B | Economic Studies at BROOKINGS

Evidence Speaks Reports, Vol 1, #24
August 4, 2016

Making use of waivers under ESSA

Susanna Loeb and Heather Hough

Executive Summary

The Every Student Succeeds Act (ESSA) devolves many decisions to states about how to design their accountability system and the measures to use in these systems in order to meet new goals of college and career readiness. Because few states presently have adequate measures for the new goals, they will need to develop the measures along with accountability structures. ESSA includes a provision that would allow district waivers to their state's programs. States can use such waivers to make use of particularly high-capacity districts' ability to innovate and test new approaches. The CORE districts in California serve as a model for the potential for waivers to benefit their state as well as other states. These districts received the only district-level waivers from No Child Left Behind and have designed a new accountability system that incorporates a wide range of measures that could also be used under ESSA. In collaboration with Policy Analysis for California Education (PACE), CORE has produced a series of reports about their experiences with the new measures that are important with respect to future state and federal policy. Among the significant findings are the large impact on which schools are identified as underperforming that follows from decisions about the minimum number of students that must be included to calculate and report on the academic outcomes of subgroups of students, such as African-Americans. In a similar vein is the finding that only 13 percent of schools that are identified as low-performing based on end-of-year test scores in a given year are likewise identified as low-performing when the measure is growth of test scores from the end of one year to the end of the next. Not all states will need district waivers or have districts that could provide the kind of insights that the CORE districts have been able to. But states with both the need and such districts could use the waiver provision to help them to develop and refine an accountability system under ESSA.

The Every Student Succeeds Act (ESSA), replacing the No Child Left Behind Act of 2001 (NCLB), requires each state to establish an accountability system, returning much authority to the states in the design of the systems. As part of this increased state control, ESSA includes a new waiver authority for states. While the waiver provision has received little attention in public discussions, it provides perhaps the best opportunity for states—especially those states with lower capacity education departments—to make the use of their well functioning and most innovative districts to improve the design and implementation of their new accountability system.

The waiver provision (ESSA sec. 8013) allows districts to request of the state, and then the state to request of the Department of Education, a waiver to ESSA provisions, including the accountability provisions. ESSA states that this waiver should be granted as long as the state *provides sufficient information to demonstrate how the waiving of such requirements will advance student academic achievement* and provides plans for *adequate evaluation to ensure review and continuous improvement*.

Why is the waiver provision important? The goals of these new systems are beyond the goals of prior systems. They seek to increase college and career readiness, hold schools accountable for a range of measures beyond the math and English language arts (ELA) test scores emphasized in NCLB, and develop supports for struggling schools with greater flexibility than under NCLB. Even the most knowledgeable and efficient state lawmakers are unlikely to create the best possible system on their first attempt, especially given the limited state of knowledge about how to even measure college and career readiness. Some states have been collecting rich measures of student progress and school functioning and are in position to implement the cycles of development, implementation, learning and improvement that are needed to create more effective accountability systems under ESSA. However, many states lack this capacity and even those with strong capacity may benefit from district-level innovation.

Consider only the measurement needs in ESSA. ESSA requires states to include multiple measures of student academic achievement, including academic performance as measured by ELA and math tests, academic growth, graduation rate, development of English Learner (EL) proficiency, and an additional indicator of "School Quality or Student Success," which can include measures of student engagement, educator engagement, student access to and completion of advanced coursework, post-secondary readiness, or school climate and safety. States are expected to create a summative rating from a set of at least five indicators; ensure that this summative rating yields variation across schools within the state such that schools can be differentiated and identified for comprehensive and targeted support and improvement; and identify underserved subgroups within each school. ESSA requires a more comprehensive approach to measurement than was required under NCLB, with the intention of including more measures and moving away from negative consequences of NCLB's measurement system, such as the narrowing of the curriculum towards tested subjects and content, strategic gaming of accountability structures, and cheating.[i]

Most states do not now have such measures ready for use. In fact, many states have collected only a very limited set of measures on student and school performance. However, quite a few districts have collected rich measures based on direct assessments of students, observations of schools and surveys of staff, students and parents. The CORE Districts[ii] in California are among such districts. Their work illustrates how district innovation can support state policymaking and, particularly, the development of new accountability systems.

