A Proposal for School Accountability under ESSA

Morgan S. Polikoff, University of Southern California

Matthew Duque, Baltimore County Public Schools

Stephani Wrabel, University of Southern California

We are pleased to submit this proposal for redesigned school accountability under ESSA. In the past, when states have been given the opportunity to implement new and creative accountability systems better designed to target the schools most in need of intervention and improvement, they have largely failed to do so (Polikoff, McEeachin, Wrabel, & Duque, 2014). The ESSA again offers states a great deal of flexibility in the design and implementation of school accountability. States should rise to the challenge by designing thoughtful systems that maximize the likely benefits of accountability and minimize the negative unintended consequences. In what follows, we describe the goals of our intended system and lay out a proposal for a system that meets our goals.

**The Goals of the System**

Our proposed accountability system is designed with two main goals in mind. The first goal is to incentivize schools to focus on improving both academic and nonacademic outcomes for all students. A great deal of research shows that consequential accountability can have meaningfully large positive effects on student outcomes (Braun, 2004; Carnoy & Loeb, 2002; Chiang, 2009; Dee & Jacob, 2011; Figlio & Rouse, 2006; Hanushek & Raymond, 2005; Reback, Rockoff, & Schwartz, 2014; Winters, Trivitt, & Greene, 2010); that is, accountability lifts all boats. We do not propose an accountability system focused specifically on the closing of achievement gaps, because there is weaker and less consistent evidence that school accountability has been effective at closing gaps (Gaddis & Lauen, 2014; Hanushek & Raymond, 2005; Harris & Herrington, 2006; Lauen & Gaddis, 2012). Rather, we believe that achievement gaps should be primarily targeted by other interventions (especially interventions that give additional resources to schools serving historically under-served populations to attract high-quality teachers and support them to be effective).

In order to minimize the negative unintended consequences of school accountability, the second goal of our proposed system is for it to be fair to teachers and schools. We have designed a system within the confines of what is allowed under the law that minimizes the extent to which the system punishes schools or teachers for things that are outside their control, such as the demographic composition of their students. In adopting this focus, we seek to reduce the negative incentives that often accompany traditional accountability systems based solely or largely on performance levels—incentives such as encouraging teachers to focus their attention on students just below the proficiency cut (Booher-Jennings, 2005; Neal & Schanzebach, 2010).

**The Design of the System**

Our system is designed to comply with the ESSA statute while seeking to meet our two main goals described above. In this section we describe in detail our measures and plans. All measures in our system are converted to a 0-to-100 scale for easy understanding and aggregation. All measures are also calculated overall and for each numerically significant subgroup, and the overall and subgroup measures are equally weighted in arriving at final scores in each area.

**Indicators of Academic Achievement**

Our indicators of academic achievement include performance in all tested subjects in aggregate, weighted by the number of grades in which they are tested.[[1]](#footnote-1) Performance is measured by conversion of students’ raw scale scores at each grade level to a 0-to-100 scale. This is superior to an approach based on performance levels or proficiency rates in that it rewards increases in performance all along the distribution (rather than just around the cut points). These 0-to-100 performance scores are reported overall and for each numerically significant subgroup (n ≥ 20). The final 0-to-100 index for academic achievement is based on equally weighting the overall 0-to-100 index and a subgroup index that averages the individual subgroup 0-to-100 indices, weighted by the number of students in each subgroup.

**Indicators of Growth**

Our growth measure uses a two-step value-added model.[[2]](#footnote-2) This model is designed to eliminate any relationship between student characteristics and schools’ growth scores. In that sense, it is maximally fair to schools. Our growth measure is otherwise parallel to our performance measure—it uses all tested grades and subjects where growth scores can be calculated and weights growth scores by the number of tested grades and subjects.[[3]](#footnote-3) Again, the growth score is on a 0-to-100 scale and is calculated and reported overall and for each numerically significant subgroup. The process of arriving at a final growth score is the same as for the performance measure—overall scores receive half the weight and an aggregate of subgroup scores receives the other half.

**Indicators of Progress toward English Language Proficiency**

We use two indicators of English Learner (EL) proficiency. First, we measure growth in EL proficiency by the average score increase on the state-selected English Language Learner assessment from the previous year. Second, we measure EL reclassification through a regression-adjusted reclassification rate (where the covariates include student characteristics such as initial EL proficiency). Again we convert both measures to a 0-to-100 score and average them to arrive at the overall EL progress score.

**Indicators of Student Success or School Quality**

Because we seek to promote non-cognitive outcomes in addition to test scores, we have devised a diverse set of additional indicators of student success and opportunity. We believe these indicators incentivize schools to focus on desirable outcomes, even if some are able to be gamed. All of these indicators are placed on a scale of 0-100 and averaged to create an overall rating.

The first indicator is focused on absenteeism. Specifically, we measure both overall attendance rates and chronic absentee rates. It is especially important to measure chronic absenteeism because of its well-documented associations with poor long-term student outcomes (Balfanz, Herzog, & MacIver, 2007; Plank, Farley-Ripple, Durham, & Norman, 2009).

The second indicator is focused on student engagement and happiness, which has been shown to be related to achievement and likelihood of graduating from high school (Fredricks, Blumenfeld, Friedel, & Paris, 2005). This metric is to be measured by a state-administered survey, delivered under controlled conditions to minimize the opportunity for schools to game the measure. Students in all grades for which validated measures of student engagement and happiness have been constructed will be included in this measure.

