**ESSA Accountability Proposal**

1. Design Objectives:
2. **Clear and Easy-to-Digest Information for Parents.** *It is best to inform the public using simple and credible measures of student achievement levels and school climate that are different from the more nuanced measures used to inform educators about their job performance.* In the absence of easy-to-digest indicators of student performance, parents often make misguided decisions based on crude proxies for school quality. These decisions tend to reinforce negative stereotypes, like school quality being dictated by local residents’ wealth. School evaluation reports should accurately depict typical student achievement levels and parent satisfaction levels, without assigning credit or blame for those simplified metrics.
3. **Constructive Feedback for all Schools Concerning Relative Strengths and Weaknesses.** *Professional workers in any year have relative strengths and weaknesses in their job performance; by providing non-threatening objective feedback to educators, one can motivate them to shift more attention toward improving their relative weaknesses even if it means devoting less attention to their relative strengths*. This relative strengths/weaknesses feedback should not be couched within an evaluation of overall job performance. Imagine your boss telling you that you efficiently completed most of your assignments this year and that you should spend more time mentoring your co-workers. Now instead imagine your boss telling you that you received a C grade for the year, which would have been higher if you had spent more time mentoring your co-workers. Which type of feedback is more likely to illicit a productive response? All schools should receive annual feedback concerning two areas where performance was relatively strong and two areas where performance was relatively week. By breaking down performance indicators across several content areas, grade levels, and relevant student subgroups, these evaluations will encourage schools to prioritize learning where is lagging the most. For some schools, this will actually mean prioritizing better learning among their highest-achieving students.
4. **Incentivize Improvements without Inducing Over-shifting.** In any industry, if you develop and publish certain metrics, then people will shift more attention to those metrics. In Major League Baseball, improved statistical measures of players’ defensive skills have recently led teams to pay additional millions of dollars to strong defenders. In education, if we publish standardized testing in mathematics and language arts, then we can expect a shift of time and resources toward those subjects and away from other subjects. Research shows that this shift has already happened in the U.S. A reasonably-sized shift can sometimes be appropriate. *The key to optimally-designed and optimally-presented metrics is that they do not induce over-shifting from some areas to others.*
5. My proposed accountability system
6. **Indicator(s) of academic achievement. Evaluation reports will have three distinct sections intended for different audiences.** The public will see the full reports, but parents and educators will know that Section 1 is *not* being used to guide state policy decisions or designations.

**Section 1, “Information for Parents: School Resources, Student Performance Levels, and Parent Satisfaction”**

* School Resources:
	+ Information about school administrators, grade range served, schools’ admissions policies
	+ # Enrolled Students; Average Class Sizes
	+ Full-time Equivalent Staff by Type (aides, counselors, etc.)
	+ Percent of Teachers with Fewer than 3 Years of Prior Teaching Experience
* Student Performance Levels. Bar charts showing results for each grade level on Math and Language Arts tests (grades 3 and up), and Science and Social Studies tests (grade 5):
	+ Percent scoring “At-risk” (beyond 0.5 years below grade-level)
	+ Percent scoring “Proficient” (at grade-level or above)
	+ Percent scoring ”Advanced” (> 0.5 years above grade-level)
	+ Each of these 3 percentages presented next to similar percentages for:
		- remainder of school district
		- an aggregate comparison group of 10 schools in other districts with similar student demographics (% free/reduced lunch, % special education, etc.)
		- the rest of the state
	+ One/Two red asterisks anywhere < 95%/85% of relevant students contributed to calculated percentage
* Parent Engagement and Satisfaction. The school district will mail/email every parent with a hard copy/weblink of a survey concerning various issues of school climate. Responses will go directly to the state DOE for tabulation, with anonymity preserved concerning which particular parent is responding.
	+ Percent of Parents completing the survey
	+ Stacked bar charts showing, on a scale of 0-4, how satisfied parents are with various aspects of the school, including:
		- teachers’ responsiveness to their child’s learning needs
		- schools’ instructional programs for mathematics, Language Arts, **science**, **social studies**, and **art/music**
		- how much their child **enjoys** going to school
		- how well the school keeps them informed of their child’s progress
		- whether the parent attends school events, such as parent-teacher conferences
	+ Similar stacked bar charts for school’s comparison group and rest of the state

**Section 2, “Information for Educators: Two Relative Strengths and Two Relative Weaknesses during the Prior School Year”**

A list of two relative strengths and two relative weaknesses, based on how the school ranked on indicators compared to other schools in the state. For example, if a school is ranked in the top 20 percent in the state for 26 of its 28 categories but only in the top 30 percent of the state for the other two categories, then those two categories are the “Two Relative Weaknesses.” The categories used for this include:

* Growth measures for each of the annually tested subjects (Math or Language Arts) for
	+ Students by grade level (except for the lowest tested grade)
	+ Students by initially low/middle/high achieving performance: subgroups of students based on whether they were in the bottom quartile, middle two quartiles, or top quartile of achievement during the prior year
	+ Students by racial group, students from lower-income families, English learners, and disability status
* Percent of students in relevant grades mastering three content knowledge subsets each in Language Arts and Math chosen by the state for the lowest tested grade level. For example, “% mastering addition in 3rd grade.”
* Two indices for 5th grade student performance on Social Studies and Science tests:

