0.59) than Academic Progress (r = 0.30). These results suggest that Academic Performance is a stronger factor than Academic Progress in explaining the moderate, positive association between Texas' school ratings and REMIQS school composite scores. This may be the case because Academic Achievement in Texas includes not only achievement on standardized assessments but also other measures such as high school graduation and college and career readiness, measures that are also incorporated into the REMIQS school composite scores.

Table 13. Coefficients Between Texas Indicator Scores and REMIQS's School Composite Scores

	Academic Achievement (Mean Proficiency)[47]	Academic Progress (School Progress as an Alternative to Academic Performance)[48]	Non-ESSA Indicator (Closing the Gaps)[49]
REMIQS School Composite Scores	0.59***	0.30***	0.51***

***p < 0.01.

Table 14. Correlations Between Percentages of School Populations from Historically Resilient and Marginalized Student Groups and School Scores or Ratings for REMIQS and Texas

	School-level Percentages of Historically Resilient and Marginalized Student Groups[50]
REMIQS School Composite Scores (Texas)	-0.41***
Texas School Ratings	-0.38***

***p < 0.01.

HISTORICALLY RESILIENT AND MARGINALIZED GROUPS

The research team conducted pairwise correlation tests to explore the association between a school's percentage of historically resilient and marginalized student groups and the school's REMIQS score and Texas school rating (Table 14). Similar results were expected because 30 percent of the Texas school rating was from the Closing the Gaps indicator score, which was similar to the REMIQS measures (Table 7). Findings confirmed moderate negative correlations between the percentage of historically resilient and marginalized students in a school and both the REMIQS score (r = -0.41) and the Texas school rating (r = -0.38). In both instances, a higher percentage of historically resilient and marginalized students in a school was associated with a lower score or school rating. Nonetheless, these models may not account for key contextual, systemic and environmental factors that can influence how well schools serve these students and promote their academic performance (Reardon et al., 2019).

DISCUSSION

Methodological Comparison

The REMIQS framework addresses concerns regarding the use of standardized test scores, which are strongly correlated with family background, as the dominant measure for state accountability systems (DeLuca et al., 2016). Ultimately, the REMIQS framework may provide states with ideas for measuring schools' impact on historically resilient and marginalized student outcomes in ways that depart from what may be an overreliance on test scores.

Because ESSA affords states broad latitude in developing their accountability systems, there were differences in the student groups, metrics and weights that states used in calculating their school ratings. The research team compared the REMIQS school composite scores to these differing state school ratings in order to contrast the ways they measure school quality and to explore the equity implications of those differences.

student population, making it harder to ascertain outcomes for historically marginalized and resilient student groups. REMIQS, on the other hand, included only historically resilient and marginalized student groups in its calculations, which centered the school ratings on how schools served those groups. States should explore the flexibility ESSA affords them to both assess how schools serve all students and foreground historically resilient and marginalized students' specific needs and experiences.

Second, SQSS was the most complex indicator to analyze because ESSA guidelines afford states considerable flexibility for this indicator. The result is variation in the numbers and types of metrics included in SQSS and in the weights allocated to this indicator in the states' school ratings. Further, four REMIQS SQSS metrics accounted for 75 percent of REMIQS's school composite scores, yet most of that weight (55 percent out of 75 percent) was attributed to college enrollment and persistence measures,

> Ultimately, the REMIQS framework may provide states with ideas for measuring schools' impact on historically resilient and marginalized student outcomes in ways that depart from what may be an overreliance on test scores.

Findings from this investigation uncovered four methodological decisions that may limit the ability of states to measure the extent to which schools serve historically resilient and marginalized students.

First, states included the whole student population in their calculations of student outcomes and school ratings. Given that all students are included in state calculations, the schools' impacts on historically resilient and marginalized students may be masked. For example, the performance of advantaged students can skew the mean when states average outcome measures for the whole

which no state included in its school ratings. This difference may help explain the relatively low correlation rates between REMIQS and the state systems. Also, the REMIQS school composite rating formulas varied slightly across states due to data availability, which may have influenced the findings.

Third, states in this study applied high weights (30–70 percent) to the Academic Achievement indicator (i.e., proficiency on state-mandated standardized tests) in calculating school ratings. Although there is variation in the way that states defined the metrics used and the weight attributed to the Academic Achievement

indicator, standardized test scores are highly correlated with family background, meaning that advantaged students have higher test scores on average than students from historically resilient and marginalized groups (Hanushek et al., 2019; Reardon et al., 2014). Consequently, schools with high percentages of historically resilient and marginalized students tend to have lower average academic achievement scores than schools that serve predominately advantaged students.

