Identification of Comprehensive Support and Improvement Schools in Florida

Influence of Selected Measures and System Design

Drew Atchison
Damon Blair
Katie Hyland
Umut Ozek
Kerstin Le Floch

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 IES or the U.S. Department of Education.

DECEMBER 2023

Contents

Identification of Comprehensive Support and Improvement Schools in Florida 1

Influence of Selected Measures and System Design 1

Executive Summary 1

Key Findings 1

Background and Policy Context 3

Overview of Florida's Accountability System 4

Objectives and Research Questions 6

RQ1: Which components, as defined and measured within Florida's accountability system, are most commonly used to evaluate school performance? 7

RQ2: How are Florida's accountability components related to each other and to school demographics? 7

RQ3: To what extent are ratings on components consistent from year to year? 7

RQ4: What is the influence of individual accountability components on schools' CSI designation under Florida's accountability system? 8

Methodology 8

Data and Sample 8

Analytic Approach 9

Results 10

RQ1: Indicators Evaluating School Performance in Practice 10

RQ2: Correlations Among Accountability Components and Between Components and School Demographics 13

RQ3: Year-to-Year Consistency of Accountability Components 20

RQ4: Influence of Accountability Components 22

Discussion 23

References 26

Appendix A. Additional Detail on Florida's Accountability Components 27

Exhibits

Exhibit 1. Federal Percentage of Points Index, by School Level 5

Exhibit 2. Components Used in Florida's Accountability System 6

Exhibit 3. Percentage of All Schools and CSI Schools by Number of Components Graded..... 11

Exhibit 4. Percentage of All Schools and CSI Schools Missing a Given Component 12

Exhibit 5. Correlations Among FPPI Components for Elementary Schools 14

Exhibit 6. Correlations Among FPPI Components for Middle Schools 14

Exhibit 7. Correlations Among FPPI Components for High Schools 15

Exhibit 8. Variation Explained by Other Components 16

Exhibit 9. Correlations Between Achievement Components and School-Level Student Demographic Percentages 17

Exhibit 10. Distribution of Performance and Average Performance of Schools in the Highest and Lowest Poverty Quartiles..... 19

Exhibit 11. Correlations Between 2018 and 2019 Component Points 21

Exhibit 12. Percentage of Schools That Would Be Reclassified if Certain Components Were Dropped 23

Executive Summary

For the past quarter century, federal law has required states to measure school performance, identify the lowest performing schools, and provide support to identified schools. Earlier accountability systems were criticized for overemphasizing student achievement in reading and mathematics, excluding nonacademic features of school performance, and relying on measures closely associated with student characteristics. However, the Every Student Succeeds Act (ESSA) of 2015 ushered in a new era of accountability, enabling states to design complex systems with multiple measures, with the intention of providing a more holistic picture of a school's overall performance. Based on 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 provides states substantial latitude in selecting, weighting, and aggregating various measures to identify low-performing schools. These state policy decisions have the potential to influence which schools are identified for support. They also have implications for how school and district leaders respond to accountability designations, ultimately changing student outcomes. In this report, we examine how specific design choices in Florida's ESSA accountability system influence the set of schools identified as CSI.

Key Findings

Florida uses many different components to evaluate school performance and identify CSI schools: student achievement in English language arts (ELA), mathematics, social studies, and science; learning gains (academic progress in ELA and mathematics for the entire school and for the lowest performing 25% of students); middle school acceleration and college and career readiness; English learner progress (ELP); and the 4-year graduation rate.[1] In practice, however, many schools in Florida are not rated on all of these accountability components due to not meeting minimum N size for data measurement (e.g., student missing the test). **We found that CSI schools, especially CSI high schools, were far more likely to have missing data than was typical across all schools. Across school levels, ELP was the most likely to be missing from accountability calculations. As a result, CSI schools were often identified based on a partial picture 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 measures are strongly correlated with each other, they do not provide a great deal of new information. In Florida, **we found strong, positive correlations between the achievement and**

[1] Definitions of each component are included in Appendix A.

learning gain components, revealing that they provide similar information. In contrast, the ELP, middle school acceleration, college and career acceleration, and 4-year graduation rate measures 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 to simply reflect the features of the student population. Measures that are strongly correlated with school demographics may be seen as unfair for identifying schools based on the characteristics of the students they serve, rather than on how well they serve their students. **In Florida, although school performance on achievement and overall learning gains were closely associated with the poverty level of enrolled students, the measure of learning gains of the lowest performing 25% of students was only weakly associated with student poverty levels, putting schools on a more even playing field.**

In addition to the fairness of accountability components, the reliability of these measures is an important consideration. Fluctuations in measurement from year to year could signal noise or randomness, while consistency over time may suggest that a 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 Florida measures are stable. Measures of student achievement, college and career readiness, and the 4-year graduation rate were the most highly correlated from year to year.**

Within a multiple-measure 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 more commonly rated. **In Florida, no single measure—when looking across school levels—had substantially more influence in determining CSI status. However, individual elementary and middle school measures had more influence than high school measures. Specifically, the middle school acceleration and social studies achievement measures had great influence in sparing schools from designation.** As a reminder, the middle school acceleration component was the least associated with other accountability 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. Earlier accountability systems were criticized for overemphasizing student achievement in reading and mathematics, excluding nonacademic features of school performance, and relying on measures closely associated with student characteristics. In December 2015, however, 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, English learner progress (ELP), 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 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-measure systems work and, in particular, how design choices influence which schools are identified for CSI. Our study team examined these accountability-related questions in a larger study, through a grant from the National Center for Education Research at the U.S. Department of Education. In this report, we focus on one state: Florida.[2]

Specifically, we examine how design choices regarding the measures included in Florida's ESSA accountability system and the method for aggregating those measures affect which schools are identified as CSI. Drawing on administrative data from the 2017–18 and 2018–19 school years, we conducted several analyses examining how many measures each school was rated on and which measures were most often missing, correlation among measures and between measures and student demographics, year-to-year correlations of measures, and simulations designed to evaluate the influence of specific measures on CSI identification.

[2] For information and publications on other components of the broader study, see our project page: https://www.air.org/project/impact-csi-designation-multiple-measure-essa-accountability-systems.

Overview of Florida's Accountability System

To evaluate school performance and identify CSI schools, Florida uses an index-based system, which combines multiple measures into a single index score to describe the overall performance of a school. This method stands in contrast to "business rules" approaches like those used in California.[3] Under such approaches, performance measures are considered individually and schools are identified as CSI if they meet a certain number of criteria. In both cases, the law specifies that specifies that schools with a graduation rate below 67% are designated for CSI, regardless of other measures.

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

Indicators Versus Components

In this report, we use the terms **indicator** and **components** to describe the broad elements on which schools are rated.

We use the term **indicator** to describe the broad performance categories that each accountability system must include under ESSA (e.g., academic achievement).

Components are the specific measures or data points that contribute to a school's overall performance within a given indicator. In Florida, each ESSA indicator includes one to six components.

Florida's Federal Percent of Points Index (FPPI) is an average score of up to 12 different components. Each component is assigned points that contribute to one of five ESSA performance indicator categories: (a) academic achievement, (b) academic progress, (c) school quality and student success, (d) English learner progress, and, if applicable, (e) graduation rate (Exhibit 1). Because certain measure and indicator combinations apply to specific school levels, very few schools are graded on all measures. At most, a typical elementary school would be graded on 8 measures, a typical middle school would be graded on 10 measures, and a typical high school would be graded on 11 measures. Non-standard schools have varying numbers of measures based on grade range, up to a maximum of 12 measures.[4] Measure scores are not reported if fewer than 10 students are tested. Florida's final FPPI score is effectively a simple average of these non-missing measure scores. The indicators and measures used in Florida's accountability system are shown in Exhibits 1 and 2.

