Identification of Comprehensive Support and Improvement Schools in Ohio

The Influence of Selected Measures and System Design

Drew Atchison
Katie Hyland
Umut Ozek
Kerstin Le Floch
Steven Hurlburt

Institute of Education Sciences Award No: R305A200254

The research reported here was supported by the Institute of Education Sciences (IES), U.S. Department of Education, through Grant R305A200254, to the American Institutes for Research. The opinions expressed are those of the authors and do not represent the views of the IES or the U.S. Department of Education.

June 2023

AIR®
Advancing Evidence.
Improving Lives.

Identification of Comprehensive Support and Improvement Schools in Ohio

The Influence of Selected Measures and System Design

Drew Atchison
Katie Hyland
Umut Ozek
Kerstin Le Floch
Steven Hurlbut

June 2023

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

Contents

Exhibits

Executive Summary

For the past quarter-century, federal law has required states to measure school performance, to identify the lowest performing schools, and to provide support to identified schools. Such systems have invited critiques—that they overemphasize student achievement in reading and mathematics, exclude nonacademic features of school performance, and rely on measures closely associated with student characteristics. The Every Student Succeeds Act (ESSA) of 2015 ushered in a new era of accountability, enabling states to design complex accountability systems containing multiple measures intended to provide a more holistic picture of a school's overall performance. Based on these evaluations of school performance, states are required to identify the lowest performing 5% of Title I schools for comprehensive support and improvement (CSI).

Federal law under ESSA extends wide latitude to states regarding the selection, weighting, and aggregation of various measures to identify low-performing schools. These state policy decisions have the potential to influence which schools are identified for support. These decisions also have implications for how school and district leaders respond to accountability designations, and ultimately, they change student outcomes. In this report, we examine how specific design choices regarding Ohio's ESSA accountability system influence the set of schools identified as CSI.[1]

Key Findings

To evaluate school performance and identify CSI schools, Ohio incorporates six components: student achievement in reading, mathematics, science, and social studies; student progress in tested subjects; an equity measure (referred to as "gap closing"); literacy achievement in early grades (K–3); graduation rate; and a career readiness indicator ("prepared for success"). In practice, however, many schools in Ohio have not been rated on all accountability components. **In particular, because CSI elementary and high schools were less likely to meet requirements for minimum numbers of students, they were often held accountable for fewer components than what was typical of all schools. So, CSI schools were often identified based on a partial portrayal of their performance.**

State accountability systems include multiple—ideally complementary—measures, each of which should provide information on a different facet of school performance. However, if these

[1] In 2022, Ohio's accountability system experienced significant changes in how measures and ratings are calculated and presented. See the accountability system crosswalk for more information on the new system compared with the prior one. The findings presented here reflect Ohio's accountability system that was in place at the time of initial CSI school identification in 2018.

measures are strongly correlated with each other, they provide little new information. Indeed, in Ohio, **we found strong, positive correlations between the *achievement* and *gap closing* components, revealing that they provide similar information. In contrast, the *progress* and *K–3 literacy* components showed weaker relationships with the other components, suggesting that they provide more distinctive information about schools compared with the other components.**

School accountability systems are intended to measure school performance, not simply to reflect the features of the student population. Measures that are strongly correlated with school demographics may be seen as unfair—identifying schools based on the characteristics of the students they serve rather than on how well they serve their students. **In Ohio, although school performance on three accountability components—*achievement*, *graduation rate*, and *prepared for success*—was closely associated with the poverty level of enrolled students, the *progress* measure was only *weakly* associated with student poverty levels, thus putting schools on a more even playing field.**

In addition to the fairness of accountability components, the reliability of these measures is also an important consideration. Fluctuations in measurement from year to year could signal noise or randomness; consistency over time suggests that the measure is stable and not random. Of course, one would not want to see perfectly stable measures; very high correlations would suggest that schools had failed to improve. **For all components, there were moderate to strong correlations between ratings in 2018 and 2019, providing evidence that the Ohio measures are stable. *Achievement*, *graduation rate*, and *prepared for success* were the most highly correlated from year to year.**

Within a multiple measures accountability system, each performance measure should contribute information to determine a school's overall performance and have some amount of influence in determining which schools ultimately become CSI schools. However, some measures may be more influential in determining school performance if they are distinct from others or are rated more often. **In Ohio, the *progress* component had the most influence on schools' rankings.** As a reminder, this indicator was least associated with other accountability components and the poverty levels of students. The *progress*, *gap closing*, and *K–3 literacy* components each played a larger role than *achievement* in determining which schools were identified as CSI.

To generate an overall rating of school performance, Ohio first assigns points to individual performance measures based on ranges of performance in the underlying raw measure. We found that Ohio's approach to assigning points did not always mirror the distribution in performance on the raw measure. More importantly, **the *method* of assigning points**

influenced which schools were identified as CSI—and, in fact, had greater influence compared with most individual components.

Background and Policy Context

For the past quarter-century, federal law has required states to measure school performance and identify the lowest performing schools. Such systems have invited critiques—that they overemphasize student achievement in reading and mathematics, exclude nonacademic features of school performance, and rely on measures closely associated with student characteristics. In December 2015, the Elementary and Secondary Education Act (ESEA) was reauthorized as the Every Student Succeeds Act (ESSA), ushering in a new era of education accountability.

ESSA introduced a system of multiple measures, collectively intended to provide a more comprehensive picture of a school's overall performance. Like previous policies, ESSA requires state accountability systems to include student achievement in reading and mathematics, but it expands the set of required indicators to include another academic indicator, graduation rates, progress in English language proficiency (ELP; for English learners), and a measure of school quality and student success (Every Student Succeeds Act, 2015). ESSA also provides states substantial latitude in determining which specific measures to include under the accountability indicators for determining school performance and how to aggregate those measures to determine which schools are underperforming.

Years after the initial implementation of ESSA accountability systems, questions remain about how well these multiple-measures systems work and, in particular, about how design choices can influence which schools are identified for CSI. Our study team examined these other accountability-related questions in a larger study, through a grant from the National Center for Education Research (NCER) at the U.S. Department of Education; in this report, we focus on only one state, Ohio.

Specifically, we examined how specific design choices regarding the measures included in Ohio's ESSA accountability system and the method of aggregating measures affect which schools end up in the lowest 5% and are, hence, identified as "CSI." Drawing on administrative data from the 2017–18 school year, we conducted several analyses to better understand how Ohio's accountability system operates in practice, including descriptive analyses examining how many components each school was rated on and which components were most often missing, correlations among component points and between component points and student demographics, year-to-year correlations of component points, and simulations designed to identify the influence of specific components on CSI identification.

Overview of Ohio's Accountability System

To evaluate school performance and identify CSI schools, Ohio uses an index-based system, which combines multiple measures into a single index or score to describe the overall performance of a school. Schools ranked in the bottom 5% based on the aggregate index are designated for CSI.[2]

Describing school performance through an index is intended to provide simple, easy-to-understand information. However, states must make numerous technical decisions to operationalize the calculation of the index. Each decision influences the score that schools receive and which schools are identified as low performing. These decisions include the selection of measures, the weights each measure receives, and how they are combined into an overall score.

In the case of Ohio, before 2022, 11 accountability measures were first aggregated into six components, which were then aggregated into an overall index representing school performance.[3] Some components and measures are rather complex, and several individual measures represent the aggregation of various data points. In brief, the components and measures used in Ohio's accountability system prior to 2022 were as follows:

Components Versus Measures

For this report, we use the terms **components** and **measures** to describe the elements on which schools are rated.

We use the term **components** to describe the core elements of Ohio's (e.g., achievement).

Measures are the specific data points that contribute to performance within a given component. In the case of Ohio, each component consists of one to four measures.

- The **Achievement component** consisted of the following two measures—the Performance Index and the Indicators Met measure:
  - The **Performance Index** represented the average aggregate performance level across assessments for each subject area (English language arts [ELA], mathematics, science, and social studies) in Grades 3–8 and, for high school, end-of-course tests in ELA and mathematics.[4] Students received more points depending on their levels of performance.[5]
  - The **Indicators Met** measure represented the percentage of up to 26 possible performance indicators met by a school. The indicators were based on a series of up to

[2] High schools with a graduation rate below 67% are also designated as CSI schools.
[3] In 2022, Ohio's accountability system experienced significant changes in how measures and ratings are calculated and presented. See the accountability system crosswalk for more information on the new system compared with the prior one.
[4] Science and social studies are not included at the high school level as students have alternative options for earning graduation points in these subject areas.
[5] All tests have six performance levels—Advanced Plus, Advanced, Accelerated, Proficient, Basic, and Limited. For each subject area, the percentage of students scoring at each performance level is calculated and then multiplied by the point value assigned to that performance level (Advanced Plus = 1.3; Advanced = 1.2; Proficient = 1.0; Basic = 0.6; Limited = 0.3).

23 state assessment results that measure the percentage of students proficient or higher in a grade and subject[6] and three additional performance indicators—the chronic absenteeism improvement indicator, end-of-course improvement indicator, and gifted indicator.

- The **Progress component** reflected the academic growth of all students by using statistical modeling known as "value added" to compare students' actual progress with an expected level of progress for a test in a given subject/year based on all prior test scores for a given student. If students in a school performed higher than expected, on average, that school had positive value added. The component included the following four **value-added measures**:
  - Progress for all students in the school
  - Progress for students who are gifted
  - Progress for students with disabilities
  - Progress for students whose academic performance is in the lowest 20% of students statewide based on scores from the current and prior school years

  These four measures together generated the Progress component score, with the *all students* measure representing 55% of the total score and the other three measures representing 15% each.

- The **Gap Closing component** measured the academic performance of specific groups of students (e.g., by race, economic disadvantage status, disability status, and English learner status) against the expected performance goals for that subgroup to determine whether gaps existed and whether the subgroup was making progress (i.e., the subgroup gap is reduced). The Gap Closing component consisted of **annual measurable objective (AMO) measures** in four domains: ELA, mathematics, graduation rates, and for English learners only, progress in achieving English language progress.

  Points were awarded for each AMO domain by aggregating the points earned for each evaluated group of students divided by the total possible points. To determine the overall score for the Gap Closing component, the percentages for the four AMO domains were averaged together.[7]

---

[6] For each state assessment indicator, at least 80% of students had to score proficient or higher to receive credit for the corresponding indicator.

[7] Each AMO measure had a total of 100 possible points. Each AMO domain contributed 25% to the Gap Closing component score (provided all four apply to the school).