Arizona, Massachusetts and Texas each included year-over-year Academic Progress indicator measures with varying units of analysis. Texas and Massachusetts measured Academic Progress at the individual level, restricting their analyses to students with two years of consecutive data. This decision may have underrepresented highly mobile students who tend to be members of historically resilient and marginalized groups.[51] Arizona measured

> This approach set the REMIQS framework apart from the state approaches because REMIQS more precisely isolated schools' impact on students rather than capturing the effects of student or school demographics.

On the other hand, the REMIQS framework's Academic Achievement indicator controlled for prior performance, and the REMIQS school composite score applied only 10 percent weight to student- and school-level performance. The REMIQS framework thus de-emphasized measures that are correlated with student demographic characteristics and instead emphasized schools' impact on historically resilient and marginalized students.

Fourth, only three states in the analysis—Arizona, Massachusetts and Texas—included Academic Progress measures in their state school ratings. Although these measures use the same assessment data as the Academic Achievement indicator, Academic Progress captures change over time, which can translate into the effect of the school on students. Kentucky and Virginia did not include Academic Progress measures in their state school ratings as of 2017.

Instead of including explicit Academic Progress measures, the REMIQS framework controlled for prior performance in all measures and for student- and school-level demographics. This approach set the REMIQS framework apart from the state approaches because REMIQS more precisely isolated schools' impact on students rather than capturing the effects of student or school demographics.

Academic Progress at the student group level, which may have been helpful in understanding each school's impact on specific student groups. Although the members of these groups change each year, mitigating concerns about mobile students' underrepresentation, failing to track the same students year-over-year may have produced misleading findings on how schools served individual students. States may want to make decisions about including or excluding mobile students based on local context and may want to consider the impact of these decisions on school ratings.

School Quality Measures Comparison

Due to observed differences across state accountability systems and between these systems and the REMIQS framework, the analyses were expected to reveal:

» Variation in the extent to which state accountability systems' summative ratings and indicator measures correlate with school REMIQS scores

» Weaker correlations of states' indicator scores and REMIQS school composite scores with the school-level percentages of historically resilient and marginalized students than with the state school ratings

The correlation between state school ratings and REMIQS school composite scores was weak for Kentucky and moderate for Massachusetts and Texas, reflecting the variation in methodologies used. The REMIQS indicator measures controlled for prior student performance to emphasize the impact of schools and minimize the effect of student demographics. Hence, researchers expected to observe stronger correlations between REMIQS school composite scores and Academic Progress indicator scores.

The analysis found, however, that Academic Achievement scores were more highly correlated with REMIQS school composite scores than Academic Progress was. Although this finding may be explained by variation in measures and weights of Academic Achievement and Academic Progress, this observation warrants further investigation into states' Academic Progress methodologies, their influence on state school ratings and other contributing factors.

As expected, there were differences in correlation levels between the school-level percentages of historically resilient and marginalized student populations and the states' school ratings. This correlation was stronger in Kentucky and Massachusetts, states that heavily weighted student achievement on standardized assessments. Conversely, this correlation was relatively weak in Texas, where school ratings substituted Academic Progress for Academic Achievement when Academic Progress had a higher score. REMIQS school composite scores had a weaker correlation with the percentages of historically resilient and marginalized students than did any of the three states' ratings. Although this finding suggests that the REMIQS framework may be more effective than the state systems at isolating school effects on student performance, further investigation is needed.

The relatively close correlations between REMIQS's composite scores and Texas school ratings suggest that REMIQS and the Texas state system include similar measures. Namely, Texas explicitly accounts for the performance of individual student groups as part of its state summative ratings through its Closing the Gaps metric. This approach may offer other states examples for measuring how schools serve historically resilient and marginalized students.

Policymakers can use these findings to understand the differences between their state accountability methodology and that of the REMIQS framework and perhaps revise their measures and weights to more accurately reflect school-level quality for historically resilient and marginalized students.

FUTURE RESEARCH

Several research areas could build on these analyses to deepen the field's understanding of the equity implications of how states measure school quality.

The analyses described in this paper used data from 2018–19 and methodologies that were established in 2017. Future research could replicate these analyses with more recent information to examine how methodological changes may result in different outcomes.

Relatedly, this study centered on five states. Replicating this methodological approach with data from other states would strengthen researchers' ability to identify patterns and expand the field's understanding of the relationship between state accountability systems and the REMIQS framework.

The REMIQS framework included data only on historically resilient and marginalized students, whereas state systems also included data on advantaged students. Findings from future studies that include comparable data across states and REMIQS may help validate or further clarify the findings from this study.