[3] For our analysis of California's accountability system, see https://www.air.org/sites/default/files/2023-07/California-Measures-Report-Final.pdf.

[4] For more detail on how indicators are measured and the underlying data, see Appendix A or the Florida Department of Education's ESSA website: https://www.fldoe.org/academics/essa.stml.

Exhibit 1. Federal Percentage of Points Index, by School Level[5]

School level	ESSA indicator performance category	Components	Designated points
Elementary	Academic achievement	Achievement ELA Achievement mathematics	100 100
	Academic progress	Learning gains ELA Learning gains mathematics Learning gains ELA – lowest performing 25% of students[a] Learning gains mathematics – lowest performing 25% of students	100 100 100 100
	School quality and student success	Science achievement	100
	English learner progress	English learner progress	100
Middle	Academic achievement	Achievement ELA Achievement mathematics	100 80
	Academic progress	Learning gains ELA Learning gains mathematics Learning gains ELA – lowest performing 25% of students Learning gains mathematics – lowest performing 25% of students	100 100 100 100
	School quality and student success	Science achievement Social studies achievement Middle school acceleration[b]	100 100 120
	English learner progress	English learner progress	100
High	Academic achievement	Achievement ELA Achievement mathematics Learning gains ELA Learning gains mathematics Learning gains ELA – lowest performing 25% of students Learning gains mathematics – lowest performing 25% of students	100 100 100 100 100 100
	School quality and student success	Science achievement Social studies achievement College and career acceleration[c]	100 100 120
	English learner progress	English learner progress	100
	Four-year graduation rate	Four-year graduation rate	80

[a] The lowest performing 25% of students within a school are determined by ranking the prior years' Florida Statewide Assessments (FSA) and Florida Standards Alternate Assessment (FSAA) ELA or mathematics assessments of each eligible student by grade level and identifying students who have a percentile ranking below 25.5% in each grade for each test.

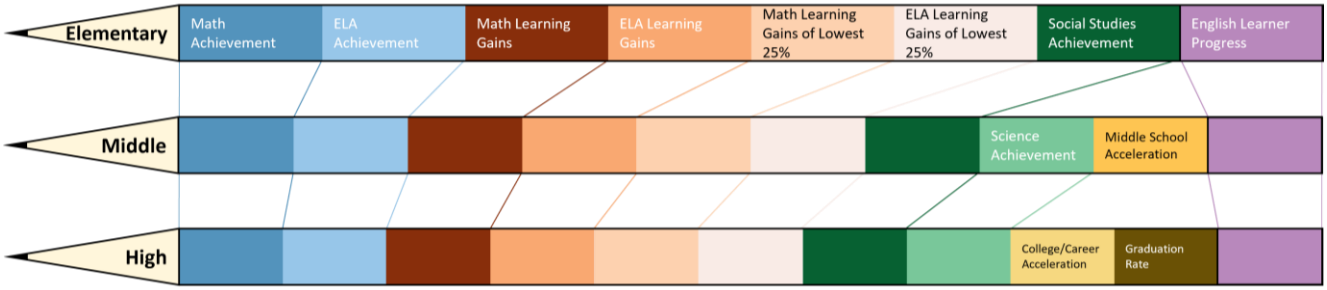
[b] The middle school acceleration component is based on the weighted sum of two components: acceleration success (100 points) and mathematics achievement (20 points). Acceleration success is calculated based on the percentage of students passing high school end-of-course (EOC) assessments or industry certifications.

[5] See Appendix A for further explanation of FPPI calculation

[c] The college and career acceleration component is based on the percentage of a school's graduates who did any of the following: (a) passed an acceleration exam (Advanced Placement, International Baccalaureate, or Advanced International Certificate of Education); (b) passed a dual enrollment course that qualifies for a college credit; (c) earned 300 hours through career dual enrollment courses; or (d) earned an industry certification.

After assessing each school based on the above components, overall FPPI scores are calculated by adding up points for all individual components that the school has data for and dividing that sum by total possible points. For decades, Florida has maintained a state accountability system through which the state applies letter grades based on school performance. Under ESSA, the letter grades are incorporated into CSI designations. These state accountability system measures are nearly identical to the FPPI but omit ELP, which leads to ELP having marginally less influence in determining CSI identification than other measures. Florida has opted to re-designate CSI schools on a yearly basis. The conditions for receiving a CSI designation in Florida are as follows: (a) attain an FPPI score of less than 40%; (b) attain a D or F in Florida's state accountability system (equivalent to scoring less than 40% of points); or (c) attain a graduation rate of less than 67%.

Exhibit 2. Components Used in Florida's Accountability System



Note. Each component is scored on a scale from 0 to 100 and averaged to create a full index score. If all the relevant components are included in the respective calculations, a typical elementary school's components are weighted at 12.5% each, a typical middle school's components are weighted at 10% each, and a typical high school's components are weighted at 9.09% each.

Objectives and Research Questions

This report is part of a broader study examining the underlying theory of action for 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 Florida in designing its ESSA

accountability system and the implications of these choices regarding which schools are identified as CSI schools. In particular, we addressed the following research questions (RQs).

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

Under ESSA, schools are meant to be rated on a variety of school performance measures, providing a more holistic and comprehensive assessment of school performance compared with prior accountability policies, which were largely based on student test performance. However, not all schools receive ratings on all measures. Some measures may not apply to all schools because they are grade-specific (such as graduation rates), and schools only receive a rating for a measure if they have a sufficient number of students contributing to the calculation of performance for that measure. To address this research question, we investigated the number of Florida accountability components for which schools receive ratings and which components are most commonly unrated.

RQ2: How are Florida'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 additional information about school performance for the purpose of designating schools; distinctive measures may yield new insights and influence a school's accountability rating. The inclusion of an 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 school performance ratings (Harris & Liu, 2018). In other words, measures should have more influence if they add unique information about school performance that is not already represented in other accountability measures. For this research question, we conducted an exploratory analysis to determine how distinctive each of Florida's accountability components is from the other components in the accountability system.

School accountability systems are intended to measure school performance, not to simply 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 that need high levels of support. As such, we investigated the extent to which each accountability component is related to school demographics.

RQ3: To what extent are ratings on components 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 a measure, in which case the information

provided by the measure may not be meaningful or reliable. Given that accountability systems are intended to foster school improvement over time, one would not want perfect consistency in school performance as that would mean schools are not improving. On balance, however, we would expect some amount of stability over time, given that academic performance tends to change incrementally. For this research question, we examined whether the components in Florida's accountability system exhibited consistency over time, signifying stability.

RQ4: What is the influence of individual accountability components on schools' CSI designation under Florida'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, policymakers and educators could reasonably question whether the school ratings are indeed capturing multiple aspects of performance. Several of the previous research questions investigate aspects of influence. RQ1 examines which components are most commonly unrated. All things equal, components that are more commonly rated for schools will have more influence. RQ2 examines the uniqueness of components. As explained, components that are more distinctive should have more influence on school ratings and whether or not schools are designated as CSI because they provide new and different information. For this research question, we examined influence more explicitly by analyzing how the exclusion of individual components from the accountability system would change the set of schools designated as CSI.

Methodology

To address the research questions about Florida'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 Florida Department of Education's school report cards, including school-level data on the following:

- School performance on the components of Florida's accountability system (e.g., achievement, learning gains, and acceleration) for the 2018–19 school year
- Accountability designations for the 2018–19 school year[6]

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

Additional data on school demographic composition (e.g., racial/ethnic composition, economic disadvantage, and students with disabilities), total enrollment, and school level (elementary, middle, high, combination[7], and juvenile justice school) for the 2018–19 school year were drawn from the Florida Department of Education.