- The **Graduation Rate component** consisted of two measures: the **4-year adjusted-cohort graduation rate** and the **5-year adjusted-cohort graduation rate**.[8] The two measures were combined to arrive at the overall component score, with the 4-year rate weighted at 60% and the 5-year rate weighted at 40%.
- The **Improving At-Risk K–3 Readers component** consisted of a single measure, which represented the percentage of kindergarten through Grade 3 students who were not on track to reading proficiency but improved to become on track to reading proficiency using a combination of diagnostic reading assessments in Grades K–3 and the Grade 3 Ohio State Test.
- The **Prepared for Success component** included multiple **measures of college and career readiness**. The component was graded based on the percentage of all students from a school's four- and five-year combined graduation cohort who demonstrated college and career readiness. A school earned 1 point for each student who earned any of the following:
  - A remediation-free score on the ACT or SAT
  - An Honors diploma
  - At least 12 points through an industry-recognized credential or group of credentials in one of 13 high-demand career fields

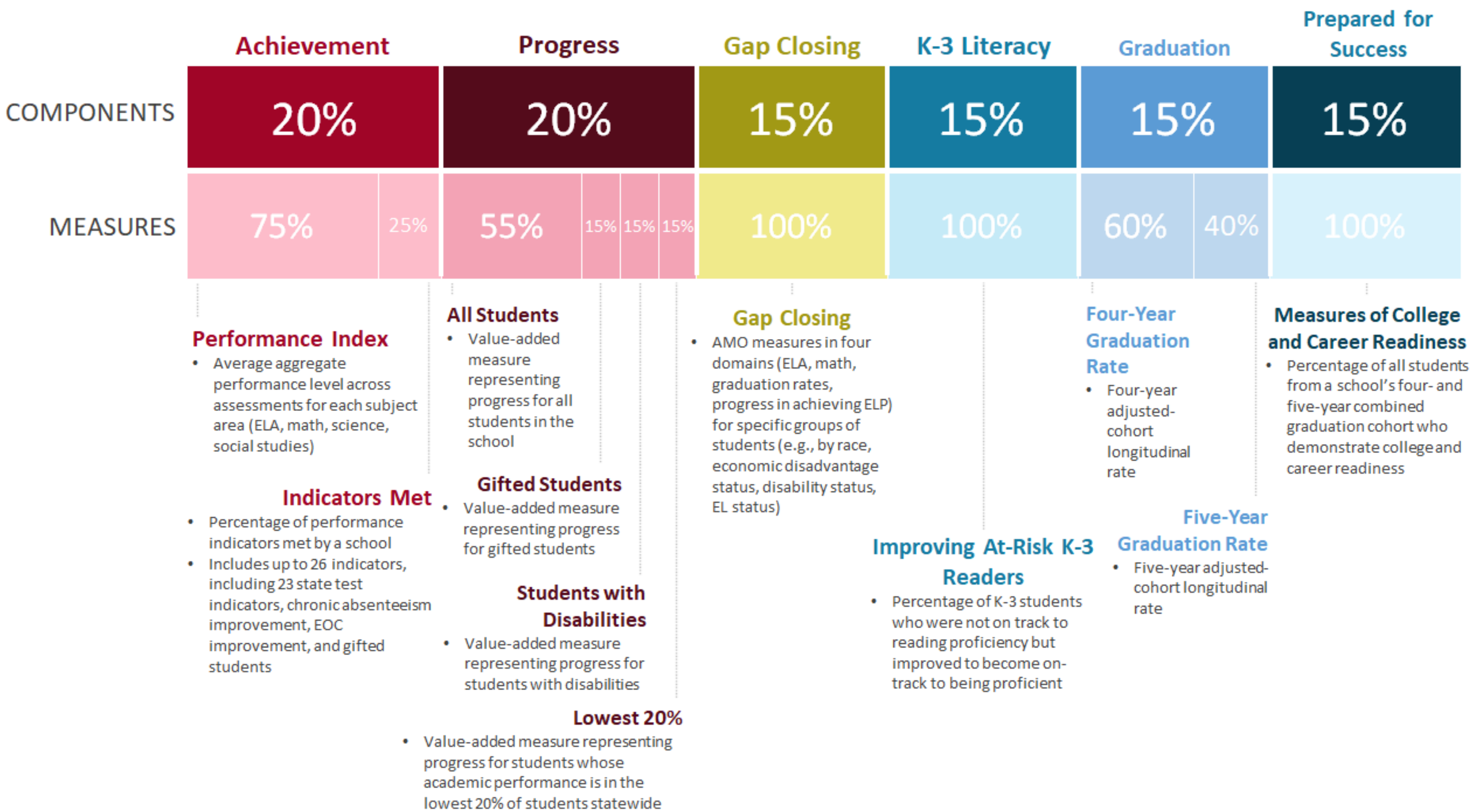
  Schools received an additional 0.3 points for each student who met one of the three conditions outlined above and one of the following: scoring a 3 or higher on an advanced placement (AP) test, scoring a 4 or higher on an international baccalaureate (IB) test, or earning at least three dual enrollment credits.

After assessing each school based on these components and measures, overall school ratings were then determined using the following steps:

- Within each component, measures were converted into points on a 0 to 5 scale in 0.25-point or whole-point increments based on a school's level of performance. Once converted, the points were weighted based on the percentage that each contributes to the component rating, as detailed in Exhibit 1, and then summed to determine the component ratings.
- Points for each applicable component were similarly weighted based on the percentage that each contributes to the overall school rating (see Exhibit 1) and then summed to achieve the overall rating.

[8] The 4-year (and 5-year) adjusted-cohort graduation rate is calculated by dividing the number of students who graduate in 4 years or less (5 years or less) with a regular or honors diploma by the total number of students in the final adjusted cohort for the graduating class.

Exhibit 1. Measures and Components Used in Ohio's Accountability System and Their Weights (2019)



Note. When one or more measure is not rated, the remaining measures contribute proportionally to the overall component rating. If a component has no rated measures, then the component is not rated. Likewise, when one or more component is not rated, the remaining components contribute proportionally to the overall rating. EOC is end-of-course assessments. For more information on how measure and component ratings are calculated, see Ohio Department Education's resources and technical documents.

Exhibit 2 shows how Ohio's components and measures satisfy the required indicator performance categories under ESSA.

Exhibit 2. ESSA Indicator Performance Categories and Ohio's Components and Measures

ESSA indicator performance categories	Ohio's components/measures
Student achievement in reading and mathematics	Achievement component, performance index measure
State-selected academic indicator for elementary and middle schools	Progress component, value-added measures
Graduation rate	Graduation rate component
ELP progress for English learners	ELP improvement measure (contributes to gap closing component)
School quality or student success	Science and social studies achievement (included within achievement and progress components) Indicators met measure (includes chronic absenteeism and performance of students who are gifted) Gap closing component Prepared for success component

Situating Ohio's Approach Within the Broader Landscape of ESSA Accountability Systems

Although Ohio uses measures that are similar to those in other states, the complexity of the Ohio accountability system is distinctive. Like most states, Ohio uses an index-style method to aggregate individual performance measures into an overall score. In addition, Ohio incorporates chronic absenteeism into its accountability system and, as required by ESSA, English learner progress. However, neither chronic absenteeism nor ELP progress is a stand-alone measure; rather, each is one of several elements within broad components. Specifically, chronic absenteeism is one of a possible 26 elements within the "indicators met" measure of the achievement component, and English learner progress is only one of 31 specific achievement targets within the AMO measure of the gap closing component.

In addition to its unusual degree of complexity, Ohio's approach to measuring school performance differs from that of other states in several ways. First, with respect to academic achievement, unlike many states that focus only on mathematics and ELA, Ohio also incorporates assessments into social studies and science (Ohio Department of Education [ODE], 2020b). Second, Ohio is one of a few states that incorporates measures of early learning as a performance measure (English, 2017). Third, Ohio also incorporates measures of subgroup performance in a prominent way, which is also uncommon across states. In particular, the gap closing component measures progress toward subgroup-specific goals in ELA, mathematics, and graduation rates, as well as measures English learner progress. Ohio's progress component also explicitly measures growth (as value added) for several student subgroups, including the lowest performing 20% in each school, students with disabilities, and students who are gifted.

Objectives and Research Questions

This report is part of a broader study examining the underlying theory of action of accountability in the context of ESSA. Other study components examine how principals in CSI schools approach school improvement, the supports provided to CSI schools, and whether student outcomes in CSI schools improve. The primary objective of the analyses presented in this report is to better understand the choices made by Ohio in designing its ESSA accountability system and the implications of these choices for which schools are identified as CSI. In particular, we addressed the following research questions:

RQ1: Which components, as defined and measured within Ohio's accountability system, are most often used to evaluate school performance?

Under ESSA, schools are intended to be rated on a variety of school performance measures, providing a more holistic and comprehensive measure of school performance compared with prior accountability policies, which were largely based on students' test performance. However, not all schools receive ratings on all measures. In particular, some measures may not apply to all schools because they are grade specific (such as graduation rates), and schools only receive a rating if they have sufficient numbers of students contributing to the calculation of performance for a given measure. For this research question, we investigate the number of accountability components for which schools receive ratings and which components are most often not rated.

RQ2: How are Ohio's accountability components related to each other and to school demographics?

ESSA requires states to include multiple indicators of school performance with the intent that each will provide complementary—not duplicative—information. Highly correlated measures provide limited information about school performance; novel measures may yield new insights and influence a school's accountability rating. The inclusion of each additional accountability measure will have less influence on a school's CSI status if the additional indicator is highly correlated with other indicators already included in rating school performance (Harris & Liu, 2018). In other words, measures should have more influence if they add unique information about school performance not already represented in the other accountability measures. For this research question, we conducted an exploratory analysis to determine how distinctive each of Ohio's accountability components is from the other components included in the accountability system.

School accountability systems are intended to measure school performance, not simply to reflect features of the student population. Measures that are strongly correlated with demographics may not be good measures of school performance (Di Carlo, 2019) and may be seen as unfair—particularly to high-poverty schools (Wright & Petrilli, 2017). However, these measures may be useful for identifying schools needing high levels of support. As such, we also investigate the extent to which each accountability component is related to school demographics.

RQ3: To what extent are ratings on measures and indicators consistent from year to year?

Measurement of school performance should exhibit some consistency over time; large fluctuations from year to year could signal noise or randomness in the measure. In such cases, the information provided by the measure may not be meaningful or reliable. Because accountability systems are intended to foster school improvement over time, one would not want perfect consistency in school performance over time as that would mean schools would have little hope of improving their performance ratings. However, on balance, we would expect some amount of stability over time given that academic performance tends to change incrementally. For this research question, we examine whether the measures included in Ohio's accountability system exhibit consistency over time signifying the stability of the measures.

RQ4: What is the influence of individual accountability components on schools' CSI designation and performance ranking under Ohio's accountability system?

Multiple measures are included in accountability systems to provide a multifaceted perspective on school performance. If certain measures have little or no influence on schools' ratings, however, policymakers and educators could reasonably question whether the school ratings are indeed multidimensional. Several of the prior research questions investigate aspects of influence. Specifically, RQ1 examines which components are most often unrated. All things equal, components that are more often rated for schools will have more influence. In addition, RQ2 examines the uniqueness of components. As explained, components that are more distinctive should have more influence on school ratings as they provide new and different information. For this research question, we examine influence more explicitly by analyzing how the exclusion of individual components from the accountability system would change school performance rankings and the set of CSI schools.

RQ5: How does Ohio's approach to assigning points to measures influence schools' CSI designation?

Before aggregating measures into a single index score, each measure must first be placed onto a common scale. To do so, Ohio assigns points to each measure on a 0 to 5 scale. Because Ohio's method for assigning points is unusual, we also investigate the extent to which the method of point assignment influences both CSI designation and performance rankings.

Methodology

To address the research questions about Ohio's school accountability system and its effects on the identification of CSI schools, we analyzed data from extant data sources using a variety of analytic methods.

Data and Sample

This report draws primarily on school-level data reported through the Ohio Department of Education's (ODE) school report cards, including school-level data on the following:

- School performance on the components and measures used in Ohio's accountability system (e.g., achievement, progress, and K–3 literacy) for the 2017–18 and 2018–19 school years[9]
- Accountability designations for the 2018–19 school year[10]

Additional data on school demographic composition (e.g., racial/ethnic composition, economic disadvantage, and students with disabilities), total enrollment, and school level (elementary, middle, and high school) for the 2017–18 school year[11] were drawn from the ODE.