Future studies could also help answer questions that emerged from the findings in this analysis. For example, why was there a stronger correlation between REMIQS and Academic Achievement than between REMIQS and Academic Progress in Massachusetts and Texas?

States are increasingly focusing on improving their students' CCR to prepare the workforce for the estimated 70 percent of jobs requiring at least some postsecondary education or training beyond a high school degree (Carnevale & Cheah, 2018). Future research could investigate how states measure CCR (as a subset of the SQSS indicator) and compare the outcomes of those measures with the REMIQS CCR measures. Findings from such a comparison could help states identify approaches to measure a school's impact on CCR for historically resilient and marginalized students.

Finally, in the aftermath of the COVID-19 pandemic, researchers and policymakers have issued calls to redesign accountability systems to better promote high standards and equity (Finn, 2022). Future research could explore how these changes may or may not align with REMIQS and explore the association between changes to accountability frameworks and student outcomes.

ENDNOTES

1. The REMIQS project defines traditional settings as nonselective admissions public schools. This definition excludes magnet and special admissions public schools, which ensures REMIQS focuses on the schools that most students in a jurisdiction have access to attend.

2. KnowledgeWorks staff: Eric Toshalis. REMIQS Advisors: Juan Carrillo, Arizona State University; Adai Tefera, University of Arizona; Ivory Toldson, Howard University; Angela Valenzuela, University of Texas at Austin. REMIQS Stakeholder Committee members: Eric Brooks, Yuma Union High School District; Asha Dane'el, Consultant; Charles Davis, Jr., Evolve502; Jakira Rogers, Massachusetts Advocates for Children; Keesa McCoy, Roosevelt Institute; Sabine "BiNi" Coleman, 212 Catalysts; Leah Dozier Walker, Waterford.org; Andrew Daire, Virginia Commonwealth University.

3. National studies consistently show Black and Hispanic or Latino students scoring lower on standardized math and English language arts (ELA) assessments than non-Hispanic White students. Likewise, students from low-income families, on average, score lower on these assessments than students from higher income families. Hanushek et al. (2019); Reardon et al. (2014).

4. This paper uses the term "English Learners" (ELs or EL students) because it is a statutorily defined term under ESSA and in each state's accountability system. The authors prefer the term "emerging bilingual" to foreground an asset-based framing of those students whose first language is not English. To become multilingual is to think within and access meaning from more than a single range of vocabularies, symbols, linguistic structures, cultures, and histories, the results of which can greatly enhance a student's understanding in and contributions to learning environments. The terms "English Learner" and "English language learner" are the typical categories used by governments and schools and refer to the same population that can be referred to as "emerging bilingual."

5. The Department of Education recognizes the following racial/ethnic groups: American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, two or more races, and White (https://nces.ed.gov/ipeds/report-your-data/race-ethnicity-definitions).

6. For more detailed information, see the REMIQS Quantitative Filtering Technical Report: https://knowledgeworks.org/wp-content/uploads/2023/11/remiqs-quantitative-filtering-technical-report-2022.pdf. The REMIQS project selected these states based on their demographic and geographic diversity.

7. Based on data availability, the REMIQS statistical models defined historically resilient and marginalized students as those who were Black or African American, Hispanic or Latino, Indigenous or multiracial; were from low-income families; qualified for special education services; or were designated as EL students.

8. The 2018–19 accountability system data is used in this analysis because it was the most recent data included in the development of the REMIQS framework and methodology and precedes the methodological and data disruptions caused by the COVID-19 pandemic.

9. In Texas, if the Academic Progress indicator score is greater than the Academic Achievement indicator score, the Academic Achievement indicator score was not included in the school rating (Table 7).

10. Virginia did not publicly report the weight for academic achievement or any other domain.

11. Course-based exams are statewide assessments in certain subject areas, such as Algebra, Geometry or Biology.

12. The two exceptions were students without a grade eight assessment and students who were grade nine repeaters.

13. REMIQS did not incorporate academic progress as part of the framework.

14. Student progress in Texas may be substituted with relative performance score if the relative performance score is higher than the progress score. See discussion on how metrics are aggregated to form accountability score.

15. Virginia did not publicly report the weight for academic progress or any other domain.

16. First grade nine year cohort is the number of students entering grade nine for the first time plus any students entering at a later point during the grade nine year or at any point up until the expected year of graduation, minus any students exiting the state public school system with a validated reason during the same time.

17. Massachusetts' extended engagement rate equaled the sum of the percentage of students graduating in five years plus the percentage of students still enrolled after five years. Massachusetts Consolidated State Plan Under the Every Student Succeeds Act (2017).

