

**CONSIDERATIONS FOR STATE LEADERS IN THE DESIGN OF SCHOOL  
ACCOUNTABILITY SYSTEMS UNDER THE *EVERY STUDENT SUCCEEDS ACT***

**Scott Marion**

**National Center for the Improvement of Educational Assessment**

**February 19, 2016**

The Elementary and Secondary Schools Education Act (ESEA) was finally reauthorized as the Every Student Succeeds Act (ESSA) this past December. The reauthorization was long overdue and with its passage comes much hype and some misinformation about what the law permits and does not permit. The purpose of this brief is to outline some of the key accountability provisions of ESSA and offer some considerations for states as education leaders move to meet the accountability requirements of the law that capitalizes on what has been learned in the state previously and how an ESSA accountability system can best support the state's policy and educational goals

ESSA Accountability Overview

The ESSA-required accountability system must be **operational in the 2017-2018 school year**, which likely necessitates a pilot in the 2016-2017 school year. Therefore, states are now in a position of preparing for a pilot in less than nine months. This requires a fairly quick design and development process. There are two main components of the ESSA accountability system:

1. Reporting requirements: States must continue to report by all required subgroups specified under NCLB.
2. School accountability determinations: States must categorize schools based on state-determined goals and methodology.

Accountability Indicators

It is important to keep in mind that while ESSA outlines the basic structure of state accountability systems, the specifics of the accountability design will need to be worked out in the rule making process. However, it is important to begin planning for the accountability system because, as noted, the timeline for implementing the new system is quite short. The goal setting requirements described above are a key aspect of the accountability design. Additionally, the law describes five types of indicators to be included in a school accountability system:

1. **Academic achievement** is also referred to as status or point-in-time indicators. Under NCLB, achievement was reported as the percentage of students scoring at the proficient level or higher. Percent above cut (e.g., proficient) has been criticized for many measurement (e.g., reduction of information) and consequential (e.g., focusing on “bubble kids”) reasons, but it does have the advantage of familiarity and relative ease of understanding. While states are still required to report percent proficient, ESSA may allow for approaches that rely on information throughout the achievement distribution such as an index system or average (i.e., mean) scale scores.
2. **Another valid and reliable academic indicator** must be included in the accountability system. The law offers student growth and achievement gap closure as two potential examples, but it is not limited to those examples. That said, measuring achievement gaps is one of the trickiest things to do well in educational measurement. Simple approaches such as computing the differences in percent proficient are almost always wrong, while more technically correct approaches such as computing the area between two performance distributions or even effect sizes are a bit more challenging to explain. As difficult as it is to measure achievement gaps at any point in time, the measurement challenges associated with measuring changes in achievement gaps are enormous. On the other hand, there are well-established methods for documenting student growth such as student growth percentiles (SGP).
3. **Graduation rate** must be part of the accountability system for high schools. Further, extended graduation rates such as five and six year rates can be included at the state’s discretion.
4. **English language proficiency** rates and progress is a new accountability requirement under ESSA, at least under Title I accountability. This is largely because Title III accountability has now been rolled into Title I. This is one of the aspects of ESSA that will need rules to help us better understand the requirements. For example, one of the key tenets of accountability design is that the results of applying the accountability rules should not privilege or reward schools based on the demographic characteristics of the school. Given that English language proficiency is a relevant indicator in only those schools with enough ELL students to meet the minimum group size, state leaders are going to have to do some thoughtful design work to ensure that schools are held responsible for developing English language proficiency in their students, but the presence of this indicator does not automatically disadvantage such schools’ accountability determinations. Applying this indicator will likely require additional insights gained through subsequent rules and/or guidance.
5. ESSA also requires the use of an **indicator of school quality or success** that meaningfully differentiates and is valid, reliable, and comparable. My colleague, Chris Domaleski, has termed this the “unicorn indicator” because it is something we have all heard about but never really seen. It is clear that the authors of ESSA wanted to broaden notions of school quality by including indicators in the system other than those based on

test scores. Because of its relative novelty in current accountability systems, I discuss some ideas for this fifth indicator in more detail below.