The schools included in specific analyses vary, depending on the research questions addressed. Most analyses in this report include all public schools in Ohio, overall or by school level. Analyses related to RQ1 examine all public schools as well as CSI schools only. Most analyses focus on performance data aggregated to the component level to have a manageable number of variables and allow for clearer and more succinct presentation of results. However, at times, we also examine outcome data at the level of individual measures.

[9] School performance data for 2018–19 were used to measure year-to-year consistency (RQ3).

[10] School accountability designations in a given year are based on performance in the prior year. A school's performance in the 2017–18 school year, for example, would affect its accountability designation for 2018–19.

[11] The report uses school demographic and characteristic data for 2017–18, the year for which 2018–19 CSI designations are based.

Analytic Approach

Descriptive Analyses

The study used descriptive analyses to investigate which accountability components are used to evaluate school performance (RQ1). Specifically, the study examined the number of performance components on which schools were rated and, for each component, the percentage of schools that were not rated due to having an insufficient number of students contributing to the calculation of performance.[12] Results for these analyses compare all public schools with CSI schools by school level.

Descriptive analyses were also conducted to explore the connection between economic disadvantage and accountability components (RQ2) by examining differences in performance between schools in the highest and lowest poverty quartiles, as defined by the percentage of economically disadvantaged students. For each accountability component, we examined differences between the two groups of schools in terms of the distribution of the number of points assigned.

Correlational Analyses

The study used simple correlational techniques to examine the relationship among accountability component points and between accountability component points and school demographics (RQ2) and year-to-year consistency in the points earned by component across schools (RQ3). In addition, multiple regression analyses were performed to determine whether other accountability components, student demographics, and school characteristics were good predictors of school performance (points earned) on a given component (RQ3).

Simulation Modeling

To examine the influence of individual accountability components on the likelihood of CSI designation (RQ4), the study used simulations comparing the actual set of CSI schools (under existing accountability rules) with the set that would be identified if a specific component were excluded from school performance ratings. The difference between the set of actual and simulated CSI schools—measured as both the percentage of newly identified schools (those identified under the simulation but not actual calculations) and the percentage of schools no longer identified (those identified using actual calculations but not under simulation)—reflects the degree of influence for a given indicator.

Although differences in schools' CSI designation between the actual and simulated accountability rules is an important measure of influence, whether a school changes CSI designation is also a function of how close a school is to the CSI cutoff under the existing

[12] For most components, a rating is calculated if a school has at least 10 accountable students with reportable data.

accountability rules. As a second way to assess the influence of components through these simulations, we examined the extent to which schools' performance rankings change by ranking each school based on its actual and simulated performance index score.[13]

In addition to examining the influence of individual components on CSI identification and school performance rankings, the study also investigated the extent that the method of assigning points to measures influences CSI identification and school performance rankings. To do so, we first descriptively examined the distributions of performance in the underlying measures and how these distributions changed once points were assigned. We then conducted the following two simulations that model school rankings under alternative systems of assigning points to measures:

- **Simulation A** assigned points in a continuous uniform distribution, assigning points proportionally to each schools' percentile rank on a given measure.[14]
- **Simulation B** preserved the existing shape of the distribution in the measures prior to converting to points (termed as the "natural distribution") and assigned points based on standard deviations from the mean.[15]

Results

In this section, we present the results of our analyses organized by research question.

RQ1: Indicators of School Performance: Intent vs. Reality

Under Ohio's school accountability system, each school is rated based on up to six accountability components. Given that some components only apply to certain grade levels, most schools are rated based on three to five components rather than on all six. Here, we

[13] For this analysis, a value of 1 represented the ranking of the lowest performing school (or schools in the case of a tie) and schools were ranked up to a maximum of 3,423—the total number of schools represented in the performance data. We then calculated the absolute value of the difference between actual and simulated ranking and converted this into a percentile by dividing the difference in ranking by the total number of schools. In contrast to the percentage of CSI schools changing designation, this measure is not dependent on whether schools are initially close to the CSI cutoff.

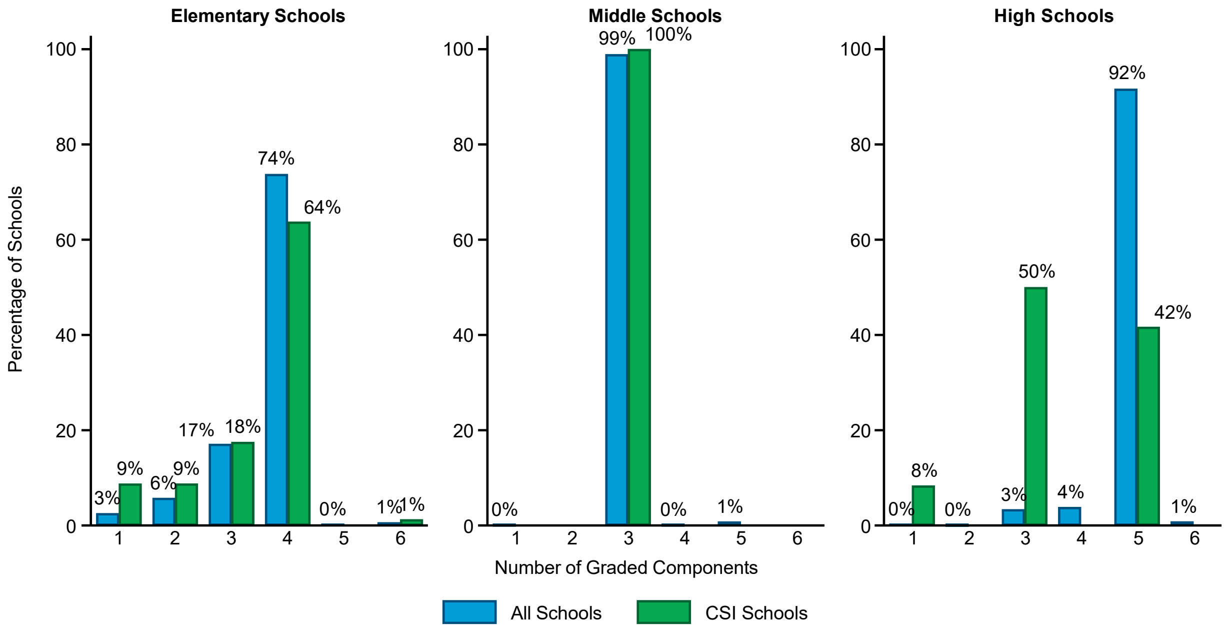
[14] For this simulation, the highest ranked school or schools for a given measure received 5 points and the lowest ranked school or schools received 0 points and points increased evenly according to percentile rank. In other words, we multiplied the precise percentile rank on a scale of 0 to 1 by 5 to get points.

[15] To operationalize points for this simulation, we converted each measure into a *z*-score that has a mean of 0 and a standard deviation of 1. We then added 2.5 points to the *z*-score so that a school performing at the average received 2.5 points. We capped the range of points at 0 and 5 so that any school performing worse than 2.5 standard deviations below the average received 0 points and any school performing better than 2.5 standard deviations above the average received 5 points. We then calculated influence in the same way as described for the prior simulations—examining change in CSI designations and performance rankings.

present the result of analyses investigating the number of components on which schools are rated and the components that are most often unrated.

CSI elementary and high schools often had fewer graded components than what was typical of all schools. This finding was particularly acute at the high school level: Of all high schools in Ohio, 92% were rated on five components and approximately 3% of high schools were rated on only three components. Among CSI high schools, however, almost 42% were rated on only three components (Exhibit 3). Among all elementary schools in Ohio, less than 10% were rated based on one or two components, collectively, compared with 18% of CSI elementary schools.

Exhibit 3. Percentage of All Schools and CSI Schools by Number of Components

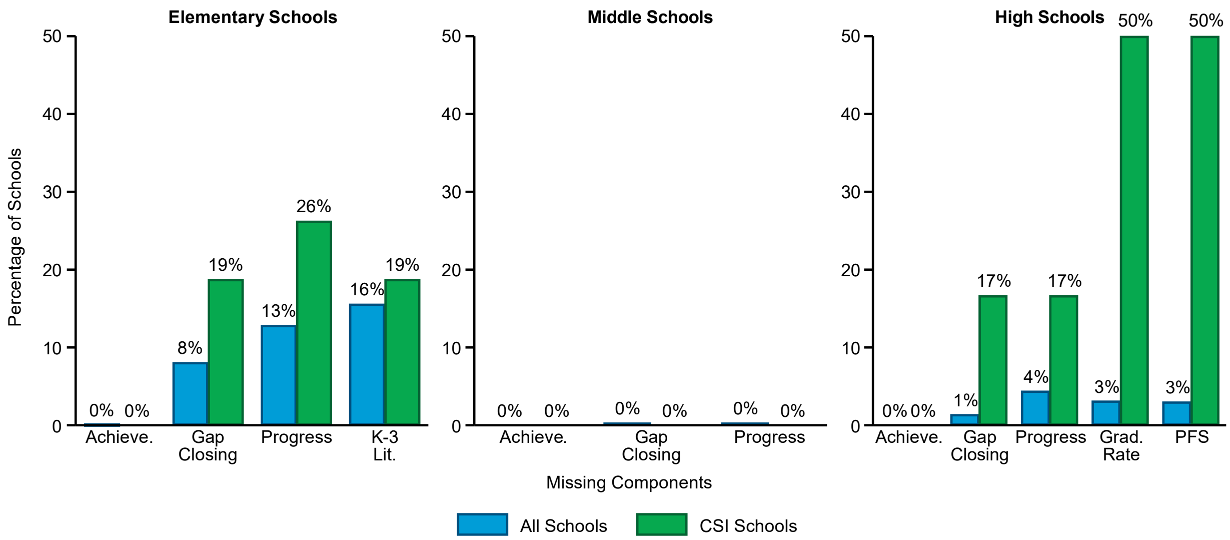


Note. *N* of all elementary schools = 1,931; *N* of CSI elementary schools = 80; *N* of all middle schools = 630; *N* of CSI middle schools = 15; *N* of all high schools = 862; *N* of CSI high schools = 12. CSI schools only include those identified based on low performance on the state index system rather than on graduation rates below 67%. Schools are categorized by grade level according to the grade range (K–5, 6–8, or 9–12) with the largest enrollment. Therefore, some schools classified as elementary or as middle schools were rated on the graduation measure because they served students in Grade 12.

CSI schools more often missed the gap closing and progress components at both elementary and high school levels and more often missed the graduation rate and prepared for success components at the high school level. Among CSI elementary schools, more than a quarter lacked data that would enable them to be held accountable for the progress component, compared with 13% of all elementary schools (Exhibit 4). Nearly 20% of CSI elementary schools were not evaluated for the gap closing and K–3 literacy components, compared with 8% and 16% for the gap closing and K–3 literacy components, respectively, for all elementary schools. No school was missing the achievement component.

Among the 12 CSI high schools identified based on performance (as opposed to those identified based solely on the graduation rate), 6 were missing the graduation rate and prepared for success components and another 2 were missing the gap closing and progress components. For all high schools, no component was missing more than 4% of the time. High schools missing multiple components were typically much smaller than those rated on five components. Several also had nontraditional grade configurations (e.g., ninth and tenth grade only).