18. Massachusetts did not publicly define its annual dropout rate calculation. Instead, it noted, "The graduation rate of a high school is certainly a key indicator of success. However, in a district and school accountability system that makes annual determinations, it is often difficult to make large gains in a graduation rate calculation in one year because much of the rate has been determined in grades nine to 11. The number of high school dropouts on an annual basis is a significant component of the graduation rate calculation. The inclusion of the annual dropout rate in a high school accountability determination allows for a more actionable indicator for high schools on an annual basis." Massachusetts Consolidated State Plan Under the Every Student Succeeds Act (2017).

19. Graduation rates were part of the Student Achievement indicator score, which overall accounted for 70 percent of school accountability ratings. This indicator could be replaced with the Student Progress indicator, which does not account for graduation, if the Student Achievement score was lower than Student Progress.

20. Virginia did not publicly report the weight for high school graduation or any other domain.

21. https://www2.ed.gov/programs/sfgp/eseatitleiiiresourceaccountelsguide.pdf, p. 4.

22. Kentucky and Virginia did not publicly report the weight for ELP.

23. REMIQS did not measure or account for ELP in the school composite score because data was unavailable.

24. The percentage of current EL students demonstrating growth or scoring Advanced on ELP assessment is included in the Academic Performance indicator score.

25. In fall 2020, EL students represented 20.1 percent of Texas's total public school enrollment. National Center for Education Statistics (2023).

26. Texas's college, career, and military readiness rating is part of the Student Achievement indicator score, which overall accounts for 70 percent of school accountability ratings. This indicator may be replaced by the Student Progress indicator, which does not account for SQSS, if the Student Achievement indicator score is lower.

27. Virginia did not publicly report the weight for SQSS or any other domain.

28. Comprehensive Support 1 (CSI) is the designation for the bottom five percent of performance of Title 1 schools; Targeted Support designations are for schools with student groups that are underperforming; Comprehensive Support 2 (CSII) is the designation for high schools with graduation rates below 67 percent.

29. The top scoring school in each stratum in each state was identified as a "REMIQS school."

30. For example, if a school's proficiency score was 80 percent, the school would receive 80 percent of the 30 points available in the Academic Achievement indicator (0.80 × 30).

31. A: 100–83.83 percent; B: 83.82–70.02 percent; C: 70.1–56.21 percent; D: 56.20–42.40 percent; F: 42.39–0 percent

32. Very low = 1, low = 2, medium = 3, high = 4, very high = 5

33. Declined = 0, no change = 1, improved below target = 2, met target = 3, exceeded target = 4

34. Eligible subgroups included race/ethnicity, economically disadvantaged students, current or former students with disabilities, current or monitored EL students, and continuous or noncontinuous enrolled.

35. See Tables 1–5 for explanation behind why certain cells are not applicable for some frameworks.

36. As of 2023, Texas' Closing the Gaps indicator evaluated the disaggregated performance of student groups, including all students, seven racial/ethnic groups, the two lowest performing racial/ethnic groups from the preceding school year, high-focus students, continuously enrolled students, and former special education students. The high-focus student group included students identified as economically disadvantaged, EL, special education or highly mobile (in foster care, experiencing homelessness, migrant). The continuously enrolled student group included students enrolled in the campus during the fall snapshot for the given school year and in the same district each of the three preceding school years. Four groups' outcomes contributed to a school's Closing the Gaps rating: all students, the two lowest performing racial/ethnic groups from the preceding school year, and high-focus students. The Closing the Gaps indicator, which was weighted at 30 percent of the school ratings, included four metrics: Academic Achievement (50 percent of indicator weight), the higher score of Academic Growth and Graduation (10 percent of indicator weight), ELP (10 percent of indicator weight), and SQSS (30 percent of indicator weight). Texas Education Agency (2023).

37. Fifteen percent of Kentucky star ratings came from an alternative academic indicator (science, social studies or writing).

38. Student Performance, Graduation Rate, and School Quality/Student Success in Texas were part of the Academic Achievement indicator score, which overall accounted for 70 percent of school ratings. This indicator could be replaced with the Academic Progress indicator if the Academic Achievement indicator score was lower.

39. States determine the minimum number of students in each subgroup to be considered "sufficient."

40. See the REMIQS Quantitative Filtering Technical Report: https://knowledgeworks.org/wp-content/uploads/2023/11/remiqs-quantitative-filtering-technical-report-2022.pdf

41. In Kentucky, mean proficiency combines the math and ELA.

42. Kentucky defined transition rate as follows: "The attainment of the necessary knowledge, skills, and dispositions to successfully transition to the next level of his or her education career." The measure is a combination of college and career readiness, science, social studies, writing assessments, and ELP. Kentucky Department of Education (2023).