In August 2013, the CORE Districts applied for and received a federal waiver to replace the NCLB accountability rules with their own School Quality Improvement System.[iii] This CORE waiver is the only district-level NCLB waiver and serves as a feasible example of district-level waivers under ESSA.[iv] Under the terms of the CORE waiver, the districts developed and are currently implementing an accountability system that focuses on both academic outcomes and non-academic measures of student success, including chronic absenteeism, suspension/expulsion, and students' social-emotional skills, as well school climate and culture. As one of the only systems that includes so many non-academic local measures in their accountability measurement system, the CORE districts provide an opportunity to learn about how such local measures can be integrated into state systems meeting ESSA regulations.

The CORE Districts have created new measures

and structures, studied their properties and effects and then altered them as needed to better meet their goals. As part of this process of improvement, they have partnered with Policy Analysis for California Education (PACE), itself a collaborative network of researchers from across California's universities. PACE has combined and organized administrative data from across the CORE Districts, collected qualitative data on implementation for further insight, and recruited researchers to analyze data and work with district leaders to interpret and build on findings, towards the goal of building capacity for improvement. While not all states have an organization such as PACE to support district work, many states have higher education capacity that could be utilized for such work.
The CORE-PACE Partnership has generated information that informs policymaking in California and across the country. For example, in April, the partnership released a policy brief examining the feasibility of using chronic absenteeism in a multi-metric accountability system, which led to California's decision to consider it as a viable indicator of school quality in the statewide measurement system.[v]

In May, the CORE-PACE Partnership released a policy brief examining the tradeoffs in using various minimum subgroup sizes, i.e., the smallest number of students in a demographic subgroup such as low-income or Asian that is allowable for the purpose of calculating and reporting student outcomes for that subgroup at the school level. When subgroups are small, the average test scores of these groups are likely to have substantial measurement error; however, when the minimum subgroups are large, many students who are members of these targeted subgroups but in schools without many others in their group will not have their scores brought to light. The May report demonstrated the substantial impact of subgroup size on the number of schools reporting subgroup scores and the number of students overlooked. In the CORE districts, six times as many schools would report results for African-American students (see Figure 1) and ten times as many schools would report results for all subgroups if the minimum group size was 20 instead of 100. When the minimum group size is large, the lowest performing racial/ethnic subgroup in the school is often excluded in the measurement. For example, only 37 percent of African American students' test scores are reported at the school-level when the minimum subgroup size is 100, but 88 percent are when the minimum subgroup size is decreased to 20. This effect is even more dramatic for students with disabilities: only 25 percent of students with disabilities are reported at the school-level when the subgroup size is 100, but 92 percent are if the minimum subgroup size is decreased to 20 (see Figure 2). These results indicate dramatic differences. Setting minimum subgroup size is a consequential decision and the report was subsequently referenced in the ESSA regulations as rationale for capping the subgroup size at 30.[vi]