Exhibit 4. Percentage of All Schools and CSI Schools Missing a Given Component



Note. *N* of all elementary schools = 1,931; *N* of CSI elementary schools = 80; *N* of all middle schools = 630; *N* of CSI middle schools = 15; *N* of all high schools = 862; *N* of CSI high schools = 12. CSI schools only include those identified based on low performance on the state index system rather than on graduation rates below 67%. Achieve. = achievement; K-3 Lit. = K–3 literacy; PFS = prepared for success.

RQ2: Correlations Among Performance Measures and Between Performance Measures and School Demographics

Under ESSA, the evaluation of school performance is intended to include multiple measures, moving away from NCLB's more narrow focus on mathematics and ELA proficiency. The intent is for each measure to provide different information about a school's performance, resulting in a more complete picture of school quality. Prior to ESSA, policymakers and researchers critiqued traditional measures of school performance (primarily achievement in reading and mathematics) as too closely associated with school demographics (e.g., McEachin & Polikoff, 2012), a relationship that could potentially be disrupted by the introduction of additional measures under ESSA. In this section, we describe the relationships among accountability measures and between accountability measures and school demographics.

Relationships Among Performance Measures

We found strong, positive correlations between achievement and gap closing across all three school levels. At the elementary, middle, and high school levels, the correlation between achievement and gap closing was above 0.7 (Exhibits 5–7). In addition, at the high school level, we found strong, positive correlations between the achievement, prepared for success, and graduation rate components. In particular, the correlations between the achievement component with both the graduation rate and prepared for success components were greater than 0.8. This finding underscores the strong relationship among these varied features of school performance.

The progress component was *least* correlated with other components across all three schooling levels. Among elementary schools, the correlations between progress and achievement and gap closing were 0.41 and 0.42, respectively. These correlations were lower for middle schools (approximately 0.35, each) and somewhat higher for high schools (0.611 and 0.543, respectively). This suggests that the progress component provides unique information about schools relative to other components. At the elementary level, the K–3 literacy component was also distinct from the others, with similar levels of correlation to achievement and gap closing as the progress component.

Exhibit 5. Correlations Among Performance Component Points for Elementary Schools

Component	Achievement	Gap closing	Progress	K–3 literacy
Achievement	1.00			
Gap closing	0.73	1.00		
Progress	0.41	0.42	1.00	
K–3 literacy	0.42	0.36	0.14	1.00

Note. N = 1,427. Includes only schools with these four components and no others.

Exhibit 6. Correlations Among Performance Component Points for Middle Schools

Component	Achievement	Gap closing	Progress
Achievement	1.00		
Gap closing	0.71	1.00	
Progress	0.35	0.36	1.00

Note. N = 886. Includes only schools with these three components and no others.

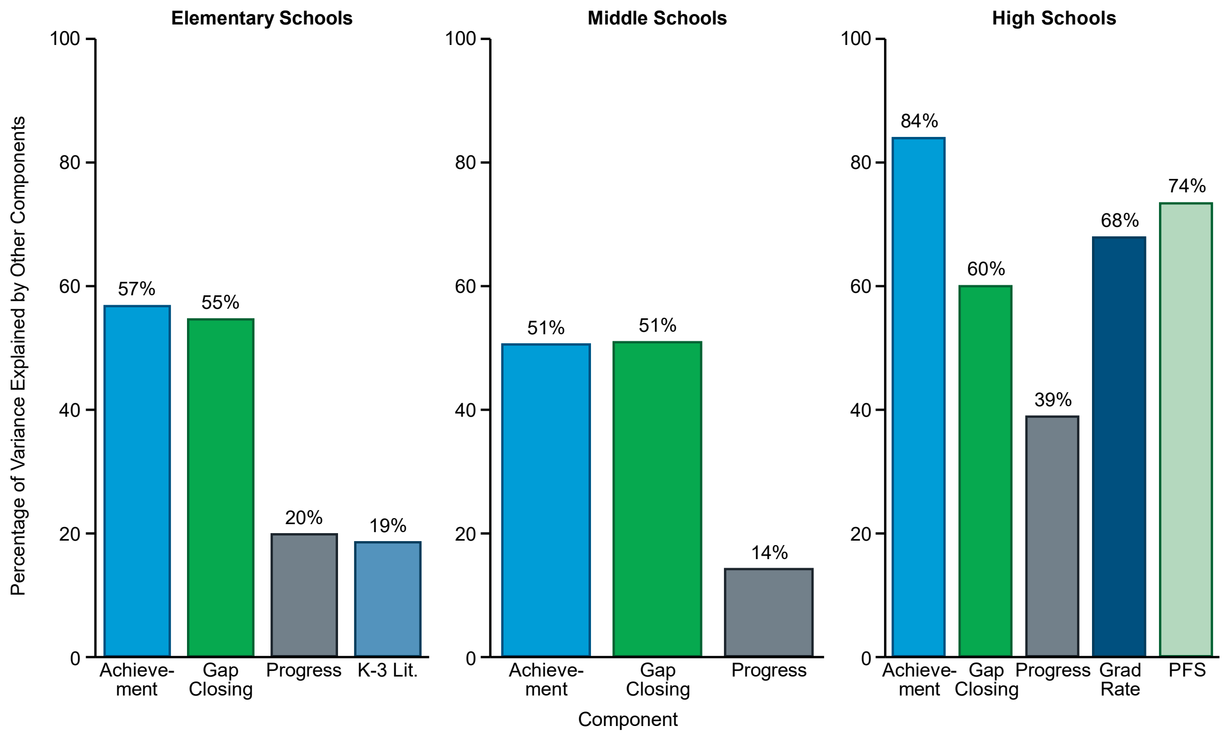
Exhibit 7. Correlations Among Performance Component Points for High Schools

Component	Achievement	Gap closing	Progress	Graduation rate	Prepared for success
Achievement	1.00				
Gap closing	0.75	1.00			
Progress	0.61	0.54	1.00		
Graduation rate	0.81	0.69	0.49	1.00	
Prepared for success	0.85	0.60	0.50	0.74	1.00

Note. N = 816. Includes only schools with these five components.

The progress component was also the most distinct component across all grade levels as evidenced by the low percentage of variance explained by other components. To complement the correlational analyses, we next accounted for each indicator's relationships with all the other indicators in the accountability system simultaneously. Consistent with the individual correlations, we found that the percentage of variance explained by other components is low for the progress component across all three grade levels—20%, 14%, and 39% of variance explained at the elementary, middle, and high school levels, respectively. For elementary schools, the K–3 literacy component provides distinct information about school performance (see Exhibit 8). In contrast, the achievement measure has the highest percentage of variance explained by other components, indicating substantial overlap.

Exhibit 8. Variance Explained by Other Components



Note. Variance explained is the R^2 from a regression where the given component is the outcome variable, and the remaining components are the explanatory variables. The number of observations is 1,427 for elementary schools, 886 for middle schools, and 816 for high schools. K-3 Lit. = K–3 literacy; PFS = prepared for success.

Relationships Between Performance Components and School Demographics

The progress and K–3 literacy components, which were most distinct from other components, were also least correlated with student demographics (Exhibit 9). The progress component exhibited moderate-to-weak negative correlations with the percentage of students who are economically disadvantaged (–0.31), the percentage of students with disabilities (–0.20), and the percentage of students who are Black (–0.29). The correlations between K–3 literacy and student demographic variables were of similar magnitude and direction.

Across all components, economic disadvantage was the student demographic variable most negatively correlated with performance. For example, the correlation coefficient between achievement and the percentage of students who are economically disadvantaged was –0.68. The magnitude of the negative correlations of the graduation rate and prepared for success components with economic disadvantage was even greater (–0.74 and –0.70, respectively).

Exhibit 9. Correlations Between Performance Measures and School-Level Student Demographic Percentages

Component	Students with disabilities	English learners	Econ. disad.	Black	Hispanic	Asian	Other non-White
Achievement	–0.33	–0.15	–0.68	–0.52	–0.21	0.25	–0.21
Gap closing	–0.31	–0.13	–0.51	–0.41	–0.18	0.14	–0.21
Progress	–0.20	–0.05	–0.31	–0.29	–0.08	0.08	–0.13
K–3 literacy	–0.17	–0.09	–0.33	–0.29	–0.09	0.03	–0.10
Graduation rate	–0.55	–0.23	–0.74	–0.57	–0.22	0.15	–0.37
Prepared for success	–0.46	–0.17	–0.70	–0.46	–0.20	0.36	–0.22

Note. Number of observations differs for each component: 3,421 for achievement, 3,252 for gap closing, 3,134 for progress, 1,639 for K–3 literacy, 858 for graduation rate, and 861 for prepared for success. Econ. disad. = economic disadvantage.

Further analyses confirmed that progress and K–3 literacy were least associated with student characteristics. We used regression analysis to examine the percentage of overall variance in outcomes that was explained collectively by student demographics, along school grade level and total enrollment (see Exhibit 10). These predictors explained just 13% of the variance in K–3 literacy and 14% of the progress component but as much as 65% of the variance for graduation rate. In addition to graduation rate, student demographics and school characteristics explained 58% or more of the variance in achievement and prepared for success.

As a reminder, Ohio converts each measure into points on a 0 to 5 scale and then aggregates measures into components. To further explore the relationship between accountability components and student characteristics, we examined how the distribution of points differs in the highest and lowest poverty quartiles of schools.

Exhibit 10. Variance in Component Performance Explained by Student and School Characteristics

Component	Variance explained (R^2)	Number of observations
Achievement	0.58	3,421
Gap closing	0.40	3,252
Progress	0.14	3,134
K–3 literacy	0.13	1,639
Graduation rate	0.65	858
Prepared for success	0.61	861

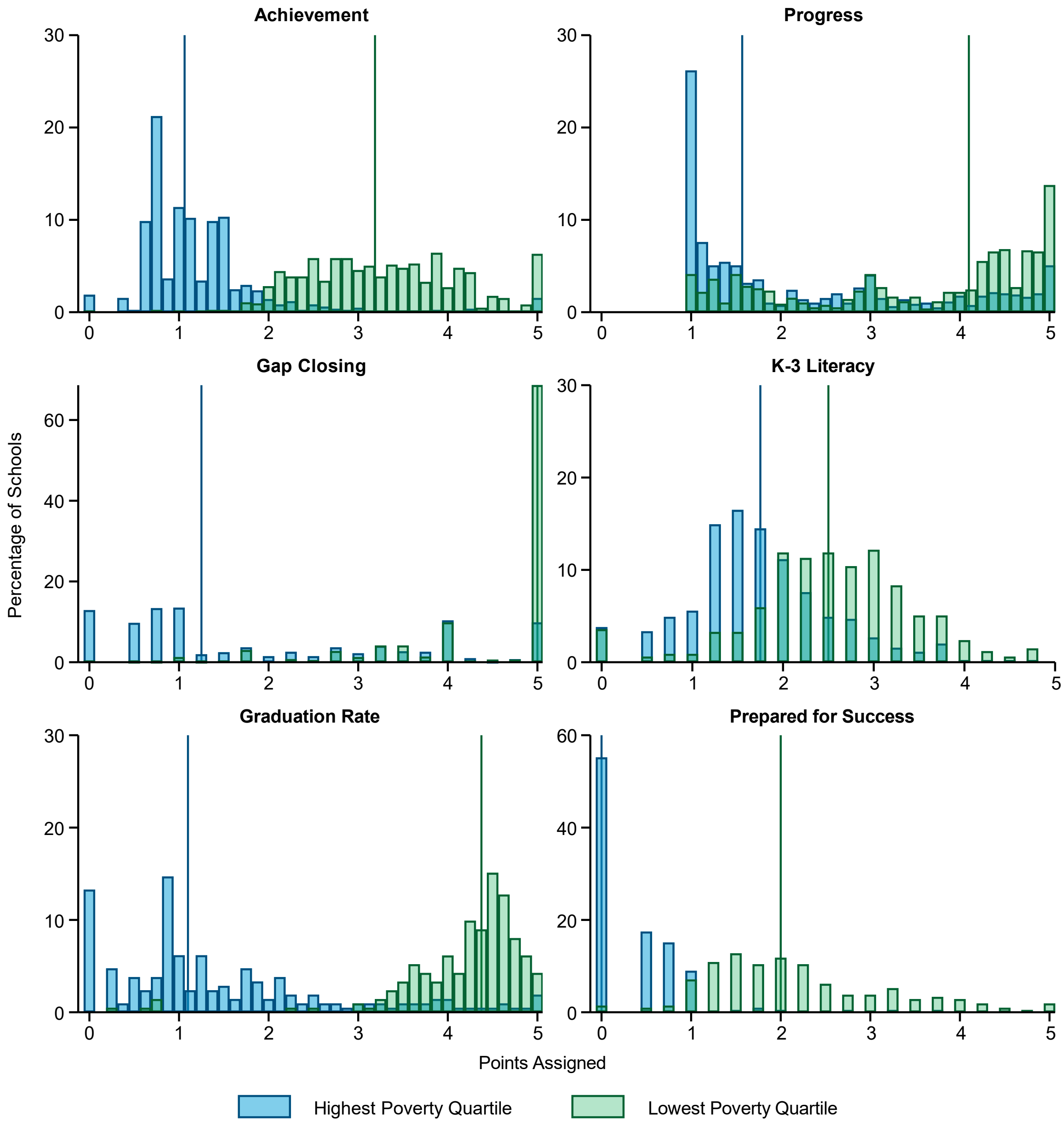
Note. Variance explained is the R^2 from a regression where the given component is the outcome variable, and student demographic percentages, school size, and school level are the explanatory variables.

