

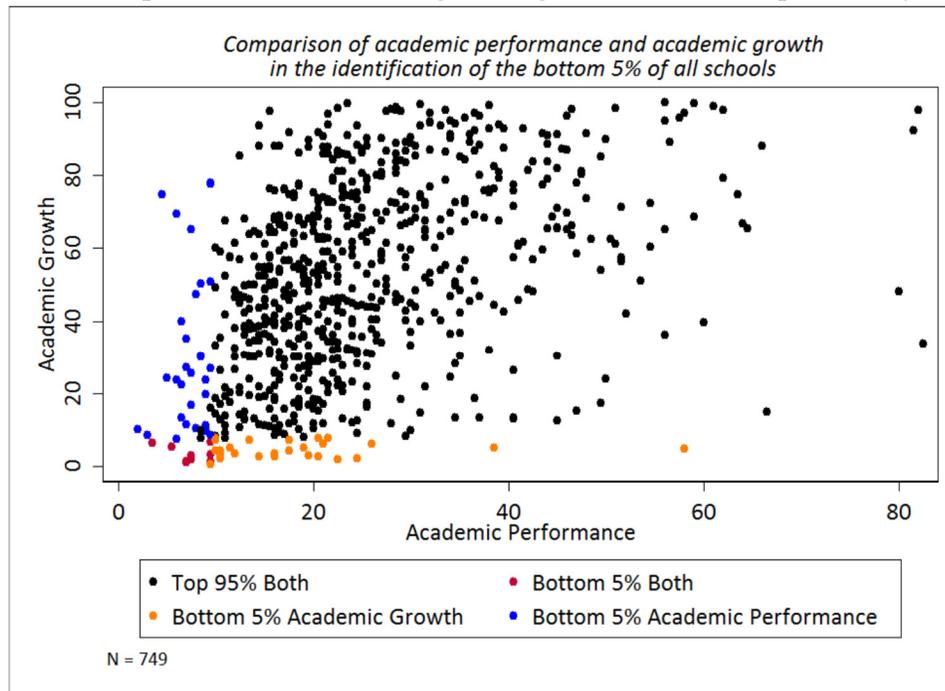
Identity Crisis: Multiple Measures and the Identification of Schools under ESSA*

Summary Submitted to the California State Board of Education on September 8, 2016

The Every Student Succeeds Act (ESSA) requires states to develop an accountability system that includes multiple measures of student academic performance and at least one additional indicator of “School Quality or Student Success” (SQSS). To support policymaking at both the state and federal level, we use the innovative measurement system developed by the CORE Districts to explore one important aspect of ESSA, the identification of schools for support and improvement using a multiple measures framework. CORE’s unique system focuses on academic outcomes alongside non-academic measures of student success, including chronic absenteeism, suspension/expulsion, students’ social-emotional skills, and school climate and culture. Given that most states and districts do not have the full set of ESSA-compliant measures ready for use, there is much to be learned from the CORE Districts about how to integrate such measures into systems of accountability and continuous improvement. We find that:

Different academic indicators measure very different aspects of school performance, and a summative score masks this variation.

ESSA requires that 5-percent of Title I schools be identified in each state for Comprehensive Support and Improvement (CSI). We investigated the extent to which the different academic measures under ESSA (academic performance, academic growth, graduation, and EL proficiency) would identify similar schools if



used independently. We found that schools in the “bottom 5-percent” on any given indicator differed dramatically from measure to measure. For example, the figure to the left shows the comparison of academic performance to academic growth in elementary and middle schools; many schools that have low academic performance are actually demonstrating high growth relative to similar schools. In fact, only 13 percent (or nine schools) are identified among the bottom 5-percent by both measures.

Summarizing the relationships across all pairs of indicators, the table below shows that in each case a low percentage of the bottom 5-percent identified is the same between indicators. When comparing EL proficiency and academic performance, for example, 82 schools are identified as being in the bottom 5-percent of all schools by either EL proficiency or academic performance, but only 4 percent of those 82

* For the full report by Heather Hough (hjhough@stanford.edu), Emily Penner, and Joe Witte, see: <http://www.edpolicyinca.org/publications/identity-crisis-multiple-measures-and-identification-schools-under-essa>