The schools included in specific analyses varied, depending on the research questions addressed. Most analyses in this report include all public schools in Florida, overall or by school level. Analyses related to RQ1 examine all public schools, as well as CSI schools only. Analyses focus on performance data aggregated to the component level. Note that Florida's data contain a high rate of missingness for some individual components due to not having enough students for which those components apply, particularly for ELP; 38% of schools had missing data in this category.

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 are 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.[8] 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, defined by the percentage of students from economically disadvantaged families[9]. For each accountability component, we examined differences between the two groups of schools in terms of the distribution of number of points assigned.

Correlational Analyses

The study used simple correlational techniques to examine (a) the relationship among accountability component points and between accountability component points and school demographics (RQ2); and (b) year-to-year consistency in 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

[7] Combination schools are schools that serve any grades between 6 and 12 but do not conform to a traditional middle or high school.
[8] For most measures, a rating is calculated if a school has at least 10 accountable students with reportable data.
[9] Economic disadvantage is determined by the percent of students eligible for free lunch or reduced-priced lunch.

predictors of school performance (points earned) on a given component (RQ3). For the regression analysis, we grouped ELA and mathematics components for achievement and learning gains into combined measures by calculating the average of the ELA and mathematics components.

Simulation Modeling

To examine the influence of certain 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 specific components 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 in the simulation but not in actual calculations) and the percentage of schools no longer identified (those identified using actual calculations but not in the simulation)—reflects the degree of influence for a given component. To better understand the collective effects of achievement and learning gain components, this analysis dropped both the ELA and mathematics components at the same time. For example, to examine the effects of achievement collectively, we dropped both the ELA and mathematics achievement components.

Results

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

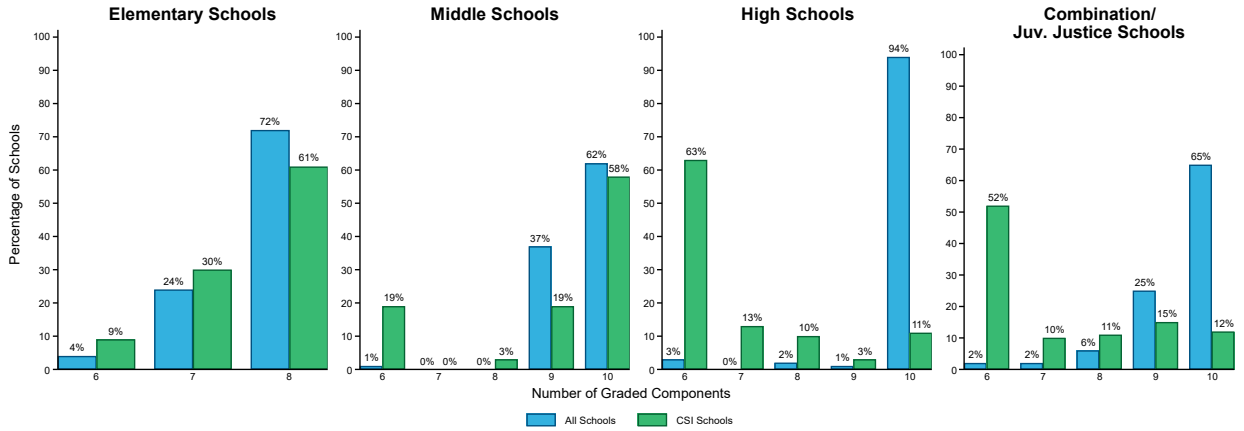
RQ1: Indicators Evaluating School Performance in Practice

Under Florida's accountability system, each school is rated based on up to 12 performance components, which are clustered within five indicators. Given that some components only apply to certain grade levels, typical schools are rated on between eight and 11 components. Here, we present the results of analyses comparing CSI schools to all schools based on (1) the number of components on which schools are rated and (2) the components on which schools most commonly did not receive a rating.

CSI schools often had fewer graded components than was typical across all schools. Among all schools in Florida, being evaluated based on fewer components was associated with being identified as CSI, although this was more pronounced among high schools than elementary or middle schools. As visualized in Exhibit 3, among all elementary and middle schools in Florida, approximately 4% were evaluated on fewer than 7 components (out of 8 and 10, respectively), compared with 10% of CSI elementary and middle schools. This contrast was amplified at the high school level. Across all high schools, combination schools, and juvenile justice schools in Florida, about 17% were evaluated on fewer than 7 components (out of 11). Among CSI high

schools, combination schools, and juvenile justice schools, roughly 61% were evaluated on fewer than 7 components.

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



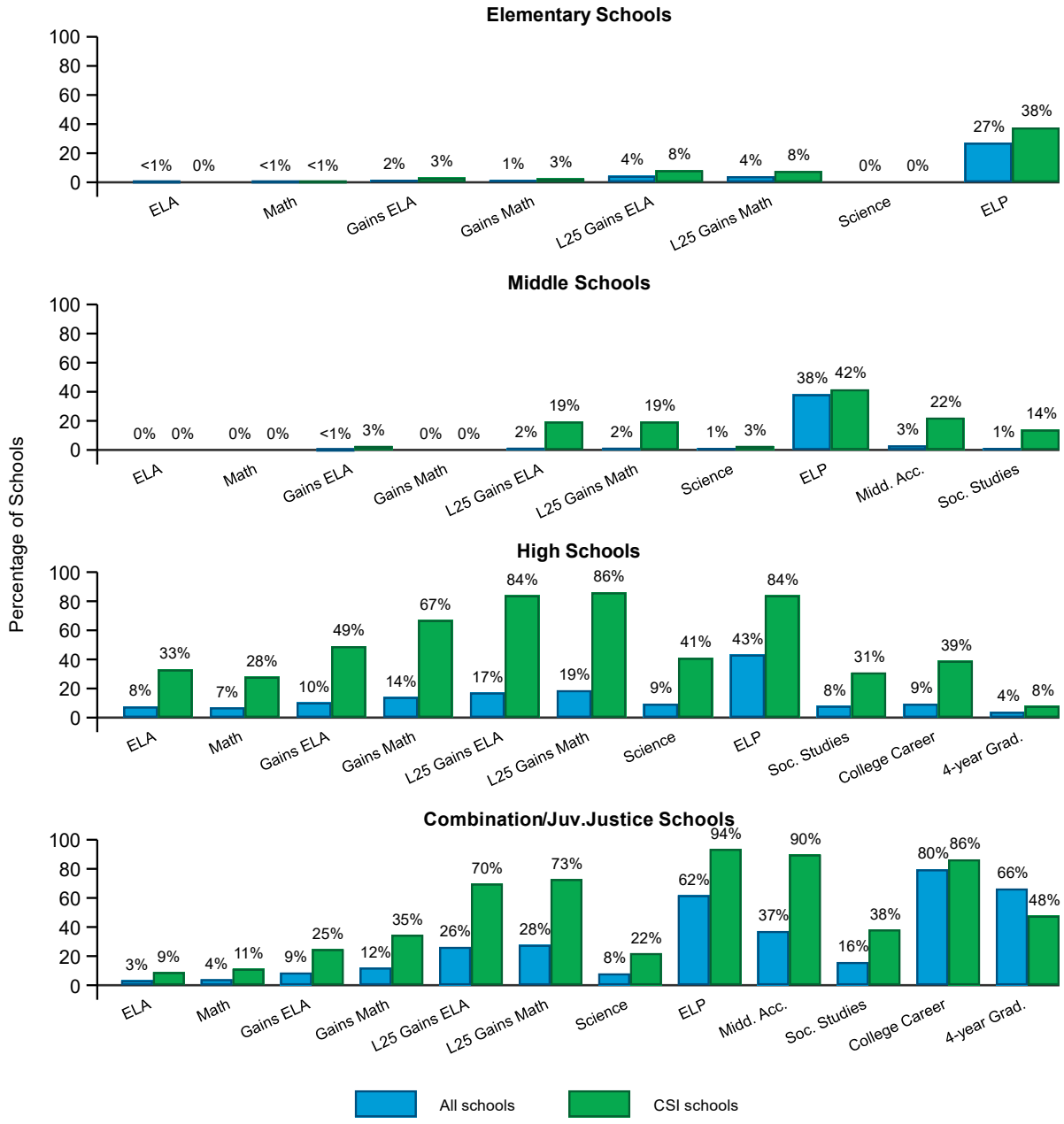
Note. N of all elementary schools = 1,665; *N* of CSI elementary schools = 194; *N* of all middle schools = 542; *N* of CSI middle schools = 36; *N* of all high schools = 459; *N* of CSI high schools = 100; *N* of all combination/juvenile justice schools = 431; *N* of CSI combination/juvenile justice schools = 187. CSI schools only include those identified based on performance rather than graduation rate. Schools that used fewer than six indicators were aggregated into the "6" components category. Schools that used more than 10 indicators were aggregated into the "10" components category.