Across all components, higher poverty schools were typically awarded fewer points than lower poverty schools (Exhibit 11). However, for some components, the differences between high- and low-poverty schools were much more pronounced.

School performance on three measures was closely associated with the poverty level of enrolled students: achievement, graduation rate, and the prepared for success component. This finding indicates that a few high-poverty schools perform as well as or better than even the lowest performing low-poverty schools in these components. For each of these components, low-poverty schools typically earn 2+ points more than high-poverty schools.

By contrast, for the progress and K–3 literacy components (and to a lesser extent, the gap closing component), notable overlap occurred in performance between high- and low-poverty schools. For these components, a sizable share of high-poverty schools performed at levels comparable with low-poverty schools. However, even for the progress component, a wide difference was found in the number of points earned by the median high- and low-poverty schools. In particular, the median high-poverty school earned slightly more than 1.5 points on the progress component compared with more than 4.0 points for the median low-poverty school. The K–3 literacy component had the smallest difference in points earned between high- and low-poverty schools, with the median low-poverty school earning less than one more point than the median high-poverty school.

Exhibit 11. Distribution of Performance and Average Performance of Schools in the Highest and Lowest Poverty Quartiles



Note. Vertical lines represent the median values for the highest and lowest poverty quartiles. Both the highest and lowest poverty quartiles have slightly more than 800 schools.

Collectively, these analyses demonstrate that even though some measures of school performance are indicative of family poverty levels, this is not always the case. Most notably, when schools are evaluated based on student progress—arguably more indicative of a school's contribution to student learning than status measures of achievement—the socioeconomic level of the student population is less influential.

RQ3: Year-to-Year Consistency of Performance Indicators

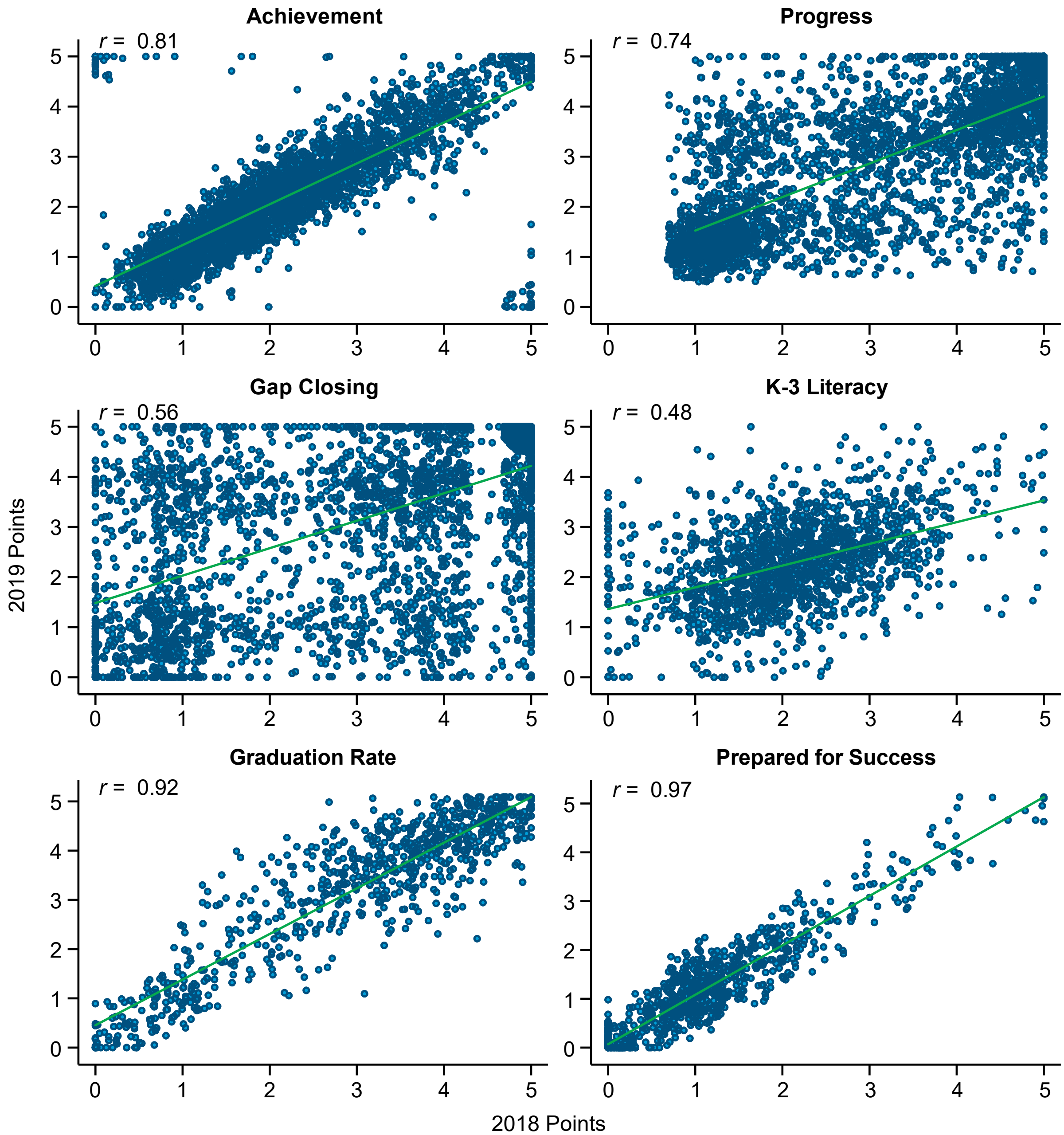
As described in the previous section, some components—particularly progress and K–3 literacy—are more distinct from other components and school demographics. Therefore, these measures could add unique information about school quality and could be fair insofar that all schools, regardless of the types of students served, have a shot at performing well. However, the less strong associations for these components could be a sign that they are infused with measurement error, making the information they provide less meaningful and reliable. To better understand the degree of randomness or unreliability in these measures, we next examined the year-to-year consistency in points earned by component across schools.

For all components, there was moderate-to-strong correlation between the points assigned in 2018 and 2019 (Exhibit 12). Across components, the correlations between points assigned in 2018 and 2019 ranged from 0.48 to 0.97.

K–3 literacy and gap closing had the lowest correlations across the 2 years at 0.48 and 0.56, respectively. However, these correlations were still moderately strong. Interestingly, the progress component had a substantially stronger year-to-year correlation (0.74) compared with both K–3 literacy and gap closing, despite having similar or lower correlations to student demographics and having a similar or lower amount of its variance explained by other measures.

Consistent with the strong relationship to demographics and other measures, achievement, graduation rates, and prepared for success were highly correlated from year to year. For these components, the correlations were at least 0.81 (achievement) and as high as 0.97 (prepared for success). Visually, the associations from year to year are apparent as demonstrated by the clear clustering of dots around the lines of best fit. In contrast, for progress, gap closing, and K–3 literacy, the dots cover most of the area of the graphs, indicating a larger range of possible outcomes in 2019 for a given outcome level in 2018.

Exhibit 12. Correlations Between 2019 and 2018 Component Points



Note. Green lines depict the line of best fit. *r* represents the correlation coefficient. The number of observations is more than 3,000 for achievement, progress, and gap closing; 1,519 for K–3 literacy; and approximately 820 for graduation rate and prepared for success.

These year-to-year correlations confirm some expected relationships and convey some good news for the Ohio system. Generally high correlations from one year to the next suggest that the measures are not displaying excessive fluctuations. Moreover, it is not surprising that measures that correlate highly with demographics tend to be stable over time, because demographics do not change much over time. However, the progress measure – that which is least associated with demographics—provides evidence that some schools are able to change their performance from year to year. Indeed, we should see some movement over time if the accountability system is working as it should.

RQ4: Influence of Performance Indicators

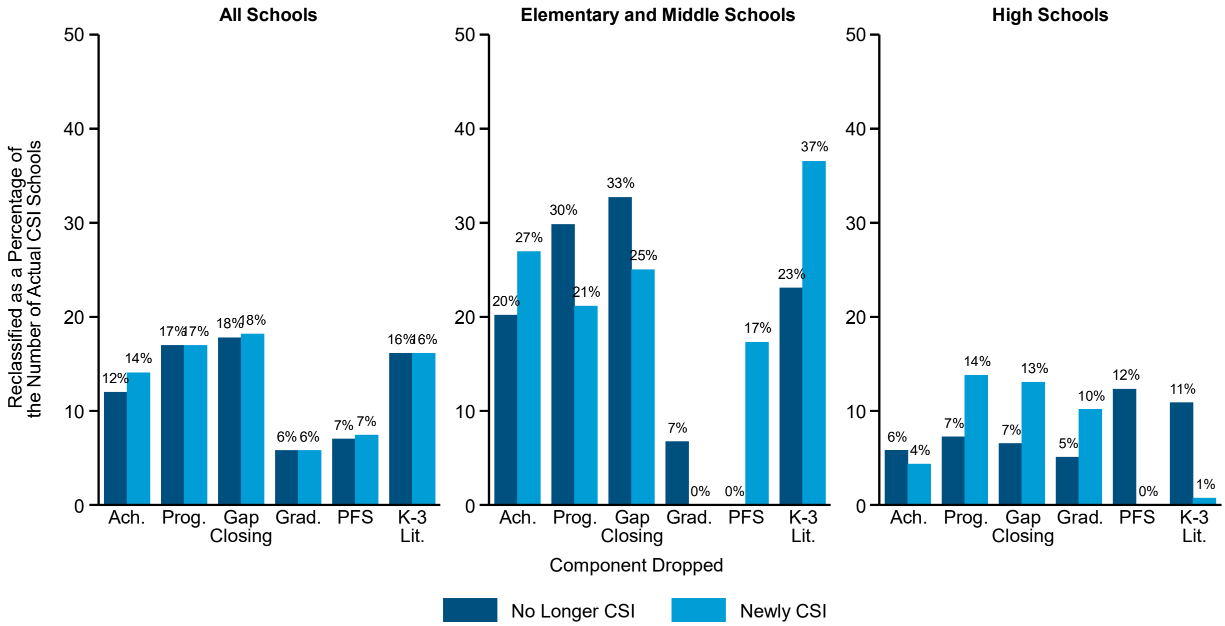
Within a multiple measures accountability system, each measure or component should contribute information to determine a school's overall performance. In other words, each component should have *some* influence in determining which schools become CSI schools. Even though the weights assigned to measures and components within Ohio's system provide some information about their expected influence, other factors may also affect how much influence specific measures or components have on school performance ratings. Components may be more influential in determining school performance if they are distinct from other measures or are more frequently rated.