Figure 1. Percent of Schools Reporting Subgroups If the Minimum were 20 vs. 100

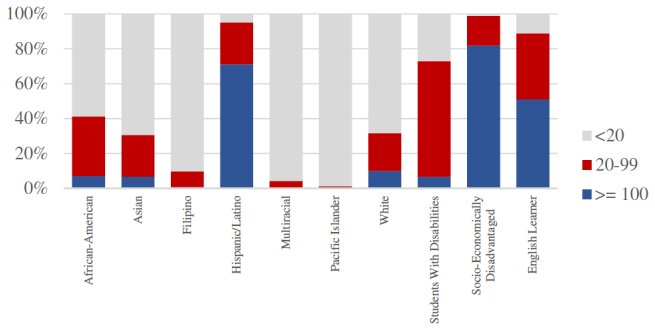
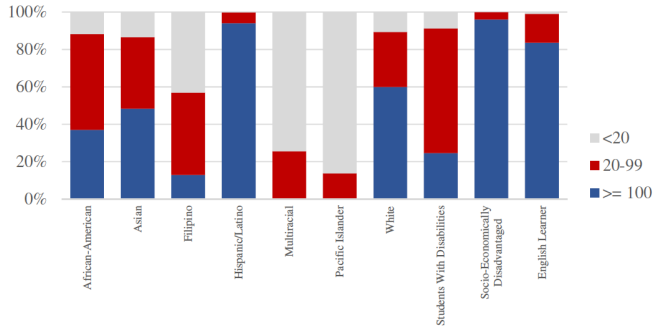


Figure 2. Percent of Students Reported If the Minimum Subgroup Size were 20 vs. 100



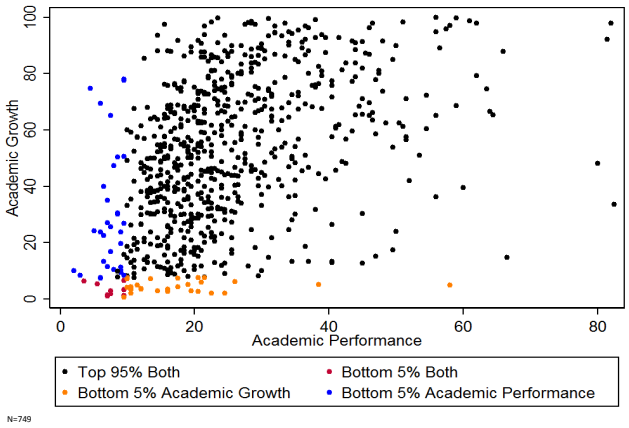
A CORE-PACE policy brief released this week further illustrates the usefulness of the district-level work, not only for the CORE districts but also for California and other states as they seek to develop their new accountability systems.[vii] This new report uses data from CORE's multiple measure system to compare the different methods that could be used to identify schools for support and improvement under ESSA. While the identification of schools for comprehensive and targeted support and improvement is only a small part of how these measures should be used over time, it is an important part, and no less important under ESSA than it was under NCLB, as substantial resources will be allocated to identified schools.[viii]

The expanded measures in ESSA include both "academic measures" and "School Quality and Student

Success" (SQSS) measures. The CORE districts collect measures of academic performance, academic growth, graduation, and English learner proficiency that meet ESSA guidelines for academic measurement. The academic performance measures include the percentage of students testing proficient in ELA and mathematics, based on Smarter Balanced Assessment Consortium test scores. The growth measure is a growth percentile (rank from 0-100) designed to measure the extent to which students in a given school have improved their performance on ELA and math tests from one year to the next relative to similar students demographically who started the school year with similar prior achievement. The graduation measure is the percentage of students who graduate compared with the number of students enrolled in the school (accounting for students who transfer into and out of the school); and the EL proficiency measure captures the percentage of students who are reclassified from English language learner status to "fluent English proficient" status.[ix]

The CORE-PACE report shows that these different academic measures identify dramatically different schools as schools in need of improvement. For example, when comparing academic performance and academic growth (for elementary and middle schools), only 13 percent of the schools identified by one measure would be identified by the other. In Figure 3 below, the red dots represent schools that would be identified in the bottom 5-percent of all schools with both measures, the blue dots represent schools that would only be identified using academic performance, and the orange dots represent schools that would only be identified in the bottom 5-percent of all schools using academic growth. When comparing academic performance and academic growth, 70 schools are identified as being in the bottom 5-percent of all schools by either measure, but only nine schools (13 percent) are identified among the bottom 5-percent by both measures. If the measures were identical, 100 percent of schools would be identified by both measures. This picture becomes even more complex when adding SQSS indicators to the measurement system. The authors show that, as the regulations currently stand, the SQSS measures will have no weight in the identification of schools, which fatally compromises their potential benefit in providing a more nuanced, comprehensive view of school performance.