Index = (% “Proficient or better”) + (% “Advanced”) - (% “At Risk”)

**Section 3, “Information for the State Department of Education: Persistent Excellence or Persistent Difficulties in Student Performance in Mathematics and Language Arts”**

For each of Language Arts and Mathematics, the school will receive one of the following summative status measures:

* Persistent Excellence
* Excellent
* Adequate Growth
* Need for Improvement
* Persistent Need for Improvement

Schools are placed into one of three, equally-sized groups based on the total number of students with valid growth measures for that year. A school receives *Need for Improvement* for a subject in any year in which:

i. It places in the bottom 6 percent of its size-group in the average student growth measure for that subject that year and had already placed in the bottom 20 percent during either of the two prior years

AND at least one of the following is true for that same subject:

iia. Fewer than 85 percent of the schools’ students are scoring at “Proficient” in the earliest tested grade.

iib. Fewer than 75 percent of the schools’ students are scoring at “Proficient” in any other grade.

iic. For Language Arts only, the school has at least 5 students who have been classified as English Language Learners for more than 2 years and are scoring “At-Risk” for Language Arts.

iid. The school ranks in the bottom half of the state in terms of parental satisfaction with instructional quality in that subject, or fewer than half of its parents completed the survey.

A school receives *Persistent Need for Improvement* if it would have qualified for *Need for Improvement* for its third consecutive years.

Similarly, a school is designated as *Approaching Excellence* in either subject if

* it places in the top 25 percent of its size-group in the average student growth measure for that subject for two consecutive years,and
* NONE of conditions iia through iid were true, and
* at least 90 percent of students in the school are tested in the subject that year.

A school receives *Persistent Excellence* if it receives *Approaching Excellence* in that subject for its third consecutive year. Schools receiving neither Excellence nor Improvement status receive *Adequate Growth*.

1. **Indicator(s) of student growth**. Student growth is first measured at the student level and adjusted for measurement error; it is based on year-to-year changes in test scores adjusting for typical variance in changes in scores among students with initial test scores in a similar prior range. Depending on the category, that student-level measure of test score growth is averaged across a specific grade or a subpopulation of students. A student subpopulation category is only used for a school if there are at least 15 students with non-missing values in the relevant grade or subpopulation. Growth scores are never combined across subjects; they are used to determine separate subject-specific indicators.

Section 3 avoids some common pitfalls with using growth measures:

* Schools serving high-achieving students may receive low growth scores. These schools might or might not need improvement plans.
	+ Solution: Schools are not designated for improvement if they have both high student proficiency rates and high levels of parental satisfaction with instruction.
* Schools with fewer tested students will inherently have more variability in their average growth due to fewer students contributing to these averages.
	+ Solutions: Compare schools with similarly-sized schools. Require minimum test participation for Excellent. Flag cases where fewer than 95/85 percent of students were tested.
* Schools may have an incentive to tank performance in the youngest tested grade or in a particular school year as a way of artificially increasing growth scores in the next grade or the next school year
	+ Solutions: Use multiple years of growth measures. Include proficiency rates in the earliest tested elementary grade as an important factor.
* Schools’ growth scores for a given year may be an aberration.
	+ Solution: When relevant, allow recent successes or failures to negate classifications based on current-year performance.

**c) Indicator(s) of progress toward English language proficiency**. Section 2 and Condition iic in Section 3 ensure that schools cannot ignore this.

**d.) Indicator(s) of student success or school quality**

Section 1 allows parents to easily compare school resources, school climate, and student performance across schools. But the focus there is differential performance based on students’ prior achievement levels and learning needs. Breakdowns by race and family’s poverty status are purposefully restricted to Section 2. This avoids punishing more diverse schools, but requires schools to focus on all types of students.

**e.) Calculating summative school grades**. See Section 3 above. To comply with *ESSA*, if the bottom 6 percent/bottom 20 percent growth rule does not produce at least 5 percent of schools with Need for Improvement in at least one of the subjects, then the “bottom 6 percent” cutoff rises until 5 percent of schools are designated.

**f.) What about schools with low-performing subgroups?**  For Section 2, student subpopulation categories are only used if at least 15 students contribute scores. If the same category appears in 3 out of 4 years as a Relative Weakness, then the school must work with the district on a Targeted Improvement Plan. Even highly-successful schools should occasionally devise Targeted Improvement Plans to improve how they serve certain student populations.

**g.) School grades or ratings**. Words matter. By avoiding “failing” designations and letter grades, derogatory “school accountability” language is replaced with constructive feedback via “school evaluation.” We want strong evaluation that provides objective feedback used to guide reforms, *and* we want educator buy-in. For Section 3, if a school receives *Need for Improvement* in either subject, their district must submit a Plan for Improvement to the state. The state would intervene in any school receiving *Persistent Need for Improvement* in either subject.

**3. Recommendations for the Department of Education**.

Make it clear that states have flexibility to use student subgroup performance in ways that don’t penalize more diverse schools. This proposal is just one example of how states can accomplish this.