Simulations Dropping Individual Performance Indicators

In Ohio, the progress, gap closing, and K–3 literacy components played a larger role than achievement did in determining which schools were identified as CSI. Taking all schools into account, progress and gap closing had the greatest influence on CSI designation. In a simulation in which we removed the gap closing component, 18% of current CSI schools would have been reclassified as no longer CSI, and an equivalent percentage would have been newly identified (Exhibit 13). Surprisingly, K–3 literacy also had stronger influence on CSI designation compared with achievement even though only elementary schools are rated on K–3 literacy.

Individual components also tended to have more influence on CSI designation at the elementary and middle school levels than at the high school level. This result was likely due to the higher rates of correlations among components at the high school level and the higher number of components, on average, at the high school level.

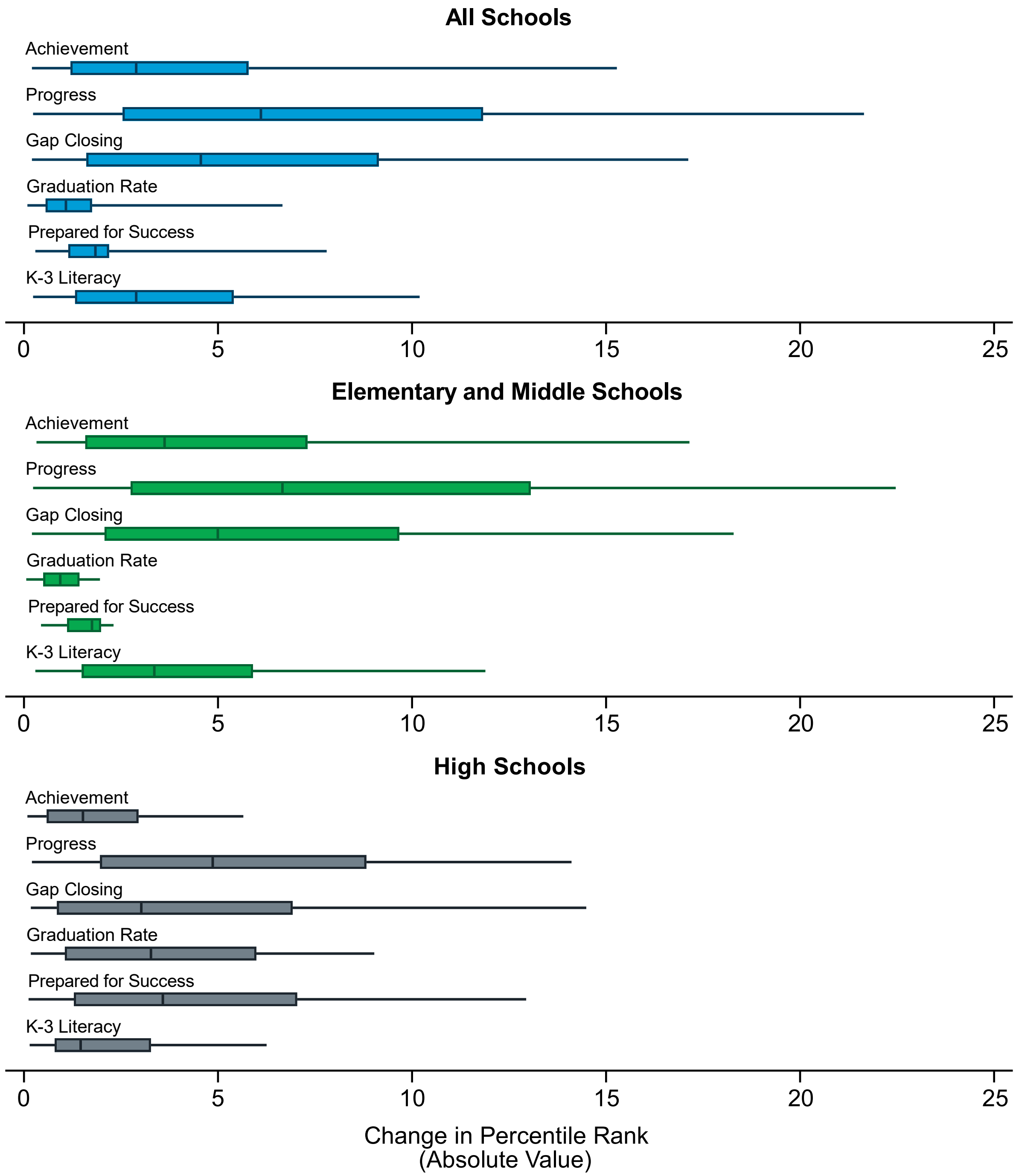
Exhibit 13. Percentage of Schools That Would Be Reclassified if Certain Components Were Dropped



Note. Percentages are calculated based on the number of actual schools below the CSI cutoff. Across all schools 242 schools were below the CSI cutoff, 104 of which were elementary or middle schools and 138 of which were high schools. Most high schools below the cutoff were identified based on graduation rate rather than on performance. Therefore, even though we are modeling reclassification of schools based on the cutoff, many high schools would remain CSI based on graduation rate. Ach. = achievement; Grad. = graduation rate; K-3 Lit. = K–3 literacy; PFS = prepared for success; Prog. = progress.

For all schools and by grade level, the progress component had the most influence on schools' rankings (Exhibit 14). The progress component had the highest median change in ranking (6 percentile points) compared with other components. This finding was true at all school levels. A quarter of schools had percentile changes of approximately 12 percentile points or more when the progress component was dropped. Percentile changes of these magnitudes are large and substantively meaningful when placed in the context of CSI identification where the bottom 5% of schools are identified. The strong influence of progress on performance rankings likely reflects its uniqueness compared with other measures and its stronger weighting. In contrast with the progress component, achievement had a median change in percentile rank of less than 3 percentile points for all schools and was among the least influential measures at the high school level.

Exhibit 14. Change in School Rankings if Certain Components Were Dropped



Note. The thin horizontal line represents the 5th to 95th percentile range. The horizontal box represents the 25th to 75th percentile range. The vertical line within the box represents the median (50th percentile). The number of observations is 3,423 for all schools, 2,561 for elementary and middle schools, and 862 for high schools.

RQ5: Influence of Ohio's Point System

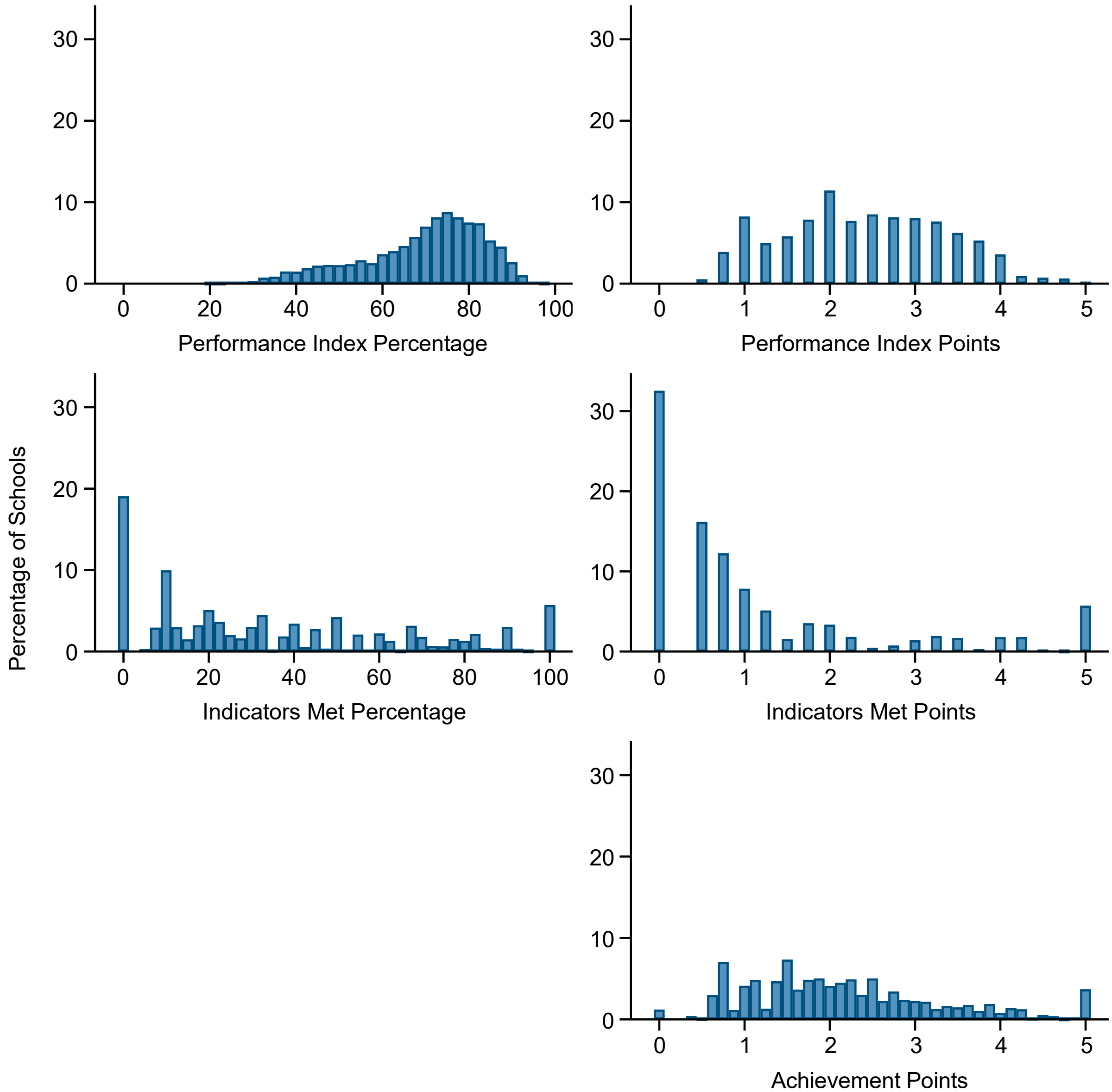
Adding further complexity to its accountability system, Ohio assigns points to measures in increments defined by performance ranges in the underlying metric. Converting various measures into a common metric is necessary to weight and average different measures. But how this conversion is done may influence the schools identified. Here we explicitly tested the influence of individual components and the method of assigning points to measures by conducting simulations that altered the calculation of school performance. We then identified the number of schools that changed their CSI status and how schools' rankings changed in the simulations compared with actual designations and rankings.