ELP was the most common component to be missing from accountability calculations across school levels. Although the ELP component was often ungraded among elementary and middle schools, the rates were even more drastic among high schools and juvenile justice schools, which were rarely graded for ELP (94% and 84% ungraded, respectively). ELP was the only component that was significantly ungraded among elementary schools (37.6% of schools ungraded). When compared with all schools, CSI schools were consistently less likely to be rated for the ELP component, with differentials up to 41 percentage points among high school designation types. In contrast, Florida high schools were least likely to have missing data on 4-year graduation rates (4% for all schools and 8% for CSI schools).

Across the board, CSI high schools were far more likely to have ungraded components than was typical across all schools. With the exception of 4-year graduation rates for juvenile justice schools, all CSI schools had more ungraded components than was typical across all schools. This difference was particularly acute at the high school level, as seen in the difference between all high schools and all CSI high schools graded on learning gains in mathematics and ELA for the lowest performing 25% of students (67.2 percentage points and 66.8 percentage points, respectively). CSI juvenile justice schools also had a high rate of unrated components compared with all schools. While CSI middle schools had far less measurement disparity between CSI

schools and all schools, four components had a large gap: learning gains for the lowest performing 25% of students in ELA and mathematics, middle school acceleration, and social studies achievement. In contrast, the measurement disparities in elementary schools were quite low, with ELP recording the largest differential (11 percentage points).

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



Note. *N* of all elementary schools = 1,859; *N* of CSI elementary schools = 194; *N* of all middle schools = 578; *N* of CSI middle schools = 36; *N* of all high schools = 559; *N* of CSI high schools = 100; *N* of all combination/juvenile justice schools = 658; *N* of CSI combination/juvenile justice schools = 222.

RQ2: Correlations Among Accountability Components and Between Components and School Demographics

Under ESSA, school performance is meant to be rated by multiple measures, moving away from the narrower focus of the No Child Left Behind Act (NCLB) 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. Traditional measures of school performance (e.g., proficiency rates) have been closely correlated 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 components and between accountability components and school demographics.

Relationships Between Components

The inclusion of an additional measure will have less influence on a school's CSI status if the additional measure is highly correlated with other measures already used to calculate school performance (Harris & Liu, 2018). We conducted an exploratory analysis to examine the association between each component and other components, and to assess the amount of variation in each component explained by other accountability components. The goal of this analysis was to determine how distinctive each component is from other components.[10] We took two approaches to this analysis: (a) calculate the simple correlations between each pair of components; and (b) identify the proportion of variance of a given component that is explained by other components, using a regression framework.[11]

Across all components, ELP and school-level-specific components had the lowest correlations with other components. In particular, the middle school acceleration component showed very low correlation levels, ranging from 0 (ELP) to .24 (for both mathematics and ELA learning gains and science achievement). Similarly, the correlations between high school college and career acceleration and the 4-year graduation rate and other components were relatively low. Florida's ELP component consistently had the weakest relationships with all other components among elementary, middle, and high schools. This suggests that the ELP component provides unique information about schools relative to other measures. Conversely, achievement components were highly correlated with one another, particularly among mathematics and ELA, but also science and social studies (Exhibits 5–7).

[10] A measure can still change the set of identified schools, even if most of its variation can be explained by other measures, because it changes the relative weights for other measures. Our goal here is to determine the extent to which the variation in new measures can be explained by other measures.

[11] In particular, we regress each measure on all other measures included in the accountability system and obtain the R^2.

Exhibit 5. Correlations Among FPPI Components for Elementary Schools

	Achiev. ELA	Achiev. math	Gains ELA	Gains math	Gains ELA: Lowest 25%	Gains math: Lowest 25%	ELP
Achievement ELA	1.00						
Achievement math	.90	1.00					
Gains ELA	.77	.73	1.00				
Gains math	.56	.73	.63	1.00			
Gains ELA: Lowest 25%	.26	.28	.65	.37	1.00		
Gains math: Lowest 25%	.36	.52	.53	.79	.49	1.00	
ELP	.34	.31	.25	.19	.08	.11	1.00

Note. N = 1,330 elementary schools with all seven components. ELA = English language arts; ELP is English learner progress.

Exhibit 6. Correlations Among FPPI Components for Middle Schools

	Achiev. ELA	Achiev. math	Gains ELA	Gains Math	Gains ELA: Lowest 25%	Gains math: Lowest 25%	Science	Soc. studies	Middle school acceleration	ELP
Achievement ELA	1.00									
Achievement math	.93	1.00								
Gains ELA	.88	.81	1.00							
Gains math	.74	.87	.74	1.00						
Gains ELA: Lowest 25%	.55	.50	.77	.52	1.00					
Gains math: Lowest 25%	.52	.62	.59	.78	.58	1.00				
Science	.89	.88	.76	.72	.45	.50	1.00			
Social studies	.71	.67	.65	.57	.43	.43	.66	1.00		
Middle school acceleration	.19	.23	.24	.24	.15	.18	.24	.19	1.00	
ELP	.46	.43	.42	.37	.32	.33	.41	.29	.00	1.00

Note. *N* = 549 middle schools with all 10 components. ELA = English language arts; ELP is English learner progress.