43. WestEd calculated mean proficiency of ELA and math proficiency rates to produce a single measure for the Academic Performance indicator to compare with the REMIQS school composite score. Kentucky did not calculate or report on mean proficiency.

44. Kentucky defined transition rate as follows: "The attainment of the necessary knowledge, skills, and dispositions to successfully transition to the next level of his or her education career." The measure is a combination of college and career readiness, science, social studies, writing assessments, and ELP. Kentucky Department of Education (2023).

45. Mean proficiency in Massachusetts combined the average scaled scores for ELA and math.

46. Massachusetts defined SGP as "a measure of the degree to which a student's achievement has changed from the prior year(s) to the current year, in comparison to other students in the same grade who performed similarly in the past." Massachusetts Department of Elementary and Secondary Education (2021).

47. Academic Performance includes student performance on standardized assessments, graduation rates, and school quality/student. This indicator may be replaced with the Student Progress indicator if the Student Achievement indicator score is lower.

48. School Progress is the best score between relative academic performance and student growth. The former refers to performance on standardized assessments relative to schools with similar proportions of economically disadvantaged students.

49. Four groups' outcomes contributed to a school's Closing the Gaps rating in Texas: all students, the two lowest performing racial/ethnic groups from the preceding school year, and high-focus students. The Closing the Gaps indicator, weighted at 30 percent of the school ratings, included four metrics: Academic Achievement (50 percent of indicator weight), the higher score of Academic Growth and Graduation (10 percent of indicator weight), ELP (10 percent of indicator weight), and SQSS (30 percent of indicator weight). Texas Education Agency (2023).

50. The student groups included in this calculation are EL students, students in special education, students who were eligible for free and reduced-price lunch, Black students, Indigenous students, Hispanic students, and students identified as multiracial.

51. Student mobility refers to any time a student changes schools during a school year for reasons other than grade promotion. Prior research has linked student mobility with lower school engagement, academic performance, and increased risk of school noncompletion. Highly mobile students are often concentrated in schools with large populations of historically resilient and marginalized students. For more information, see Rumberger (2015).

REFERENCES

Anderson, T., Blount, D.C., Lindsay C.A., Blom, E., Gebrekristos, S., & Alba, F. (2019). Robust and equitable measures to identify quality schools [Concept brief]. https://www.urban.org/research/publication/robust-and-equitable-measures-identify-quality-schools

Arizona Department of Education. (2019). A–F letter grade accountability system: Traditional schools business rules 9–12 model. https://www.azed.gov/sites/default/files/2022/02/2022%209-12%20Traditional%20Schools%20Business%20Rules%20Draft.pdf

Barrows, S., Henderson, M., Peterson, P. E., & West, M. R. (2016). Relative performance information and perceptions of public service quality: Evidence from American school districts. Journal of Public Administration Research and Theory, 26(3), 571–583. https://doi.org/10.1093/jopart/muw028

Blom, E., Anderson, T., & Lindsay, C. (2020). Identifying high-performing schools for historically underserved students: Exploring a multistate model. Urban Institute.

Carnevale, A. P., & Cheah, B. (2018). Five rules of the college and career game. Georgetown University Center on Education and the Workforce. https://eric.ed.gov/?id=ED590613#:~:text=The%20five%20rules%20presented%20in,is%20important%2C%20it%20does%20not

Chingos, M. (2017, January 18). Why the proficiency-versus-growth debate matters for assessing school performance. Urban Wire. https://www.urban.org/urban-wire/why-proficiency-versus-growth-debate-matters-assessing-school-performance

Chingos, M. M., Henderson, M., & West, M. R. (2012). Citizen perceptions of government service quality: Evidence from public schools. Quarterly Journal of Political Science, 7(4), 411–445. http://dx.doi.org/10.1561/100.00011071

Kentucky Department of Education (2017). Commonwealth of Kentucky revised consolidated state plan under the Every Student Succeeds Act. https://education.ky.gov/comm/Documents/Kentucky%20Consolidated%20State%20Plan.pdf

Massachusetts Department of Elementary and Secondary Education. (2017). Commonwealth of Massachusetts revised consolidated state plan under the Every Student Succeeds Act. https://www.doe.mass.edu/federalgrants/essa/stateplan/

Cuban, L. (2013). Inside the black box of classroom practice: Change without reform in American education. Harvard Education Press.