Figure 3. Comparison of academic performance and academic growth in the identification of the bottom 5-percent of all schools



The CORE-PACE report shows that decisions about how to identify schools are highly sensitive to the specific definitions employed. It describes how different approaches to combining the measures results in the identification of different schools for interventions. These differences point not only to the complexity of measurement, but also to the importance of transparency and full reporting in that schools and districts can use the information on their relative performance on difference measures to better target their efforts towards improvements in the areas that they need most. A single summative score would not provide this information, while a clearly presented set of indicators would. The results provide evidence that could be useful for policy makers determining features of their accountability system.

States are often not in the best position to innovate because of the complexity of policymaking at the state level and, sometimes, the lack of capacity in state departments of education. Some districts have this capacity and can be utilized for the benefit of all districts in the state. The CORE Districts are an example of how innovative groups of districts can advance understanding about what works locally and under what conditions. These districts have both the authority and the ability to innovate and test, to collaborate with researchers, and to adjust approaches in real-time that are necessary to develop accountability systems that help achieve the new goals of college and career readiness for students. The waiver provision in ESSA can allow states to make use of these district capacities by granting expanded authority to districts while maintaining their own key role in school improvement and equalization.

States pursuing this strategy will need to figure out how to accommodate both district innovation and the need for statewide comparability in how the success of individual schools is calculated and reported. After all, the needs of parents deciding where to live based on the quality of the public schools, the responsibilities of those whose goal is to improve school performance, and the information needs of taxpayers and the general public are not well served by individual districts or groups of districts using incommensurate measures to describe how their schools are doing. One solution would be to have a common set of measures and reports for all districts but allow district-specific additions in the form of other measures and ways of reporting. How to best to solve this problem is, like the finer grained questions about minimum group size and the use of growth to identify low performing, an empirical question. The rest of the nation can learn from how the state of California chooses to accommodate its reporting and accountability requirements to the innovations of the CORE districts.

---

[i] For a review see: Loeb, S., & Figlio, D. (2011). School accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), Handbook of the Economics of Education, Vol. 3 (pp. 383-423). San Diego, CA: North Holland.

[ii] The CORE districts include Fresno, Garden Grove, Long Beach, Los Angeles, Oakland, Sacramento, San Francisco, Sanger and Santa Ana Unified School Districts. CORE Districts is a nonprofit organization that supports and advances the work of the member districts.

[iii] In September 2014, the U.S. Department of Education extended the districts' NCLB waiver.

[iv] The CORE Districts implementing the waiver (Fresno, Long Beach, Los Angeles, Oakland, San Francisco, and Santa Ana Unified School Districts) together represent nearly one million students, almost 20 percent of the students served in California, and have a combined 923 schools that are Title 1 schools, which is more than 26 of the 50 states (Analysis based on data from https://nces.ed.gov/pubs2015/2015151/tables/table_03.asp).

[v] http://www.edpolicyinca.org/publications/using-chronic-absence-multi-metric-accountability-system

[vi] http://www.edpolicyinca.org/publications/making-students-visible-comparing-different-student-subgroup-sizes-accountability

[vii] http://edpolicyinca.org/publications/identity-crisis-multiple-measures-and-identification-schools-under-essa

[viii] ESSA stipulates that a district that receives funds for school improvement should receive a minimum of $500,000 for each comprehensive support school it serves and $50,000 for each targeted support school it serves, unless the state determines that a smaller amount is sufficient.

[ix] CORE's measure of EL proficiency is slightly different than what is specified in ESSA. Rather than using only test score results to determine progress on English proficiency, the CORE Districts chose to report reclassification rates, which are a combination of language proficiency scores and academic performance (Carranza, 2015).