Exhibit 7. Correlations Among FPPI Components for High Schools

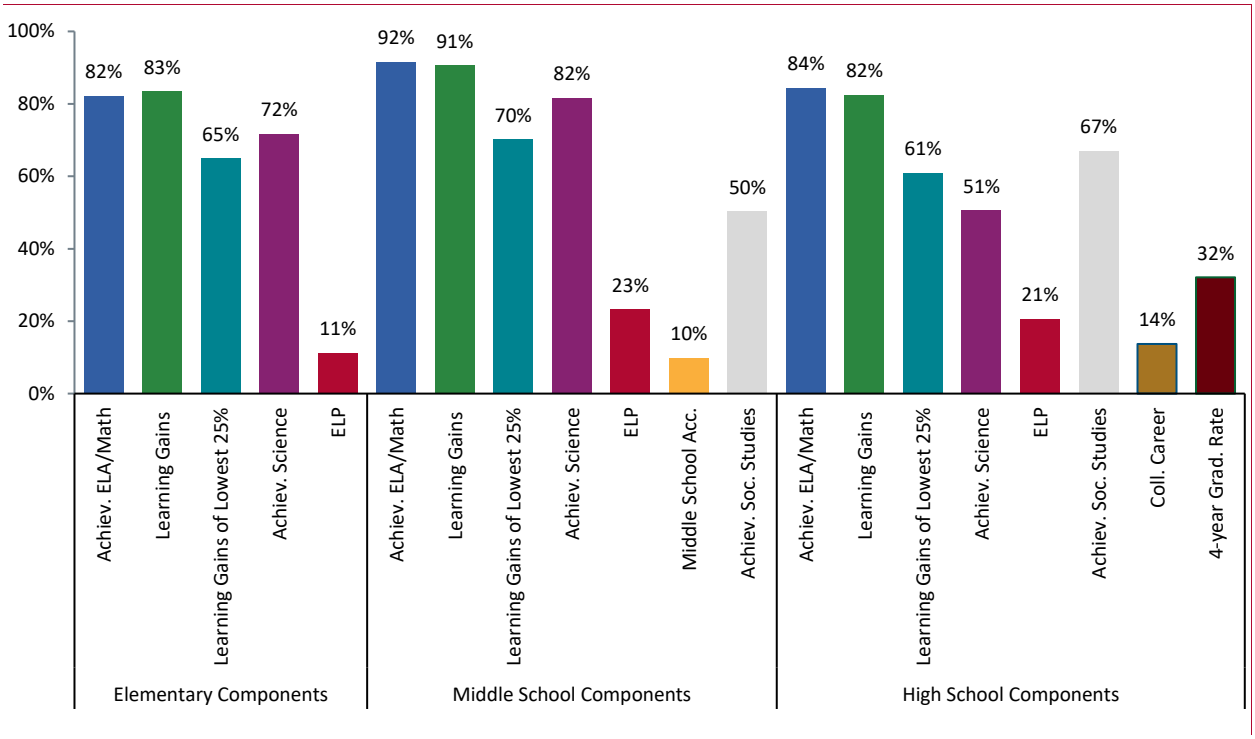
	Achiev. ELA	Achiev. math	Gains ELA	Gains math	Gains ELA: Lowest 25%	Gains math: Lowest 25%	Science	Soc. studies	College and career accelera-tion	Four-year grad rate	ELP
Achievement ELA	1.00										
Achievement math	.83	1.00									
Gains ELA	.84	.69	1.00								
Gains math	.51	.74	.51	1.00							
Gains ELA: Lowest 25%	.56	.51	.77	.41	1.00						
Gains math: Lowest 25%	.32	.48	.38	.73	.37	1.00					
Science	.68	.62	.56	.38	.40	.20	1.00				
Social studies	.81	.75	.66	.50	.46	.27	.66	1.00			
College and career acceleration	.27	.26	.18	.19	.15	.18	.22	.20	1.00		
Four-year graduation rate	.56	.42	.45	.19	.32	.05	.38	.45	.33	1.00	
ELP	.42	.40	.39	.29	.37	.22	.27	.41	.14	.22	1.00

Note. N = 331 high schools with all 11 components. ELA = English language arts; ELP is English learner progress.

Among all performance components included in the FPPI, ELP and school-level-specific components emerged as the most distinctive. To complement the correlational analyses, we accounted for each component's relationships with all the other components in the accountability system simultaneously. Consistent with the individual correlations, we found that ELP consistently had some of the lowest variances explained by other components across elementary, middle, and high schools (11%, 23%, and 21%, respectively). The same was true for middle school acceleration (10% variation), college and career acceleration (14% variation), and the 4-year graduation rate (32% variation). This low variation can be interpreted as these components contributing new and unique information to the FPPI score. However, low correlations could also be a sign that there is randomness associated with the measure. As can be expected, across schooling levels, achievement and learning gains for mathematics and ELA were highly explained by other components. The high percentage of "variance explained" signifies that the component is a valid measure of school performance – in that it is related to

other measures that are also indicative of performance. However, it also means that it does not contribute much new information to the FPPI score calculation.

Exhibit 8. Variation Explained by Other Components



Note. "Variation explained" is the R^2 from a regression where the given component is the outcome variable and the remaining components are the explanatory variables. Achievement, learning gains, and learning gains of the lowest performing 25% of students represent the averages of mathematics and ELA scores for those respective components. ELA = English language arts, ELP = English learner progress.

Relationships Between Components and School Demographics

In designing an accountability system, policymakers should seek a set of measures that does not simply reflect the characteristics of the student body, but rather the contributions of the school to student learning. Here we examine the extent to which each accountability component is correlated with school demographics.

Across the board, the percentages of Black students, students from economically disadvantaged families, and students with disabilities were at least moderately (and negatively) related to nearly all components. In contrast, the percentage of Asian students was positively related to the accountability components. The percentage of Hispanic students was weakly (but positively) related to the accountability components.

Achievement components were moderately correlated with school demographics. Most notably, the percentages of students with disabilities, students from economically disadvantaged families, and Black students were moderately negatively correlated with achievement in ELA, mathematics, science, and social studies (correlations ranging from –.30 to –.53). In contrast, the percentage of Asian students was positively correlated with these measures. The percentage of Hispanic students was almost entirely unrelated to achievement measures.

For Black students and students from economically disadvantaged families, the correlations with learning gain measures were generally weaker than correlations with achievement measures. For example, the correlation between the percentage of Black students and ELA achievement was –.50, whereas the correlation with ELA learning gains was –.33. While somewhat less associated with student demographic groups, the learning gain measures still reflected the composition of each school. The ELP measure and the middle school acceleration measure had weaker correlations with student demographics.

Exhibit 9. Correlations Between Achievement Components and School-Level Student Demographic Percentages

	Students with disabilities	Econ. disad.	English learners	Black	Hispanic	Asian
Achievement ELA	-.44	-.53	-.10	-.50	.06	.30
Achievement mathematics	-.43	-.41	-.03	-.44	.05	.26
Learning gains ELA	-.40	-.34	.07	-.33	.17	.22
Learning gains mathematics	-.33	-.30	.03	-.29	.09	.21
Learning gains ELA – lowest 25%	-.21	-.07	.15	-.04	.22	.09
Learning gains math – lowest 25%	-.19	-.24	-.02	-.15	.09	.15
ELP	-.09	-.22	-.07	-.25	.08	.14
Middle school acceleration	-.07	-.07	.07	-.07	.20	.17
Achievement science	-.38	-.49	-.12	-.49	.02	.28
Achievement social studies	-.49	-.30	-.06	-.40	.08	.18
College and career acceleration	-.42	-.18	-.24	-.36	.02	.28
Four-year graduation rate	-.51	-.22	-.17	-.41	.03	.07