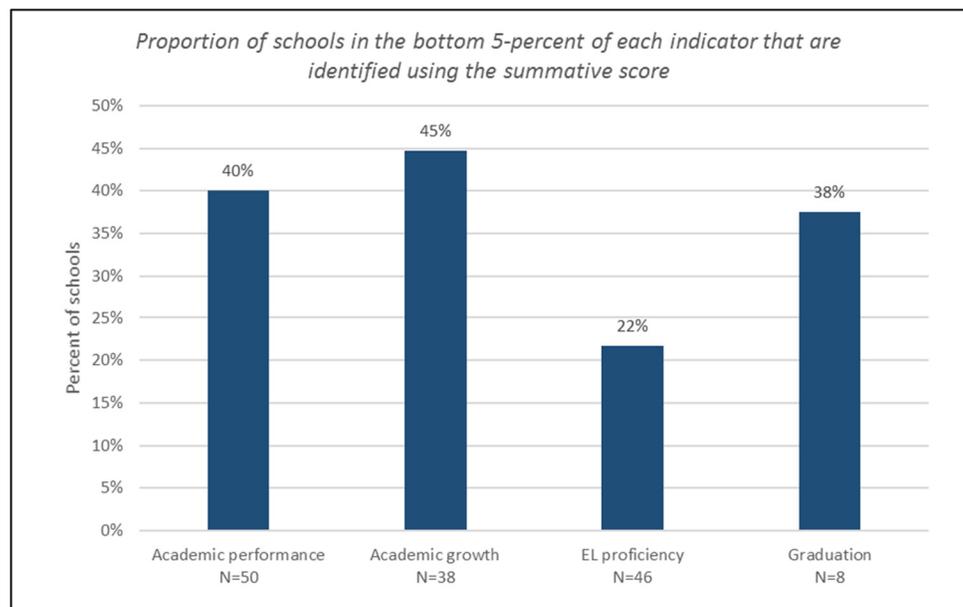
schools that are similarly identified by both measures. These comparisons show that the four academic indicators appear to be measuring very different aspects of school performance, illuminating different dimensions of schools’ strengths and weaknesses.

Percentage of schools similarly identified in pairwise comparisons

		Academic performance	Academic growth	Graduation	EL proficiency
Academic performance	%	100%			
	N	877			
Academic growth	%	13%	100%		
	N	70	749		
Graduation	%	14%	--	100%	
	N	14	--	123	
EL proficiency	%	4%	5%	7%	100%
	N	82	64	15	743

Most schools are high on some things and low on others, which means that a summative score, by definition, averages out this variation. We show that an equally-weighted summative score will identify schools that are low on all indicators, but fail to identify many schools that are very low on individual indicators. Less than 1 percent of schools are in the bottom 5-percent on every indicator, and all of these schools are identified using the summative measure.

However, by aggregating across measures that represent very different dimensions of performance, the summative score may not identify schools that are very low on one measure if they are even average on another. As shown in the figure to the right, of the schools in the bottom 5-percent of all schools on the measure of academic performance, only 40 percent are identified for CSI using the summative measure. Similarly, only 45 percent of the schools in the bottom 5-percent of all schools on academic growth are identified by the summative measure. The corresponding numbers are 22 percent for EL proficiency and 38 percent for graduation.



Which schools should be identified for CSI: those with moderately low levels of performance or those with acutely low levels of performance? This is akin to the dilemma facing a school counselor, who can only provide intensive support to a limited number of students. Should she target her support to a student with all Ds, or to a student with mostly Cs, Ds, and one F? The decision reflects a value judgement, and may depend



on other characteristics of the student. By the same token, the full information in the set of multiple measures is more informative than a single number.

SQSS indicators measure different aspects of school performance than academic indicators and should “count” in the identification of schools.

In addition to the academic indicators, ESSA specifies that states must include at least one indicator of “School Quality or Student Success.” Ahead of the curve, the CORE Districts have already begun collecting a wide range of such indicators locally. The table below shows the percentage of schools that are similarly identified in pairwise comparisons between the four SQSS indicators and the summative academic score. As with the four academic measures, schools are not often low on multiple measures simultaneously, and there is a wide range in how the measures similarly identify schools in the bottom 5-percent compared to one another and compared to the summative academic score. When comparing suspension/expulsion and chronic absence, for example, 90 schools are identified as being in the bottom 5-percent of all schools by either measure, and only 16 percent of those schools are similarly identified by both measures.

Percentage of schools similarly identified in pairwise comparisons

Similarity		Chronic absence	Suspension/ expulsion	Social-emotional skills	School culture/ climate	Summative academic score
Chronic absence	%	100%				
	N	877				
Suspension/expulsion	%	16%	100%			
	N	90	878			
Social-emotional skills	%	10%	8%	100%		
	N	81	80	748		
School culture/climate	%	10%	2%	16%	100%	
	N	87	88	69	836	
Summative academic score	%	11%	6%	9%	12%	100%
	N	91	88	70	76	878

The current draft ESSA regulations specify that non-academic measures cannot remove a school from CSI that would otherwise have been identified using the academic measures. Given how different the non-academic measures are from the summative academic score, this effectively forces states to weight the SQSS indicators so that they do not change the identification of schools using the academic measures. Across each of the four indicators, we found that SQSS measures would have to account for *less than one percent* of the summative measure to not change which schools are identified for CSI. This suggests that a summative score is particularly problematic when considering the inclusion of SQSS measures in states’ accountability systems. If the SQSS indicators are important indicators of school performance, as the law suggests they are, they should be accorded a meaningful weight in the process of identifying schools for support and improvement.