DeLuca, S., Clampet-Lundquist, S., & Edin, K. (2016). Coming of age in the other America. Russell Sage Foundation.

Duncan, G. J., & Murnane, R. J. (Eds.). (2011). Whither opportunity?: Rising inequality, schools, and children's life chances. Russell Sage Foundation.

Durodoye, R., Jr., Lacireno-Paquet, N., Tanner, S., & Vislosky, E. (2021). REMIQS quantitative filtering technical report. WestEd.

U.S. Department of Education. Every Student Succeeds Act, 20 U.S.C. (2015). Pub. L, 114–95. https://www.ed.gov/essa?src=rn

Facella, M. A., Rampino, K. M., & Shea, E. K. (2005). Effective teaching strategies for English language learners. Bilingual Research Journal, 29(1), 209–221. https://doi.org/10.1080/15235882.2005.10162832

Feng, L., Figlio, D., & Sass, T. (2018). School accountability and teacher mobility. Journal of Urban Economics, 103, 1–17.

Finn, C. E., Jr. (2022). School accountability: Yesterday, today, and tomorrow. Phi Delta Kappan, 104(3), 12–17.

Gándara, P., & Rumberger, R. W. (2007). Resource needs for California's English learners. https://cepa.stanford.edu/content/resource-needs-california%E2%80%99s-english-learners

Gándara, P., & Rumberger, R. W. (2009). Immigration, language, and education: How does language policy structure opportunity? Teachers College Record, 111(3), 750–782.

Goudeau, S., & Croizet, J. C. (2017). Hidden advantages and disadvantages of social class: How classroom settings reproduce social inequality by staging unfair comparison. Psychological Science, 28(2), 162–170.

Hanushek, E. (2014). Achievement gap. In D. C. Phillips (Ed.), Encyclopedia of educational theory and philosophy (pp. 4–7). SAGE.

Hanushek, E. A., Peterson, P. E., Talpey, L. M., & Woessmann, L. (2019). The achievement gap fails to close. Education Next, 19(3), 8–17.

Ho, A. (2016). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. Educational Researcher, 37(6), 351–260.

Howe, K. R., & Murray, K. (2015). Why school report cards merit a failing grade. National Education Policy Center. https://nepc.colorado.edu/publication/why-school-report-cards-fail

Institute of Education Sciences Statewide Longitudinal Data Systems Grant Program. (2012). Growth models: Issues and advice from the states. https://files.eric.ed.gov/fulltext/ED551302.pdf

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

Kentucky Department of Education. (2023). Postsecondary readiness. https://education.ky.gov/AA/Acct/Pages/Postsecondary-Readiness.aspx

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. Educational Researcher, 35(7), 3–12.

Massachusetts Consolidated State Plan Under the Every Student Succeeds Act (ESSA). (2017). U.S. Department of Education. https://www2.ed.gov/admins/lead/account/stateplan17/maconsolidatedstateplan.pdf

Massachusetts Department of Elementary and Secondary Education. (2021). What are Student Growth Percentiles (SGPs), and how are they usually calculated? https://www.doe.mass.edu/mcas/2021/results/sgp.docx

National Center for Education Statistics. (n.d.). Definitions for new race and ethnicity categories. Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/ipeds/report-your-data/race-ethnicity-definitions

National Center for Education Statistics. (2023). English Learners in public schools. Condition of education. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cgf

Noonan, J., & Schneider, J. (2022). Beyond "good" and "bad": Disrupting narratives about school quality. Phi Delta Kappan, 104(3), 6–11.

Owens, A. (2018). Income segregation between school districts and inequality in students' achievement. Sociology of Education, 91(1), 1–27.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), Whither opportunity? Rising inequality, schools, and children's life chances (pp. 91–116). Russell Sage Foundation. https://www.russellsage.org/news/widening-academic-achievement-gap-between-rich-and-poor

Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The geography of racial/ethnic test score gaps. American Journal of Sociology, 124(4), 1164–1221.

Reardon, S. F., Robinson-Cimpian, J. P., & Weathers, E. S. (2014). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In Handbook of research in education finance and policy (pp. 491–509). Routledge.

Scharrón-del Rio, M., & Aja, A.A. (2015). The case for Latinx: why intersectionality is not a choice. Latino Rebels. https://www.latinorebels.com/2015/12/05/the-case-for-latinx-why-intersectionality-is-not-a-choice/

Rose, B. A. (2013). Examining variation in effects of student mobility using cross-classified, multiple membership modeling. Society for Research on Educational Effectiveness.

Rothstein, R. (2004). Class and schools: Using social, economic, and educational reform to close the Black–White achievement gap. Economic Policy Institute.