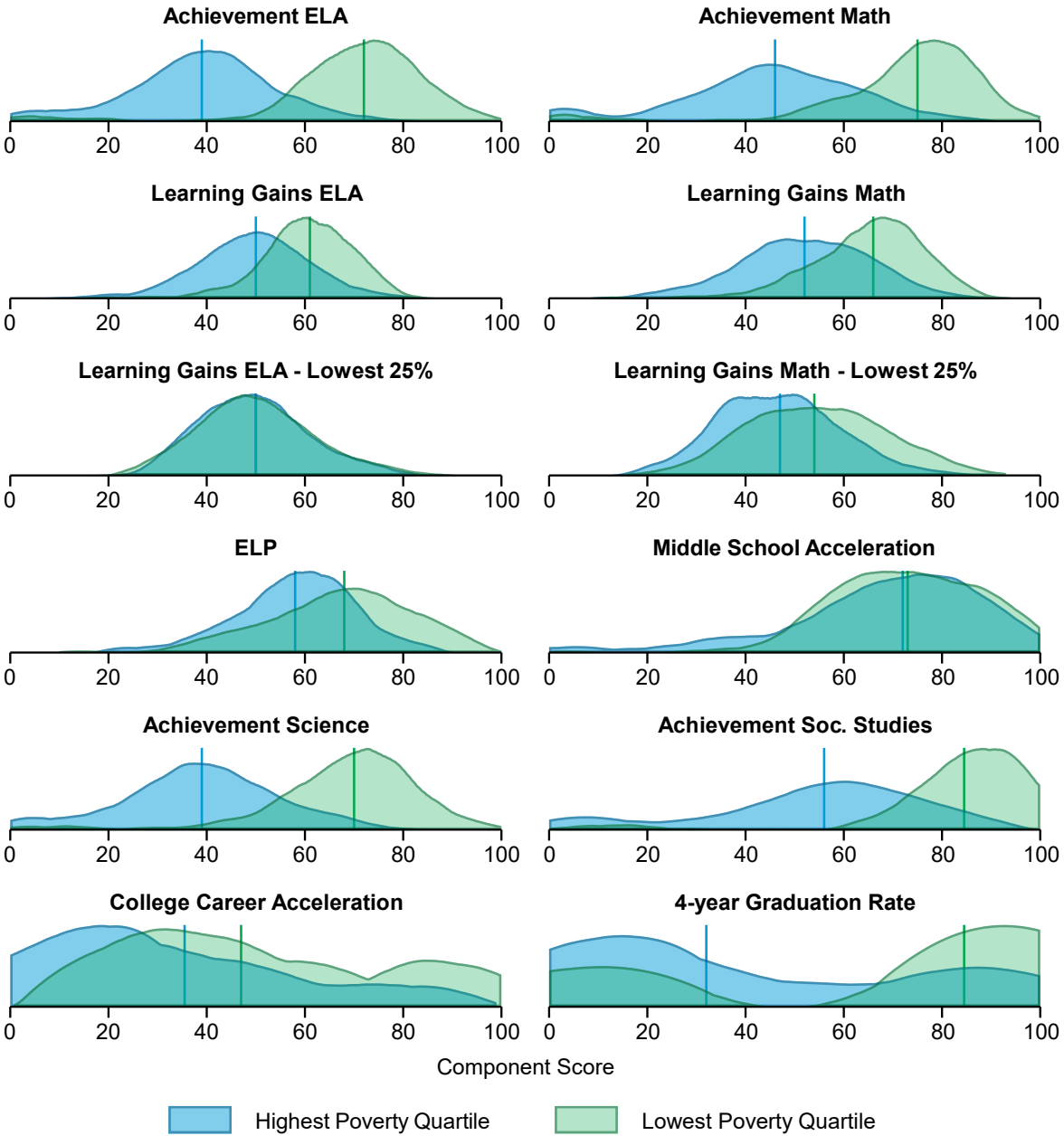
Note. ELA = English language arts, ELP = English learner progress.

To extend the correlational analyses, we examined the distribution of school performance in the lowest and highest poverty quartiles on 10 accountability measures.

For the mathematics and ELA achievement components, almost no high-poverty schools performed at the average of low-poverty schools, and vice versa. As illustrated in Exhibit 10, there was very little overlap in the distributions of these schools' respective performance. For learning gains, and especially learning gains of the lowest performing 25% of students, there was a narrower achievement gap between the highest and lowest poverty quartiles—in fact, the distribution of ELA learning gains of the lowest performing 25% of students was almost indistinguishable distribution between these groups. Similarly, component performance gaps between school-level-specific metrics (besides the 4-year graduation rate) were relatively small. In particular, there was little difference in the average or distribution of performance for high- and low-poverty schools on middle school acceleration.

There was more variation in component scores among the highest poverty quartile than the lowest poverty quartile. Although certainly present in achievement and learning gains, widespread distribution among the highest poverty quartile could be seen in the graduation rate, social studies achievement, and college and career readiness measures. This could indicate that there is greater variation in school quality among high-poverty schools. In other words, although average performance among high-poverty schools was substantially lower than the average performance of low-poverty schools, certain high-poverty schools were defying the trend and producing strong outcomes. These schools could be looked to as exemplars from which other high-poverty schools can learn.

Exhibit 10. Distribution of Performance and Average Performance of Schools in the Highest and Lowest Poverty Quartiles



Note. Vertical lines represent average performance for the highest and lowest poverty quartiles, respectively. ELA = English language arts, ELP = English learner progress.

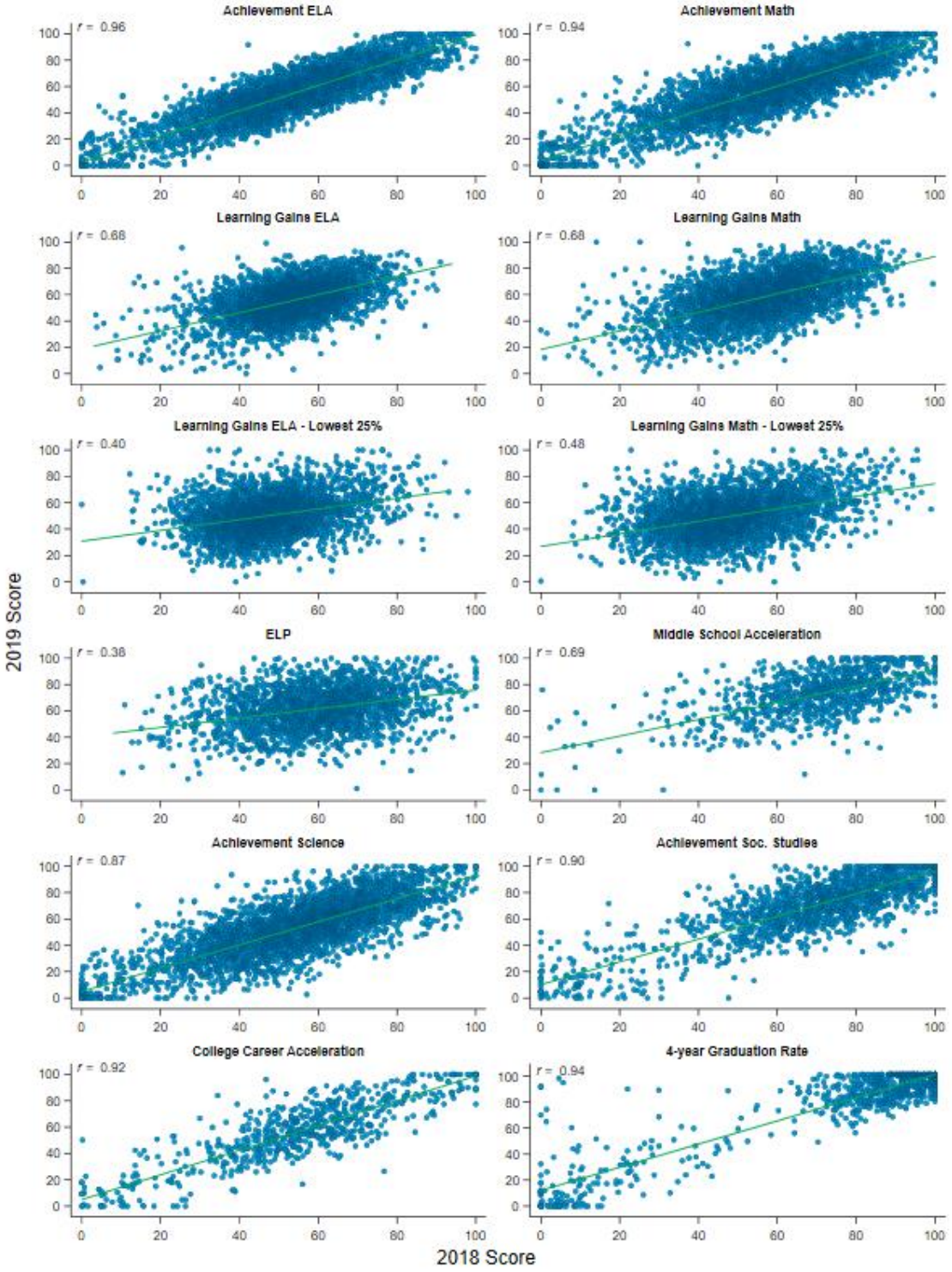
RQ3: Year-to-Year Consistency of Accountability Components

The previous results section on correlations among components and between components and student demographics reported that certain components (achievement and learning gains) were typically more correlated with other components and with student demographics, while other components (ELP and middle school acceleration) tended to be less correlated with other components and student demographics. Low correlations with other components and demographics could indicate that these measures add unique information about school quality and are a fair addition to the accountability system, in that all schools have a chance to perform well regardless of the types of students they serve. However, the weaker associations for these components could also 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 examined year-to-year consistency in component scores across schools. Given that the hope of an accountability system is that schools improve over time, one would not want to see perfect consistency from year to year. School improvement processes also take time, so one would expect somewhat modest changes from one year to the next, as opposed to large fluctuations. Some amount of consistency in ratings from year to year is also a sign of the reliability of the underlying data—that it is measured, collected, and calculated without a high degree of randomness.