The “tiered approach” provides a transparent method for identifying schools that leverages full information on each indicator.

As an alternative to a summative score, we show how the state can identify schools for CSI using a method that leverages full information on each indicator rather than aggregating the indicators into a summative score. Using a “dashboard” of measures, the state could use a tiered approach to make a series of decisions about school performance on particular indicators. This would enable the state to make judgments about whether or not schools need CSI based on a comprehensive evaluation of all the data. For example, instead

of averaging or differently weighting scores on academic performance and academic growth, the state could decide to identify for CSI only schools that have low academic outcomes and are not demonstrating growth. Similarly, of two schools with similar academic achievement, the state could choose to focus limited resources for CSI on a school with poor SQSS outcomes rather than a school with positive SQSS outcomes, since the latter school may be on a road to improvement while the former is not.

The majority of schools will be identified for Targeted Support and Improvement (TSI) under current regulations, suggesting that the rules for identification should be clarified.

States must identify schools for TSI if the performance of any subgroup falls below the bottom 5-percent level on the summative rating for the “all students” group of Title 1 schools. We find that this method of identifying additional schools for TSI has the potential to identify an enormous number of schools. To estimate the number of schools that would be identified under ESSA’s specified method for TSI, we show which subgroups would be identified using academic performance and graduation rate. As shown in the

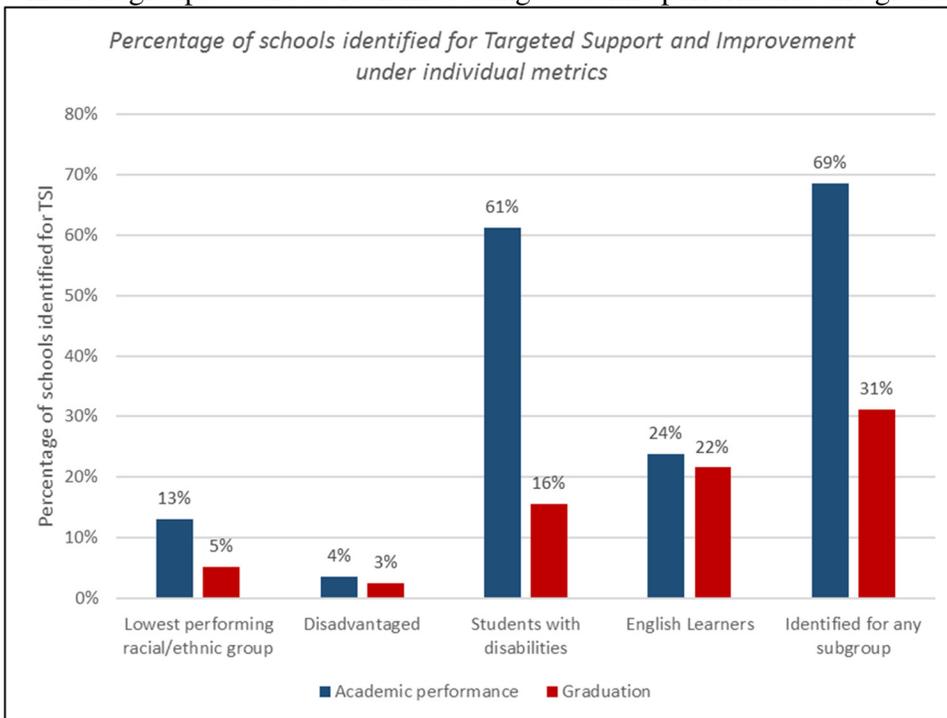


figure to the left, using academic performance, an additional 69 percent of schools (beyond those identified for CSI) would be identified for at least one subgroup. Using graduation rate, an additional 31 percent of high schools would be identified for at least one subgroup. This effect seems to be driven mainly by students with disabilities. In 61 percent of schools, students with disabilities are performing at or below the 5-percent level for all schools on academics, and 15 percent of schools have students with disabilities graduating at the 5-percent level. With infinite

resources, it would be possible, and potentially desirable, to support all such schools. The state should be aware of the magnitude of this identification to prepare accordingly.

Our analyses show that decisions about how to identify schools are highly sensitive to the specific definitions employed, and that these technical decisions reflect value judgements. Multiple measures offer multiple ways for states to identify schools for improvement, and there are tradeoffs between various approaches. Regardless of the method adopted, the state should undertake the kind of empirical tests that we have presented here to better understand the unintended consequences of specific policy choices.