### The Fifth Indicator

Most of the indicators required under ESSA are at least familiar, even if the specific metrics proposed may be new under ESSA. However, the types of metrics and indicators suggested for the fifth indicator are relatively new and generally have not been used in accountability systems. The specific passage from the law defining this indicator follows:

- (v)(I) For all public schools in the State, not less than one indicator of school quality or student success that—*
  - (aa) allows for meaningful differentiation in school performance;*
  - (bb) is valid, reliable, comparable, and statewide (with the same indicator or indicators used for each grade span, as such term is determined by the State); and*
  - (cc) may include one or more of the measures described in subclause (II).*
- (II) For purposes of subclause (I), the State may include measures of—*
- (III) student engagement;*
- (IV) educator engagement;*
- (V) student access to and completion of advanced coursework;*
- (VI) postsecondary readiness;*
- (VII) school climate and safety; and*
- (VIII) any other indicator the State chooses that meets the requirements of this clause.*

As can be seen above, there are several psychometric characteristics required of the indicator—valid, reliable, and must differentiate performance—but, in general, the options for what can be used as an indicator are fairly wide open. That said, it will be important to consider each of these technical requirements in the selection of potential indicators. While reliability is easily defined, the validity of an indicator (within a system context) is less clear but needs to be based on a well-articulated theory of action. Our current thinking about “differentiate” is that the law intends for indicators to have a fair amount of true variability among schools compared with indicators such as elementary school attendance that essentially acts as a constant in the system.

States need to thoughtfully ensure that this additional indicator fits with stakeholders' conceptions of educational accountability and school quality. Do stakeholders think this additional indicator will broaden the "construct" of school quality because previous test-based accountability systems have missed important aspects of school effectiveness? On the other hand, some might consider these indicators useful for accountability systems because they serve as precursors to the achievement and growth academic indicators. For example, some might want to include an indicator of student engagement because they think it is a precursor to higher levels of student achievement, while certain social-emotional learning indicators help broaden our characterizations of school quality. Obviously, there can be considerable overlap among these conceptions.

These distinctions are important, because it highlights how one approaches the development and validation of the indicator. If the indicator represents something distinct from traditional test-based academic achievement, then we would not necessarily expect a strong relationship between assessment performance and favorable performance on this indicator. For example, one might think of a school engagement initiative that encourages students to participate in community service or other applied projects. Such engagement may help students hone leadership skills and other characteristics associated with being responsible global citizens, but may not be well-measured on tests. It stands to reason, then, that validating the indicator with assessment data would be misplaced. Rather, we would seek other data thought to affirm our understanding of the construct. On the other hand, one might operate from a perspective that encouraging students to be engaged in community service or other applied projects increases motivation and hones critical thinking skills essential to academic success. With this view, one expects students who are more engaged to perform better on academic assessments. If not, our understanding of the construct is less certain.

There is no question that the indicators listed as examples in the statute could provide rich information to schools and districts beyond test scores. However, many of the potential indicators such as school climate, student or teacher engagement, or other social-emotional indicators are often based on self-reported information through surveys or other similar approaches. We must carefully consider "Campbell's Law" when using any indicator, but especially those easily corruptible if they are used as part of a high stakes (or at least publicly reported) accountability systems.

*The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor<sup>1</sup>.*

---

<sup>1</sup> Campbell, Donald T., [Assessing the Impact of Planned Social Change](#). The Public Affairs Center, Dartmouth College, Hanover New Hampshire, USA. December, 1976.

This “law” has been observed many times over since 1976, but especially in the past 15 years. The double-edge sword described by Campbell is that not only will the indicator be potentially corrupted, but the underlying trait or quality we are trying to measure will be distorted as well. In other words, we need to be really thoughtful and careful in our accountability system design and especially in the design of this fifth indicator.

One of the ways to minimize the corruption pressures is to consider multiple indicators for this category. For example, if this indicator was worth 15-20% of the overall rating, by using 3-4 indicators, each one would be worth only 5% of the overall score, which would lessen the risk of corruption because the potential reward is so small. Another way would be to consider indicators that required clear demonstrations of evidence where corruption may be minimized.