All components in Florida's accountability system showed a moderate to strong correlation in performance in consecutive years. The most strongly correlated components between years were the achievement components (.96 for ELA, .94 for mathematics, .9 for social studies, and .87 for science), as depicted in Exhibit 11. The correlations in learning gains for ELA and mathematics were lower but still maintained a moderate correlation between years (.68 for each). Similarly, school-level-specific measures showed strong year-to-year correlations, especially for the 4-year graduation rate and college and career acceleration measures (.94 and .92, respectively). These strong, positive correlations are signals that component scores are reliable. In comparison, we saw lower correlations for ELA and mathematics learning gains for the lowest performing 25% of students and the ELP component (.4, .48, and .38 between years, respectively). These larger fluctuations in performance between years could indicate that there is some amount of measurement error or randomness in these components, making them less reliable. Alternatively, it could mean that these measures are more malleable and responsive to schools' improvement efforts.

Exhibit 11. Correlations Between 2018 and 2019 Component Points



Notes. ELA = English language arts, ELP = English learner progress.

RQ4: Influence of Accountability Components

Within a multiple-measure accountability system, each performance measure should contribute information to determine a school's overall performance. In other words, each measure should have some influence in determining which schools ultimately become CSI schools. While each component is weighted similarly within Florida's accountability system, some may be more influential in determining school performance if they are unique or more commonly rated. We explicitly tested the influence of components by conducting simulations that dropped certain components and recalculated CSI designations to see how many schools changed CSI designation in the simulation.

Simulations Dropping Individual Accountability Components

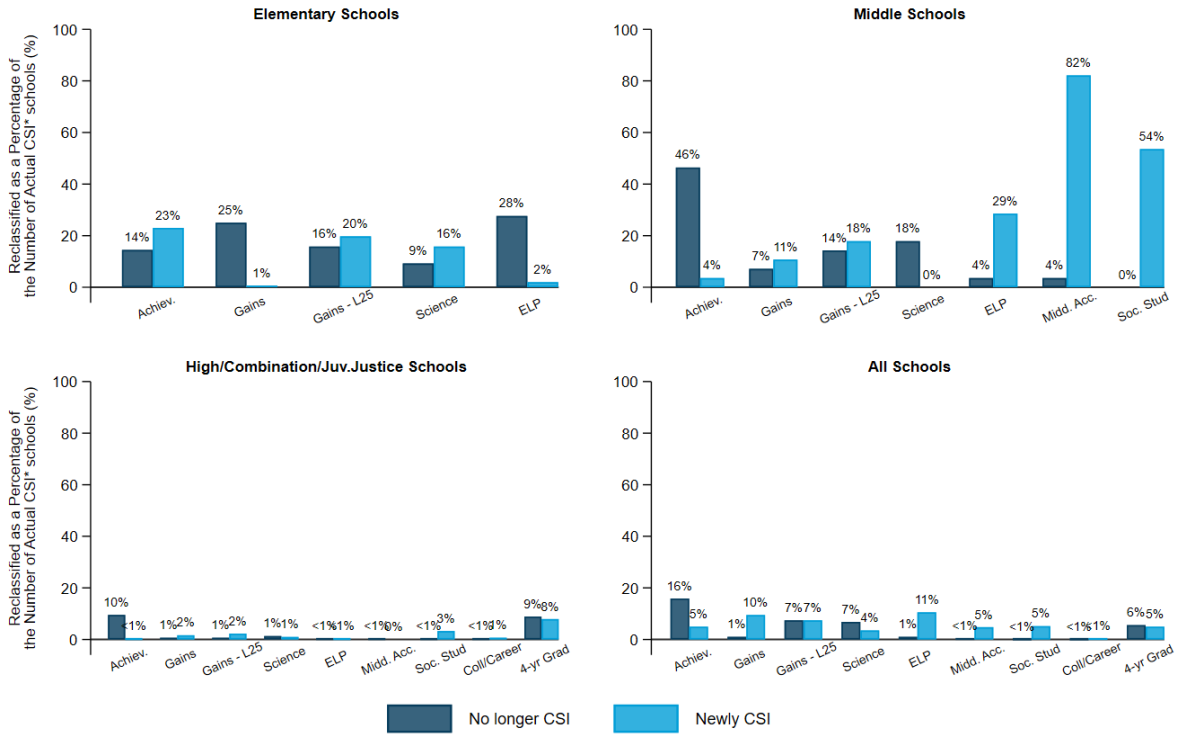
Looking across all grade levels, no single component stood out as being more influential in determining which schools were identified as CSI. Exhibit 12 shows the share of schools that would be classified differently if a given component was dropped from FPPI calculations. For example, if the mathematics and ELA achievement components were dropped from the accountability calculations, 14% of CSI elementary schools would no longer be identified as CSI (represented by the dark blue bars in Exhibit 12). In the same simulation, there would also be a number of newly designated CSI elementary schools, equivalent to 23% of existing CSI elementary schools (represented by the light blue bars). Across all school levels, achievement measures were associated with the largest shifts: If ELA and mathematics achievement measures were dropped, 16% of existing CSI schools would no longer be identified as CSI, while a smaller number of schools (equivalent to 5% of existing CSI schools) would see their FPPI score fall below 40, triggering CSI identification.

The middle school acceleration and social studies achievement components were very influential in shifting CSI identification at the middle school level. Dropping the middle school acceleration measure and social studies achievement measure from the calculation would result in a considerable number of newly designated CSI middle schools, equivalent to 82% and 54% of existing CSI middle schools, respectively. This means that these two components enable a substantial proportion of middle schools to avoid being identified as CSI. Note that only 28 middle schools were initially identified as CSI because of their FPPI score, which represents 5% of all CSI schools. If middle school acceleration was omitted from the FPPI, the number of middle schools identified as CSI based on their FPPI score would increase to just over 50. If social studies achievement was omitted, the number of middle schools identified as CSI based on their FPPI score would increase to over 40.

Dropping accountability components tended to have more influence on CSI designation at the elementary and middle school level. At the elementary and middle school level, each component had a meaningful influence on CSI identification. Conversely, at the high school,

combination school, and juvenile justice school levels, many of the components had a relatively small effect on CSI designation. This is likely due to the presence of more components and the higher rates of correlation among components at the high school level, as discussed previously.

Exhibit 12. Percentage of Schools That Would Be Reclassified if Certain Components Were Dropped



Note. Percentages are calculated based on the number of schools below the CSI cutoff in Florida's Federal Percent of Points Index. Across all schools, 484 were below the CSI cutoff. Of these, 152 were elementary schools; 28 were middle schools; and 304 were high schools, combination schools, or juvenile justice schools. Most high schools below the cutoff were identified based on the graduation rate, rather than just performance. Even though we model reclassification of schools based on the cutoff, this means that many high schools would remain CSI based on their graduation rate. For achievement and learning gains, both the ELA and mathematics components were dropped at the same time because of their conceptual similarity and relatively high correlations between the two. ELP = English learner progress.