Rumberger, R. W. (2015). Student mobility: Causes, consequences, and solutions. National Education Policy Center. https://files.eric.ed.gov/fulltext/ED574695.pdf

Schneider, J. (2014). From the ivory tower to the schoolhouse: How scholarship becomes common knowledge in education. Harvard Education Press.

Schneider, J. (2017). Beyond test scores: A better way to measure school quality. Harvard University Press.

Schneider, J., & Gottlieb, D. (2021). In praise of ordinary measures: The present limits and future possibilities of educational accountability. Educational Theory, 71(4), 455–473.

Schneider, J., & Hutt, E. (2014). Making the grade: A history of the A–F marking scheme. Journal of Curriculum Studies, 46(2), 201–224.

Schneider, J., Jacobsen, R., White, R., & Gehlbach, H. (2017). Building a better measure of school quality. Phi Delta Kappan, 98(7), 43–48.

Schwartz, A. E., Stiefel, L., & Cordes, S. A. (2015). Moving matters: The causal effect of moving schools on student performance (Working Paper #01-15). Institute for Education and Social Policy. https://files.eric.ed.gov/fulltext/ED556782.pdf

Seastrom, M. (2017). Best practices for determining subgroup size in accountability systems while protecting personally identifiable student information (IES 2017-147). U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/pubs2017/2017147.pdf

Texas Education Agency. (2023). 2023 accountability manual. https://tea.texas.gov/texas-schools/accountability/academic-accountability/performance-reporting/2023-accountability-manual

U.S. Department of Education. (2017). Every Student Succeeds Act high school graduation rate non-regulatory guidance. https://www2.ed.gov/policy/elsec/leg/essa/essagradrateguidance.pdf

Valdés, G. (2005). Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized? The Modern Language Journal, 89(3), 410–426. https://doi.org/10.1111/j.1540-4781.2005.00314.x

WestEd. (n.d.). Striving to achieve equity is integral to WestEd's mission. https://www.wested.org/equity/

Xu, D., Solanki, S., & Fink, J. (2021). College acceleration for all? Mapping racial gaps in advanced placement and dual enrollment participation. American Educational Research Journal, 58(5), 954–992.

York, T. T., Gibson, C., & Rankin, S. (2019). Defining and measuring academic success. Practical Assessment, Research, and Evaluation, 20, Article 5. https://doi.org/10.7275/hz5x-tx03

GLOSSARY

Accountability system: This term refers to the policies and practices a state uses to measure the performance of student groups, schools, districts and the state's education system as a whole.

Advantaged students: This paper uses the term advantaged students to reference students, particularly those who are White and affluent, whose backgrounds may positively impact their standardized testing and academic performance. Advantaged students tend to have greater access to high-quality schools, experienced teachers, tutoring, enrichment programs, learning materials and technology (Rothstein, 2004). Additionally, their families may have the financial means to provide additional educational opportunities outside of school (Goudeau & Croizet, 2017). Because of these privileges, advantaged students often start school with a head start compared to other students, including historically resilient and marginalized students (Owens, 2018).

Equity: This report views equity as the attainment of comparably positive outcomes for all groups within or served by any complex system. Working toward equity is an ongoing process, implementing policies, practices and procedures that remove systemic barriers and provide the support needed to ensure everyone's complete and successful participation in the system. In schools, these policies, practices and procedures involve such approaches as providing appropriate and differentiated student support, setting high expectations for all students, supporting student agency in learning and building on community strengths to create experientially relevant and culturally integrated classrooms. Equity exists when race, ethnicity, language, religion, gender identity, sexual orientation, age, national origin, physical or cognitive ability, socioeconomic status and other such characteristics are not predictors of outcomes for any group or the individuals in them (WestEd, n.d.).

Every Student Succeeds Act (ESSA): Enacted in 2015, the Every Student Succeeds Acts (ESSA) reauthorized the Elementary and Secondary Education Act (ESEA) and outlined the statutory requirements for statewide accountability systems (Every Student Succeeds Act, 2015).

Every Student Succeeds Act Indicators: ESSA requires each state to establish a multimeasure accountability system that assesses school functioning and student performance using indicators in five areas: Academic Achievement, Academic Progress, Graduation Rate, Progress in Achieving English Language Proficiency (also referred to as English Language Proficiency or ELP) and School Quality or Student Success (SQSS).

Framework: The set of business rules and logic that aggregate student performance across metrics and indicators into school scores, such as the Robust and Equitable Measures for Inspiring Quality Schools (REMIQS) school composite score and states' school ratings.