The third indicator is an equity measure—specifically, we measure disproportionality in discipline by student demographics, including race/ethnicity, socioeconomic status, and special education status.

The fourth indicator gauges the extent to which the school adequately prepares students for success in subsequent grades. This indicator is measured by the on-time promotion rate of students in the next two grades after they leave the present school, overall and by student subgroups.

The fifth indicator is intended to measure student opportunity. In particular, this measure captures the proportion of students who receive a rich, full curriculum. We define a full curriculum as access to the four core subjects plus the arts and physical education each for a minimum amount of time per week. Our goal with this measure is to ensure that schools do not excessively narrow the curriculum at the cost of non-tested subjects and opportunities for enrichment. This indicator will be verified through random audits.

**The Identification of Low-Performing Schools**

**Calculating Summative School Grades**

We propose four school ratings – one each for achievement, growth, ELL proficiency, and a composite of the other five indicators of student success/school quality. Each of the four school ratings is measured on a scale from 0 to 100. This scale provides more differentiation in ratings than an A-F grading scheme. Each rating would be an average of two ratings components - a whole-school measure and a subgroups rating, which itself would be an average of ratings for every significant school subgroup. We do not suggest that the four school ratings be combined in a summative rating, so as not to mask any important variation between them.

The achievement and growth ratings would be the primary determinants to identify schools for accountability, while the ELL proficiency and other indicators would primarily be available to diagnose problems and target interventions in the lowest rated schools (however, any school in the bottom 10% on either the EL or other indicators for consecutive years would also qualify for intervention). Interventions would be determined based on a school’s performance on all four overall ratings and, within each rating, on the numerically significant subgroups with low ratings.

**Defining Low-Performance**

Performance level targets would be set on each of the four school rating dimensions, rather than selecting a fixed percent of schools to be identified as low-performing. Requiring a certain percent of schools to be identified could either identify some schools that perform above acceptable levels or could omit schools that fall below such levels. Performance targets would be reset every five years.

Schools that fall below the whole-school targets would use their relative performance on all the metrics – whole school and subgroup – to identify areas of need to target for improvement. If an insufficient number of schools do not meet their performance targets as required by ESSA, the bottom 5% of schools on each of the achievement and growth indicators would be selected as the “lowest performing.”

**Low-Performing Subgroups**

Under ESSA, states are required to identify schools for interventions based on “consistently underperforming” subgroups. We identify any schools as such that have numerically significant subgroups that fall in the bottom 10% on growth or achievement for two consecutive years.

References

Balfanz, R., Herzog, L., & MacIver, D. (2007). Preventing student disengagement and keeping students on the graduation path in urban m iddle grade schools: Early identification and effective interventions. *Educational Psychologist, 42*(4), 223-235.

Booher-Jennings, J. (2005). Below the bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal, 42*(2), 231-268.

Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Education Policy Analysis Archives, 12*(1), 1-43.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*(24), 305-331.

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics, 93*(9-10), 1045-1057.

Dee, T., & Jacob, B. (2011). The Impact of No Child Left Behind on Student Achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446.

Figlio, D., & Rouse, C. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics, 90*(1-2), 239-255.

Fredricks, J., Blumenfeld, P., Friedel, J., & Paris, A. (2005). School engagement. In K. Moore, & L. Lippman, *Conceptualizing and measuring indicators of positive development: what do children need to flourish?* New York: Kluwer Academic/Plenum Press.

Gaddis, S., & Lauen, D. (2014). School accountability and the black-white test score gap. *Social Science Research, 44*, 15-31.

Hanushek, E., & Raymond, M. (2005). Does school accountability lead to imprvoed student performance? *Journal of Policy Analysis and Management*(24), 297-327.

Harris, D., & Herrington, C. (2006). Accountability, Standards, and the Growing Achievement Gap: Lessons from the Past Half-Century. *American Journal of Education, 112*(2), 209-238.

Lauen, D., & Gaddis, S. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis, 34*(2), 185-208.

Neal, D., & Schanzebach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics, 92*(2), 263-283.

Plank, S., Farley-Ripple, E., Durham, R., & Norman, O. (2009). *First grade and forward: A seven-year examination within the Baltimore City Public School System.* Baltimore MD: Baltimore Education Research Consortium.

Polikoff, M., McEeachin, A., Wrabel, S., & Duque, M. (2014). The waive of the future? School accountability in the waiver era. *Educational Researcher, 43*(1), 45-54.

Reback, R., Rockoff, J., & Schwartz, H. (2014). Under pressure: Job security, resource allocation, and productivity in schools under NCLB. *American Economic Journal: Economic Policy, 6*(3).

Winters, M., Trivitt, J., & Greene, J. (2010). The impact of high-stakes testing on student proficiency in low-stakes subjects: Evidence from Florida's elementary science exam. *Economics of Education Review, 29*(1), 138-146.

School Accountability under ESSA

1. For example, in a K-5 elementary school where mathematics and ELA are tested in grades 3-5, science is tested in grade 4, and social studies is tested in grade 5, mathematics and ELA would be weighted three times that of science and social studies. [↑](#footnote-ref-1)
2. More details on this model can be found at [**http://educationnext.org/choosing-the-right-growth-measure/**](http://educationnext.org/choosing-the-right-growth-measure/) [↑](#footnote-ref-2)
3. For example, if only two growth scores can be calculated in mathematics and two in ELA (fourth and fifth grades), then the growth score would equally weight these four growth scores. [↑](#footnote-ref-3)