School Ratings Must Provide Useful Information to Parents and the State

Nat Malkus and Max Eden, AEI

No Child Left Behind (NCLB) bred a broad backlash against school accountability. The Every Student Succeeds Act (ESSA) was a response to the unpopularity of NCLB and its waivers, rolling back the associated federal overreach, untenable goals, heavy-handed sanctions, and ham-fisted mandates. Now that states have regained autonomy over their accountability and school rating systems, it may be politically expedient to simply do away with any measures associated with NCLB. This would be a mistake. For all its shortcomings, NCLB marked a great leap forward in gathering and disseminating information on all students and all schools, and ESSA retained that foundation in its testing requirements. State’s school rating systems should expand and improve on that foundation by ensuring ratings provide useful indicators of school quality to the right audience. Our proposed rating system is grounded in three principles that are clear improvements on NCLB-era school rating systems.

**1. Use holistic measures to gauge performance of all schools.** NCLB’s primary measure, student proficiency, balanced the burden of its mandate with (the appearance of) a comparable measure of school performance. But in the long run proficiency proved to be an insufficient indicator of school performance that fostered perverse incentives to narrow curriculum, focus on students performing near the proficiency cutoff, and teach to a few high stakes tests. Broader based and more holisticindicators of school performance would be more informative, more difficult to game, and minimize perverse incentives.

**2. Information on schools must flow to parents and the state*.*** Heavily influenced by NCLB requirements, many states’ school reports were structured around compliance and geared to provide information to the state. Based on a flawed theory of action that state agencies are the primary locus of school reform, these systems often failed to give parents useful information. State agencies may be useful in identifying and turning around failing schools, but popular local pressure is necessary to improve decently performing schools. Moving forward, rating systems must inform parents with cogent, useful, and digestible information on schools.

**3. Do no harm.** NCLB harmed schools through its school ratings. Annual Yearly Progress (AYP) was binary, had too many tripwires for the “needs improvement” classification, created perverse incentives, and eventually labeled well-performing schools as underperforming. Accountability systems must be able to identify and dismantle the worst performing schools and provide more nuanced data on how schools fall short. They must not undermine the effective practices of average and well-performing schools, and ideally would provide guidance to take them from good to great.

## Overall School Rating and Components

Our suggested system uses school report cards that display our primary measure of school performance on a 100 point scale based on three primary domains. Schools are graded on the intuitive grade scale of “A” through “F”. We designed the scale to discriminate clearly between schools with truly average performing schools get a “C” grade. Proficiency on state reading and math exams makes up 20% of the overall score. Student growth constitutes 40% of the overall score. A measure of school climate would make up the remaining 40%. ELL students English language acquisition and achievement gaps can affect a school's overall score, but are secondary concerns. We expect that this may cut against the grain of the other plans in this forum; it certainly runs counter to the accountability systems in place over the past decade. We explain the weighting, rationale, and proposed mechanics of each domain in turn.

1. Proficiency- 20%

*Rationale:* Proficiency was a cornerstone of NCLB, but it was incapable of telling much about a school’s productivity. Proficiency tells more about the characteristics of the students entering the school than about their academic progress. It tells almost nothing about schools' operations. It is also not intuitive, particularly for parents, because the standard for proficiency is not well understood. Despite its insufficiency, proficiency is an established component of rating systems that adequately measures educational outcomes against an absolute standard. Our scale awards up to 20 points for schools’ overall proficiency, which precludes a high score for any school with many students not meeting expectations. In school report cards, the average school proficiency rate for the state would be displayed prominently as a reference, and rates for subgroups would be displayed clearly under the average for all students.

*Mechanics:* Specific mechanics of proficiency scores vary based on state tests and cut scores. However, each state should set a reasonable minimum threshold for percent proficient based on their state standards and cut-scores for proficiency. (100% a-la NCLB is not recommended!) Schools meeting that threshold would receive 15 of 20 points. Each standard deviation (of school proficiency percentages) below or above that threshold would be worth ±2 points, up to the maximum.

1. Student Academic Growth- 40%

*Rationale:* Compared to proficiency percentages, indicators of student academic growth are better suited to inform parents and the state about schools' productivity.As such, growth counts for 40 points on our scale. Using schools’ average student learning gains, our system can discriminate between the educational effectiveness of schools that take in students with large differences in academic readiness. Importantly, schools with low proficiency but admirable growth rates would be rewarded instead of unreasonably penalized. To make these measures intuitive, learning gains will be presented in terms of annual months of learning in reading and math. The measures for “months” of learning would be relative to the state average for each grade. Telling parents “students at this school make fourteen months of learning gains each year” communicates much more about school quality than “78 percent of students are proficient” does.