Historically resilient and marginalized students: The REMIQS framework's quantitative model for assessing school quality defined historically resilient and marginalized students as (a) students of color who identify as Black or African American, Hispanic or Latino or belonging to two or more races; (b) students with disabilities; (c) economically disadvantaged students; and (d) students who are English language learners. These student groups are included in the analysis based on available data. The authors of this paper also acknowledge that America's historical, social and political conditions pose systemic barriers for many other students based on racial, ethnic, socioeconomic, indigenous, linguistic, religious, (dis)ability, immigration status, gender expression and sexual orientation differences. This framing foregrounds the tremendous achievements and contributions of members of such groups despite systems that may undermine them (Duncan & Murnane, 2011; Ladson-Billings, 2006).

While these analyses use the official ESSA designations for these student groups, the authors also acknowledge the deficit framing of these designations and the importance of asset-based alternatives. For instance, it is preferable to describe students whose first language is not English as "emerging bilingual." Becoming multilingual means thinking within and accessing meaning from more than a single range of vocabularies, symbols, linguistic structures, cultures and histories, enhancing these students' understanding and contributions to learning environments (Facella et al., 2005; Gándara & Rumberger, 2006, 2009; Valdés, 2005). Similarly, "Latinx" is a gender-neutral alternative to Latino/a and is intended to capture masculine, feminine and nonbinary identities. It is also a pan-ethnic term for co-ethnics from Latin America and individuals with Spanish-speaking cultural ties (Scharron-del Rio & Aja, 2015).

REMIQS (pronounced "re-mix"): The Robust and Equitable Measures to Inspire Quality Schools (REMIQS) project was designed to identify and understand the practices of high schools in traditional settings that demonstrate strong outcomes for historically resilient and marginalized students.

REMIQS methodology: The analyses in this paper drew on the REMIQS statistical model that used individual-level student data from five states (Arizona, Kentucky, Massachusetts, Texas and Virginia) to identify schools that excel at supporting historically resilient and marginalized students to succeed in school and after high school (Durodoye et al., 2021). Student data came from public high schools with no selective admissions criteria that had at least 100 or more grade nine students, and at least 25 percent of students identified as Black or African American, Hispanic or Latino, Indigenous or multiracial; were from low-income families; qualified for special education services; or were English Learners. REMIQS included data for students from historically resilient and marginalized groups. Inputs to the REMIQS model varied by state based on data availability. The REMIQS hierarchical mixed models nested students in schools within districts. Separate models estimated each of the following outcome variables of interest:

» **Academic outcomes**
These are variables that the REMIQS statistical model used to estimate the impact of high school on student outcomes, including attendance, eighth grade assessment scores, high school graduation, high school assessment scores and advanced coursework.

» **Postsecondary outcomes**
These are variables that the REMIQS statistical model used to estimate the impact of high school on student outcomes, including enrollment in postsecondary education institutions and college persistence or graduation. None of the state accountability systems described in this analysis included postsecondary outcomes.

» **Workforce outcomes**
The REMIQS statistical model included post–high school wages earned for Virginia in the REMIQS statistical model to estimate the impact of high school on student outcomes. None of the state accountability systems described in this analysis included workforce outcomes.

Student group: Demographic categories including race/ethnicity (American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino of any race, Native Hawaiian or Other Pacific Islander, two or more races, White), economically disadvantaged students, English Learners and students with disabilities. (See also: Historically resilient and marginalized students.)

Summative school rating: This term refers to the rating assigned to each school, also referred to as the REMIQS school composite score or the state school rating.

APPENDIX

Table A1. Subgroups Reported On and/or Accounted for in Overall Ratings, by Methodology

Subgroup	REMIQS	Arizona	Kentucky	Massachusetts	Texas	Virginia
Economically Disadvantaged	●	●	●	●	●	●
English Learner (EL)	●	●	●	●	●	●
Former EL				●	●	
Non-EL		●				
Special Education	●	●	●	●	●	●
Former Special Education					●	
American Indian or Alaska Native			●	●	●	
Asian		●	●	●	●	●
Black or African American	●	●	●	●	●	●
Hispanic or Latino	●	●	●	●	●	●
Two or More Races	●	●	●	●	●	●
Native Hawaiian or Other Pacific Islander		●	●	●	●	
White		●	●	●	●	●
Military		●				
Homeless		●				
Continuously Enrolled					●	
Noncontinuously Enrolled					●	
Minimum *N*-size	*	10	10	20	25	30

*REMIQS scores are only produced for schools in which at least 25 percent of enrolled students are historically resilient and marginalized.