Simulations Changing How Points Are Assigned to Measures

The method of assigning points created discrete jumps in points around cutoffs between one point range and the next and created distributions that did not always mirror the distribution of performance in the raw measures. As an example, we present the distribution of measures for the achievement component before conversion to points and after conversion to points (Exhibit 15). The achievement component consists of the performance index and indicators met measures. Prior to the conversion to points, the performance index and the indicators met measures are shown as percentages. For the performance index, the distribution had a smooth shape, with schools most frequently scoring in the 70% to 80% range on the measure. After the conversion to points, gaps appeared in the distribution because schools were assigned points at 0.25-point increments. The same process of converting the measure into points applies to the indicators met measure. However, as can be seen, the indicators met measure in its original form also had a good deal of lumpiness (meaning that schools tend to cluster at certain percentages) due to the way the measure is calculated (total indicators met divided by the total possible indicators for which a school is eligible).[16] In particular, spikes occurred in the percentages of schools with specific percentages of indicators met (e.g., at 0%, 10%, and 20%). After the performance points and indicators met points were averaged, with a weight of 0.75 given to the performance index and a weight of 0.25 given to the indicators met measure, the resulting distribution of points had fewer gaps but retained some of the lumpiness or spikes as certain points.

[16] See ODE (2020b) for performance index documentation. See ODE (2020a) for indicators met documentation.

Exhibit 15. Measures Contributing to the Achievement Component—Distributions of Original Measures and Points



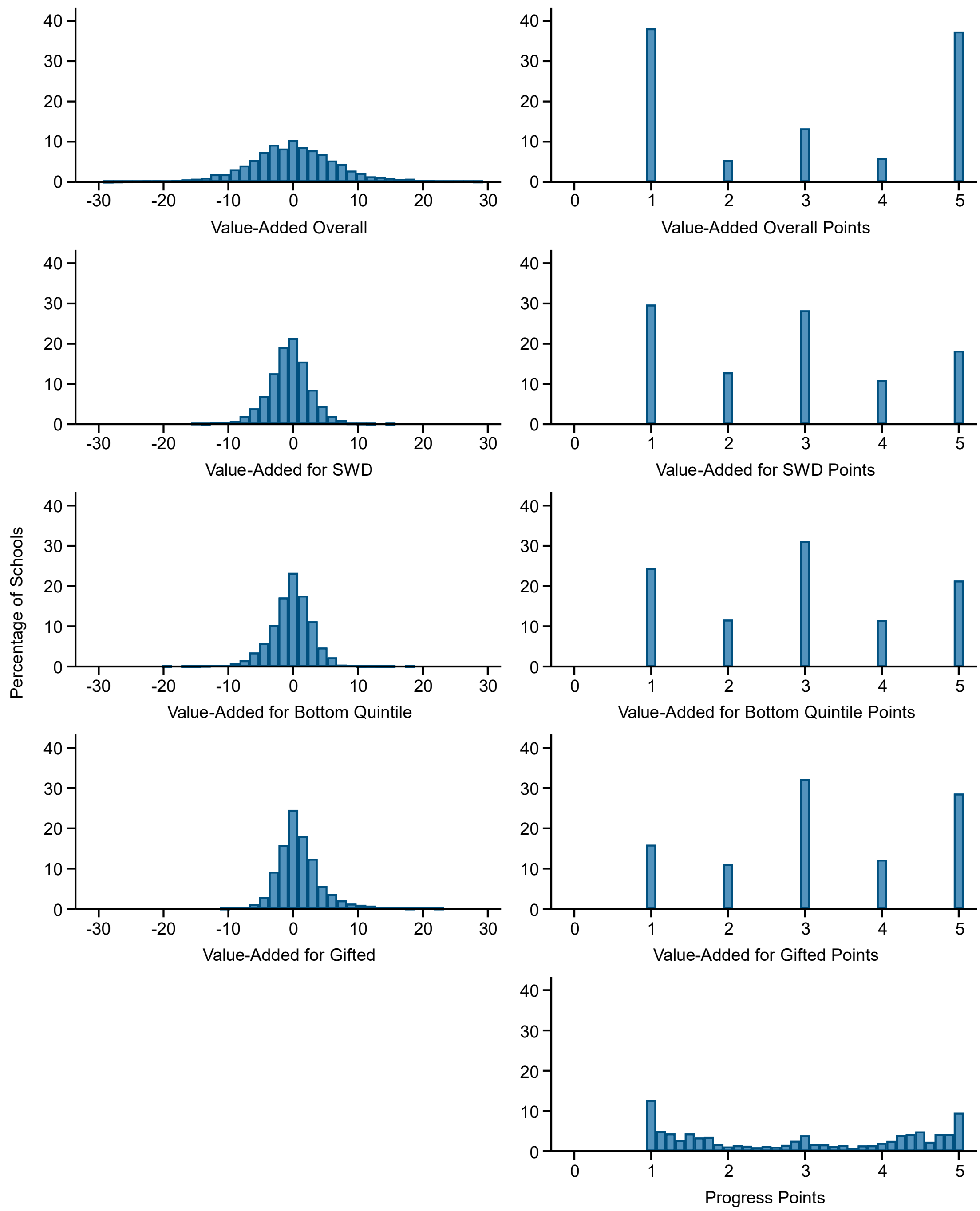
Note. The bars in the histograms of percentages occur at 2.5 percentage point increments. The bars in the histograms of points occur at 0.125-point increments. More than 3,000 schools are represented in each histogram.

As a second example, we show how points are assigned to schools for the progress component (Exhibit 16). The progress component combines school-level, value-added scores across all students, students with disabilities, students who are gifted, and students in the bottom quintile. Rather than assign points to measures in 0.25-point increments, points are assigned in whole-point increments. Although each original progress measure had a smooth distribution of approximately normal shape, the conversion to points assigned most schools 1, 3, or 5 points for each measure and fewer instances of 2 or 4 points. After points for each measure were averaged together into the progress component (with the bulk of weight assigned to the overall value-added measure), the result was a "W"-shaped distribution in which many schools had very low or very high points and a moderate number of schools ended up in the middle with very few schools earning points in between.

The two simulated ways of assigning points appear to be less subjective than the state's own approach. Compared with actual point assignment, Simulation A resulted in a more even distribution of schools across the range of possible point values (Exhibit 17). In other words, the share of schools that have 2 points was approximately equal to the share of schools that have 4 points. Because points were averaged across measures and components, some curvature to the distribution occurred for overall points, but it was generally flat.

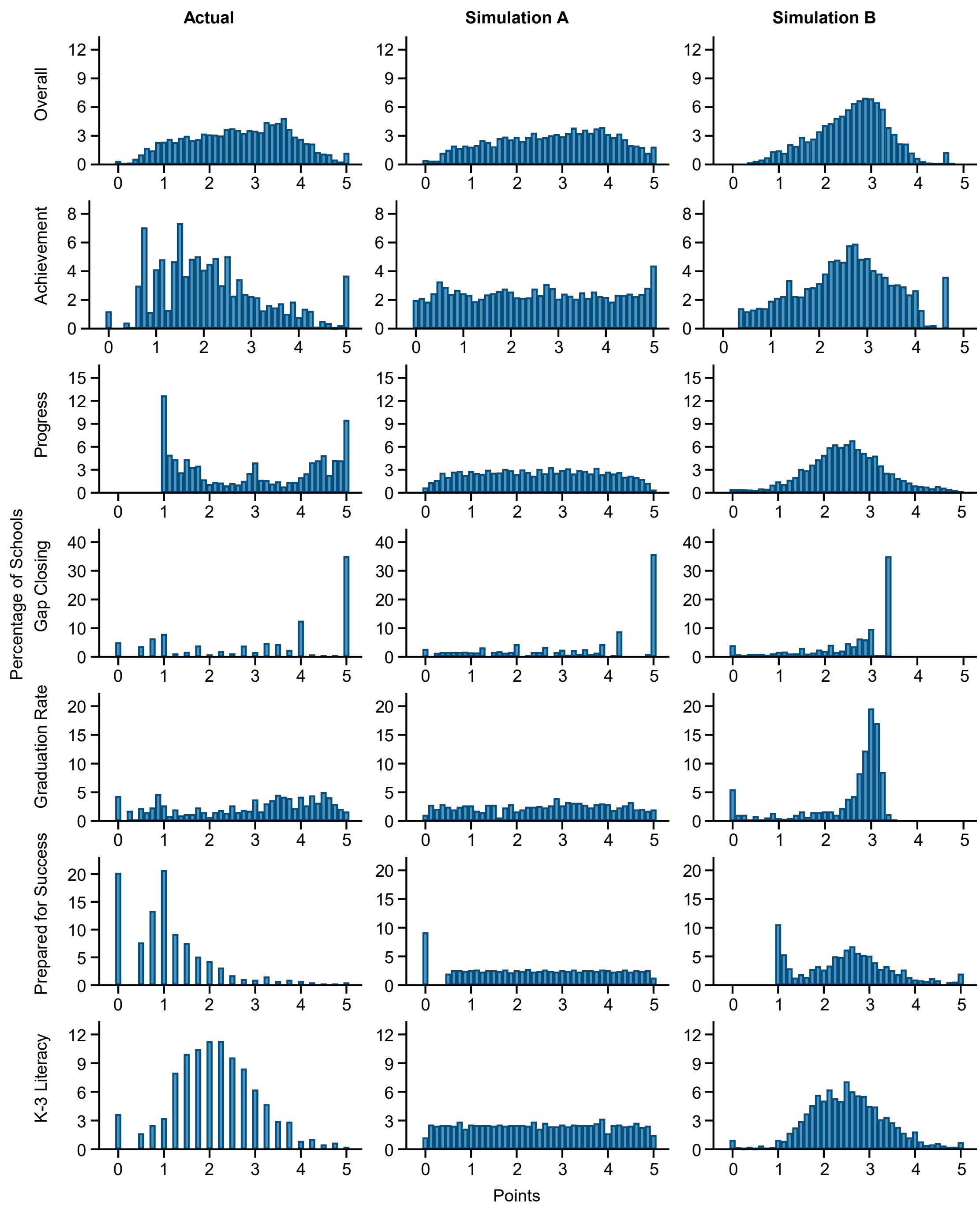
Simulation B, by contrast, had more curve compared with the actual distribution of points. The most common number of points awarded overall was slightly less than 3. As points increased or decreased from 3, the share of schools receiving a given number of points decreased, such that very few schools received fewer than 1 or more than 4 points overall.

Exhibit 16. Measures Contributing to the Progress Component—Distributions of Original Measures and Points



Note. The bars in the histograms of value-added scores occur at 1.5 score increments. The bars in the histograms of points occur at 0.125-point increments. Numbers of schools represented in each histogram range from approximately 1,800 for value-added for gifted to more than 3,000 for overall value added.