*Mechanics:* These growth measures are relative, indicating students’ progress compared to the state average. Schools averaging one year of learning would receive 30 of 40 possible points. Each month above or below average would be scaled to gains equivalent to raising math and reading performance to the average of the next grade. Specific determination of growth will depend on state assessments, but each standard deviation difference would be worth 4 points, up to the maximum, and subgroup scores would be displayed under the school’s average.

1. School Climate- 40%

*Rationale*: Referring to ESSA, one reformer asked incredulously, “…could a state give, say, 49% weight in school ratings to only the factors of student engagement, parent engagement, and school climate?” As if that were a bad thing. We view school climate as an important component of school quality and weight it at 40 percent. Quality school climate metrics look beyond outputs and can reflect whether schools have the necessary precursors for effective instruction. Communities that see their school is at 72 percent proficient are unlikely to take much action. If the rating system shows that students don’t feel safe in the halls, there is a focal point for community pressure and administrative action. The diagnostic value of specifying schools’ particular strengths, weaknesses, and improvements over time will inform parents about important aspects of school quality and help schools build their capacity to excel.

School climate measures are still developing. However, states could use private survey companies, such as the trailblazing startup [Panorama Education](https://www.panoramaed.com/), or a resource like the Department of Education’s free School Climate Survey platform ([EDSCLS](https://safesupportivelearning.ed.gov/edscls)). Both gather data from students, teachers and non-instructional staff, and parents, measure multiple domains of school climate that are comparable across schools and over time, and use a broad base of respondents. Such surveys are not easily gamed and allow schools to identify specific safety, engagement, and culture problems, as well as improvement over time. School climate’s prominent place in our school grade allows parents, the public, and administrators a look inside schools across the quality spectrum, to assess and remediate problems.

*Mechanics:* States have significant discretion to define the instruments they use, but effective climate surveys would measure multiple aspects by surveying school personnel, students and parents. Schools meeting the normed acceptable climate threshold score should would receive 30 of 40 points, with each standard deviation from the threshold worth 4 points, up to the maximum. School climate scores would be reported with percentile ranks and details, including differential responses by subgroups, would be available in a separate report.

Schools with **academic growth gaps** by student subgroups (race, ELL status, disability and poverty) or whose **ELL students** make insufficientprogress towards proficiency would lose growth points. Proficiency, growth, and climate affect all students, and if growth is equitable, schools’ overall scores would accurately reflect quality. Measurable equity gaps should be transparent and should attenuate schools’ points for growth. However, consistent with our conviction that schools’ primary goal is productively educating all students, the maximum penalty would be limited to 13 points.

**ELL students’** progression towards English language proficiency is weighted in proportion to their presence in the school. ELL progression, as defined by states, would count more in schools with many ELL students, and less in schools with few. Student Academic Growth points would be deducted so schools with average ELL progression would lose 25 percent of eligible points, ±10 percent of points for each standard deviation from the state average. ELL progression could count for one point for every 5 percent of all students classified as ELL, with a maximum of 5 points.

**Subgroups’ growth** will be displayed “above the fold” on the rating card for any subgroup comprising 8 percent of students. Growth gaps over 1 month below the school average would flagged and would reduce the growth score by 1 point per subgroup, and by 2 points for over 2 months. The total penalty for growth gaps is capped at 7 points. This method highlights learning gaps for parents and the state and would prevent schools with equity issues from scoring well, without crippling a school’s overall score. The transparency in months of learning would create pressure for equity without creating outsized and perverse incentives.

## School Report Cards

The school report card is a vital component of an effective school rating system because it provides a single source for interpretable information. It will display all of the components of our rating system. In addition, the bottom of our school report card provides limited space for school voice by making 420 characters (three tweets’ worth of text) available for principals to describe what makes their schools successful. Limited only to comments about their school, this space allows principals to pursue and trumpet their own version of excellence.

 *{Insert Mock-up of report card here. Forthcoming and delayed by snowstorm in DC.}*

Our school rating system reflects a holistic and balanced view of school quality, focusing on productivity more than proficiency and incorporating climate as well as academics measured by test scores. The balanced and intuitive overall score and transparent, detailed sub-scores will provide useful information to parents and states. Importantly, it avoids harm to schools by using balanced measures, identifying schools in need of immediate intervention and support, avoiding perverse incentives, and giving principals a voice.

Our ratings would clearly identify underperforming schools by the schools’ percentile rank in the state with normative and relative benchmarks. As ESSA demands, the lowest 5 percent (or more, as the state determines) would receive interventions and additional support. This system is presented as a guide for each state to calibrate to fit its needs.

Though our focus is on school ratings, we include two suggestions for intervention consistent with our principles. Providing information to parents undergirds choice as a strong regulatory mechanism for improving school quality. However, that information isn’t useful for parents without choices. Particularly for persistently low scoring schools, the state should ensure parent choices in better performing schools, be they traditional public, charter, or private options. Secondly, struggling schools need more than admonishment. These ratings should trigger some administrative flexibility, technical assistance, and resources to increase schools’ capacity to improve.