Discussion

In designing its accountability system under ESSA, Florida made choices that influence which schools are designated as CSI and therefore receive support. Florida, like many other states, uses an index-based approach which aggregates multiple components into a single score. Florida education officials decided which measures to include in the index, how the index

should differ by school level, how to weight each component, where to place the thresholds for designations, and how often to designate schools.[12] Those choices ultimately influenced how schools have been rated and which schools have been identified for 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 was intended to foster state accountability systems that are comprehensive, reliable, and fair.

Florida's accountability system was created to convey a comprehensive picture of school performance across multiple components. It certainly evaluates school performance across a wide range of components (Exhibit 1) and, on paper, should produce a multifaceted assessment, identifying schools truly in need of comprehensive support. However, some of our findings suggest that this is not always the case in practice. We concluded that CSI schools were often rated on fewer components than would be typical across all schools because fewer data points are used for some components (see Exhibit 3). Although this finding is not unique to Florida (Le Floch et al., 2023; Atchison et al., 2023), the discrepancy means that CSI schools are often rated on a less-than-complete picture of school performance, particularly at the high school level.

Our analyses also indicate that some components have disproportionate influence in school identification, likely because of the amount of unique information they provide relative to other measures. Some components (e.g., ELP and school-level-specific components) provide unique information about school performance and are therefore less influenced by variation in other components. Other components (e.g., mathematics achievement) are highly correlated with others and provide information that is mostly duplicative. The middle school acceleration component is unique from other components (see Exhibit 8) and has a sizable influence in determining which middle schools are identified as CSI (see Exhibit 12).

There is some evidence that Florida's learning gain components are reliable and, arguably, fairer to high-poverty schools than other components (see Exhibit 10). NCLB accountability was frequently criticized for its reliance on student proficiency levels—which are strongly associated with student poverty levels—rather than students' academic progress, which better reflects a school's contribution to student learning. In Florida, we found that the percentages of Black students, students from economically disadvantaged families, and students with disabilities were negatively correlated with nearly all components (see Exhibit 9). This means that most Florida accountability components are doing as much to identify schools with high percentages of these student groups as they are to find schools with truly low performance.

[12] States can opt to re-designate CSI schools on a yearly basis or in 3-year cycles.

In contrast, the correlations between learning gain components (and especially learning gains for the lowest performing 25% of students) and specific student groups were generally weaker than correlations with achievement components, particularly for Black students and students from economically disadvantaged families. Middle school acceleration and college and career acceleration were also less strongly related to school demographics than the achievement components. Including learning gain and acceleration measures rewards schools for improvement and perhaps creates a fairer accountability system by incorporating measures that are less strongly linked to school demographic characteristics.

References

Atchison, D., Hyland, K., Ozek, U., Le Floch, K., & Hurlburt, S. (2023). *Identification of Comprehensive Support and Improvement schools in California: The influence of selected measures and system design.* American Institutes for Research. https://www.air.org/sites/default/files/2023-07/California-Measures-Report-Final.pdf

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

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.* Education Research Alliance. https://educationresearchalliancenola.org/files/publications/052518-Harris-Liu-What-Gets-Measured-Gets-Done-Technical-Paper.pdf

Le Floch, K., Atchison, D., Ozek, U., Hyland, K., Blair, D., & Hurlburt, S. (2023). *Multiple measure accountability under ESSA: Early findings from three states.* American Institutes for Research. https://www.air.org/sites/default/files/2023-04/23-20801-CSI-NCER-ESSA-measures-brief-FMT-ed_rev.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.

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

Appendix A. Additional Detail on Florida's Accountability Components

- The **student achievement** indicator consists of two components: student achievement in **English language arts (ELA)** and student achievement in **mathematics**. Both components represent the percentage of students who achieve grade-level proficiency[13] in their respective subjects, as measured by standardized assessments. Each included component is scored between 0 and 100.[14] The student achievement indicator includes all learning gain components at the high school level, though each individual component retains its original 100 points in terms of total weight.
- The **academic progress** indicator consists of four components, each contributing a maximum of 100 points to the FPPI: learning gains in ELA and mathematics, as well as the learning gains of the lowest performing 25% of students in each subject.
    - Each **learning gain** component represents the percentage of students who achieve a learning gain between years in respective subjects. Students can demonstrate a learning gain by meeting any of the following criteria, based on Florida's proficiency levels (a discrete scale between 1 [lowest] and 5 [highest], each of which contains various subcategories):
        - Maintaining an achievement score of level 5
        - Improving by one or more achievement levels
        - Maintaining an achievement score of level 3 or above and increasing at least one subcategory score
        - If a student remains in levels 1 or 2, increasing a subcategory score within their level
    - The **lowest performing 25% of students**, who contribute 200 (combined) of 400 total points to this indicator, are determined by ranking the prior years' Florida Standards Assessments (FSA) and Florida Standards Alternate Assessment (FSAA) mathematics or ELA assessments of each eligible student by grade level. Students

[13] Florida's proficiency standards are on a scale from 1 (lowest) to 5 (highest). Grade-level proficiency is achieved at level 3 on this scale.

[14] At the middle school level, mathematics achievement is broken into two separate weights: 80 points are assigned to the student achievement indicator and the remaining 20 points are assigned to the middle school acceleration measure (thereby allocating 100 total points for both mathematics achievement and middle school acceleration).

who have a percentile ranking below 25.5% in each grade for each test are identified and the percentage of their learning gains is calculated.

- The **school quality and student success** indicator consists of one to three of the following four components: science achievement, social studies achievement, middle school acceleration, and college and career acceleration.
    - Like other achievement components, **science achievement** and **social studies achievement** are calculated based on the percentage of students who achieve grade-level proficiency on the respective standardized assessments. These components are measured on a scale from 0 to 100.
    - The **middle school acceleration** component, contributing a maximum of 100 points of unique information,[15] is the *percentage of eligible students*[16] who meet either of the following criteria:
        - Achieving proficiency in a high school level Algebra 1, Geometry, Biology 1, or United States History standardized high school end-of-course assessment
        - Attaining an approved industry certification
    - The **college and career acceleration** component is a weighted average of the *graduation rate* (20 points) and the *percentage of graduates who meet any of the following criteria* (100 points):
        - Passed an acceleration exam (Advanced Placement, International Baccalaureate, or Advanced International Certificate of Education)
        - Passed a dual enrollment course that qualifies for a college credit
        - Earned 300 hours through career dual enrollment courses
        - Earned an approved industry certification
- The **English learner progress** indicator consists of a single component, scored between 0 and 100, which represents the percentage of English learner students who meet either of the following criteria:
    - Increased their composite proficiency level by at least one point on ACCESS for ELLs assessments
    - Retained a composite score above 4

[15] As noted earlier, 20 of the 120 total point middle school acceleration measure comes from math achievement.
[16] Students eligible to be included in the denominator are those who (a) are 8th graders who scored at Level 3 or above on mathematics statewide assessment or end-of-course assessment in the prior year; (b) are 6th, 7th, or 8th graders who are enrolled in a high school course with an end-of-course exam and has a valid assessment score for that course; or (c) are 6th, 7th, or 8th graders who took a test for a high school industry certification.

- The **graduation rate** indicator consists of a single component and represents the total percentage of high school seniors who graduate. The graduation rate has an aggregate weight of 100 points in the FPPI, with 80 points assigned to the graduation rate indicator and 20 points assigned to the college and career acceleration measure.

About the American Institutes for Research®

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit institution 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**

23295_12/23