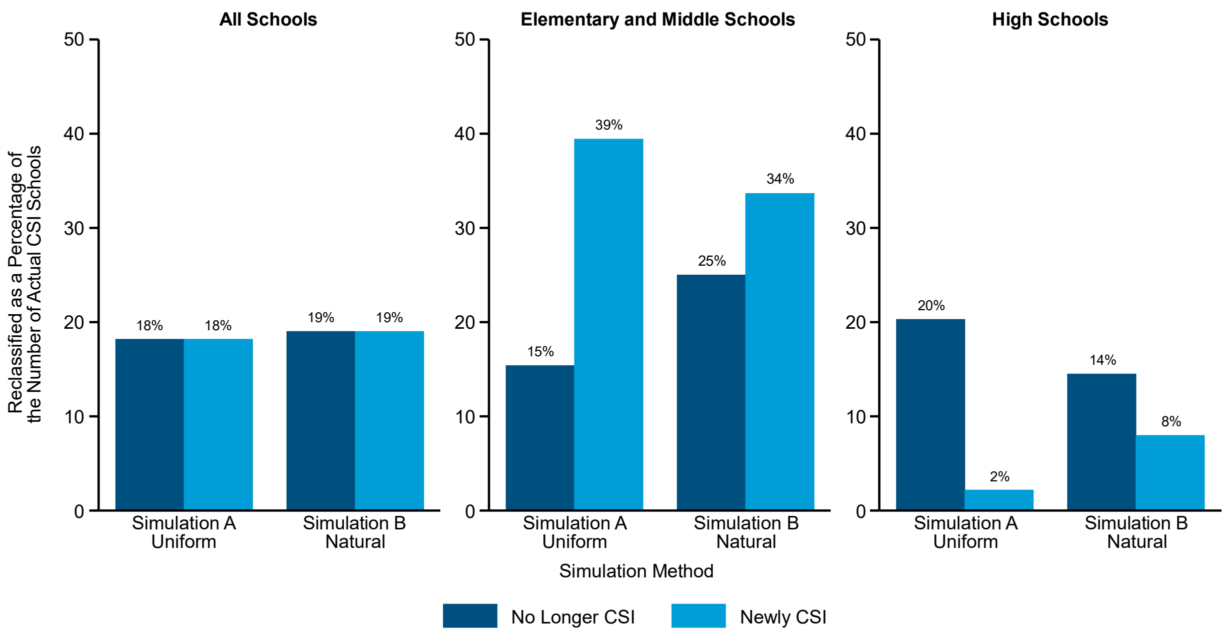
Exhibit 17. Overall Points and Points by Component for the Actual System of Assigning Points, Assignment Points Using a Uniform Distribution (Simulation A), and Assignment of Points Using the Natural Distribution of Measures (Simulation B)



Note. The bars in the histograms occur at 0.125-point increments. Numbers of schools represented in each histogram range from approximately 850 for graduation rate and prepared for success to more than 3,000 for overall, achievement, progress, and gap closing.

The method of assigning points to measures had as much influence on CSI identification as did the most influential component. Using these simulations, we calculated the number of schools that would be reclassified, expressed as a percentage of actual CSI schools (Exhibit 18). For the simulation using a uniform distribution (Simulation A), 18% of actual CSI schools were no longer classified as CSI and an equivalent number became newly CSI under the simulation. The reclassification rate under the natural distribution (Simulation B) was slightly higher at 19%. These rates of reclassification were as high or higher than when progress or gap closing components were dropped—the two most influential components on CSI identification.

Exhibit 18. Percentage of Schools That Would Be Reclassified Assigning Points Using a Uniform Distribution (Simulation A) and Assigning Points Using the Natural Distribution of Measures (Simulation B)

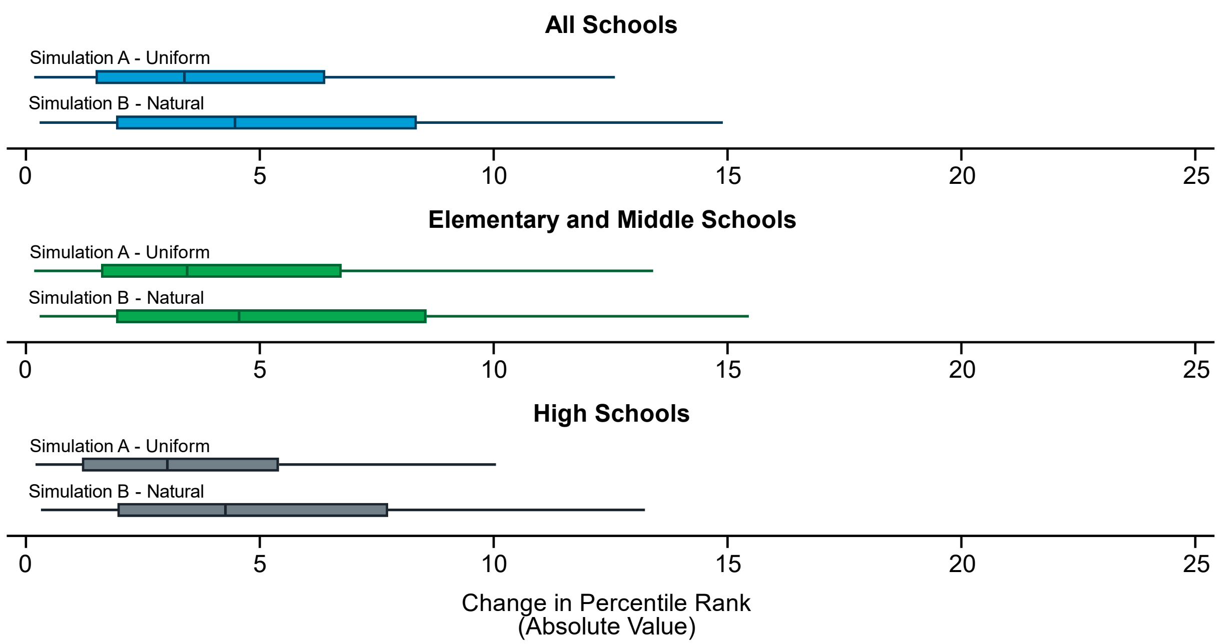


Note. Percentages are calculated based on the number of actual schools below the CSI cutoff. Across all schools, 242 schools were below the CSI cutoff, 104 of which were elementary or middle schools and 138 of which were high schools. Most high schools below the cutoff were identified based on graduation rate rather than on performance. Therefore, even though we are modeling reclassification of schools based on the cutoff, many high schools would remain CSI based on graduation rate.

As with the simulations dropping components, the simulations of alternative point assignments resulted in higher reclassification rates at the elementary and middle school levels compared with high schools. However, patterns by level differed across the two simulations. For both simulations, at the elementary and middle school levels, the numbers of no longer CSI schools were lower than the numbers of newly CSI schools. The opposite was true at the high school level. However, for Simulation A, the pattern was exacerbated with newly CSI schools far exceeding the number of no longer CSI schools at the elementary level and no longer CSI schools far exceeding the number of newly CSI schools at the high school level. For Simulation B, the numbers of no longer and newly CSI schools were more similar within schooling level.

Use of a natural distribution resulted in a greater median change in rank compared with a uniform distribution for assigning points to measures (Exhibit 19). Under the simulation where points were distributed using measures' natural distribution (Simulation B), more schools showed larger differences in their performance rank compared with a uniform distribution (Simulation A). Under Simulation B, the median change of rank was almost 4.5 percentile points compared with 3.4 percentile points under Simulation A. The median change in rank under Simulation B was similar to that under the simulation dropping the gap closing component from the calculation. The only other simulation more influential than Simulation B was the dropping of the progress component.

Exhibit 19. Change in School Rank Assigning Points Using a Uniform Distribution (Simulation A) and Assigning Points Using the Natural Distribution of Measures (Simulation B)



Note. The thin horizontal line represents the 5th to 95th percentile range. The horizontal box represents the 25th to 75th percentile range. The vertical line within the box represents the median (50th percentile). The number of observations is 3,423 for all schools, 2,561 for elementary and middle schools, and 862 for high schools.

Compared with individual components, which had much less influence on high schools, the simulations modeling methods of assigning points had equal influence across school levels. Under the simulations of points, all components were affected simultaneously. Whereas under the simulations dropping components, having more components at the high school level and having components more strongly correlated acted as a buffer to change in rank, this is not the case for the simulations of points. Therefore, for high schools in particular, the method of assigning points has a greater influence on CSI designation than any of the individual components.

Conclusion

In the design of accountability systems under ESSA, each state made choices with regard to which measures to include and how to aggregate those measures to evaluate schools' overall performance. Those choices ultimately influenced how schools have been rated and which schools have been identified as CSI. Despite the flexibility afforded by ESSA—or perhaps because of this flexibility—the law embedded principles to anchor the design of state accountability systems. Specifically, ESSA sought to foster state accountability systems that are comprehensive, reliable, and fair.

Ohio designed an accountability system intended to be comprehensive. Indeed, Ohio evaluates school performance on a wide range of measures embedded within components (see Exhibit 1). On paper, the system should produce a multifaceted assessment of school performance, identifying schools truly in need of comprehensive supports. However, some of our findings suggest this intent may be compromised. First, CSI schools tend to be held accountable for fewer components than are other schools—hence, their identification is based on fewer facets of school performance. Moreover, the complicated method of aggregating individual data points into measures and components potentially obscures the influence of individual measures on CSI designations. For example, even though chronic absenteeism and English learner progress are included as data points, they are aggregated into the indicators met and annual measurable objective measures, along with many other data points, and then further aggregated into components.

Some evidence exists that Ohio's progress measure is reliable and, arguably, more fair to high-poverty schools. NCLB accountability was frequently criticized for reliance on student proficiency levels—which are reflective of students' poverty levels—rather than on students' academic progress, which better reflects a school's contribution to student learning. In Ohio, we found that the progress component was the least correlated to other components, as well as to student demographics. The progress component is also one of the most influential in determining a school's aggregate performance, likely related to its uniqueness and its strong weight. Progress also proved to have good year-to-year stability, meaning its measurement is not overly error prone or random.

One feature of Ohio's accountability system merits further consideration by policymakers: Ohio's method of assigning points based on discrete point values, rather than on a more continuous method, strongly influences schools' performance rankings and CSI status. In fact, our analyses show that the decisions on how to translate performance levels into numbers of points received by schools can have just as much influence as the measures themselves. Hence, for some schools, a technical feature of the system, rather than actual differences in school performance, determine whether or not they are identified as CSI.

References

Di Carlo, M. (2019). *The false choice of growth versus proficiency.* Albert Shanker Institute.

English, D. (2017). Proposed state accountability systems under the Every Student Succeeds Act: A summary of fall 2017 submissions. American Institutes for Research (AIR). https://www.air.org/sites/default/files/Summary%20of%20State%20Accountability%20Snapshots_rev-Dec-2017.pdf

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Harris, D. N., & Liu, L. (2018). *What gets measured gets done: Multiple measures, value-added, and the next generation of accountability under ESSA.* Education Research Alliance. https://educationresearchalliancenola.org/files/publications/052518-Harris-Liu-What-Gets-Measured-Gets-Done-Technical-Paper.pdf

McEachin, A., & Polikoff, M. (2012). We are the 5%: Which schools would be held accountable under a proposed revision of the Elementary and Secondary Education Act? *Educational Researcher, 41*(7), 243–251.

Ohio Department of Education. (2020a). *2019-2020 indicators table*. https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Achievement-Measure/Indicators-Table.pdf.aspx

Ohio Department of Education. (2020b). *Statewide indicators—percent proficient—(by Subject)*. https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Achievement-Component/Technical-Documentation-PI-Score.pdf.aspx?lang=en-US

Wright, B. L., & Petrilli, M. J. (2017). *Rating the ratings: An analysis of the 51 ESSA accountability plans*. Fordham Institute.

About the American Institutes for Research®

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of education, health, and the workforce. AIR's work is driven by its mission to generate and use rigorous evidence that contributes to a better, more equitable world. With headquarters in Arlington, Virginia, AIR has offices across the U.S. and abroad. For more information, visit AIR.ORG.

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

20936_03/23