We, at the Center for Assessment, have been advocating for quite some time, that accountability systems need to be designed according to a well-articulated theory of action that clearly lays out the intended goals and outcomes as well as proximal and intermediate indicators and the mechanisms and processes necessary to realize these goals. This is a critical first step in designing the system, but it is especially critical when selecting/creating an indicator or indicators in this category. Further, we and others, notably Hargreaves and Braun<sup>2</sup> have offered concrete suggestion for designing improvement-based (compared with punitive) accountability system. While the design of the entire system should attend to these recommendations, the selection of the 5<sup>th</sup> indicator (or indicators) offers opportunities to try to meet these improvement intentions. If we agree that accountability systems should incentivize the types of behaviors we want to see and disincentivize the behaviors we do not want to see, then we need to think about this fifth indicator in this light.

Part of the thinking about the theory of action for an accountability system is that it is both constrained and informed by the political, educational, and financial context in which the accountability system sits. Therefore, state leaders should consider the selection of the 5<sup>th</sup> indicator or indicators to advance key policy and/or educational goals. Before considering potential indicators or types of indicators, it is important to remember the ESSA requirements related to this indicator. In particular, the statute requires “the same indicator or indicators [to be] used for each grade span,” which limits the flexibility a state has in tailoring different indicators to different school districts depending on need. A narrow reading of the law suggests that the same measures (e.g., school climate survey, assessment of student engagement) must be used at least for each grade span. However, a more flexible reading would suggest as long as the same indicator is used (e.g., student engagement), schools/districts may be able to use different specific measures of the same indicator. While this may be addressed in the rules, we can make a case for this latter position if we can document some degree of comparability across schools.

---

<sup>2</sup> Hargreaves, A. & Braun, H. (2013). *Data-Driven Improvement and Accountability*. Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/publication/data-driven-improvement-accountability/>.

The important point is that we should view this fifth indicator as an opportunity to further important state policy goals rather than as a burden of “just one more thing” to include in the accountability system. Further, states should use the time prior to 2017-2018 to try out a variety of indicators to evaluate the quality of data received and the burden associated with collecting such data.

### School Accountability Determinations

School accountability categorizations must be based on state-determined goals and methodology, but like much of ESSA, there are certain federal requirements and constraints, particularly in terms of reporting. There are fewer requirements about goal setting and combining the multiple indicators. We suspect that more details will be forthcoming through the rule-making process.

Goal setting is an important aspect of ESSA. Under NCLB, the goal of 100% proficiency by 2014 was established for all states statutorily, with a limited number of alternative approaches allowed under NCLB waivers. Under ESSA states are required to determine status (point in time) and improvement goals for at least three sets of indicators:

- ✓ Academic achievement (status or improvement),
- ✓ Graduation rate, and
- ✓ Sub-groups that are behind.

The law makes reference to ambitious goals so we doubt that goals with a 25 year timeframe would pass muster. That said, states have an opportunity to be thoughtful in setting ambitious, but reasonable goals. What is not clear is the degree to which accountability determinations need to be based separately (conjunctively) on these three set of goals or whether reporting on these goals will satisfy the law. ESSA is explicit, however, that states must continue to report by all required subgroups specified under NCLB, but we are not yet sure regarding how the results of individual subgroups need to factor into accountability determinations.

ESSA offers little guidance regarding how the various accountability indicators should be combined to produce an overall accountability determination other than to require that the first four indicators in the aggregate must have “much greater” weight in the overall determination compared to the fifth indicator. Interestingly, this requirement shifted from “greater” to “much greater” in one of the last iterations of the law. We know that the rules and guidance will provide more specificity on overall determinations.

However, states need to consider the ways in which they combine indicators (or not) to be coherent with the specific goals and the overall system theory of action. For instance, not all indicators need to factor into an accountability determination as long as the required indicators are included. This means that the state can have a rich reporting system to help support

accountability decisions, which may allow for the state to try out indicators that may be easily corrupted if used for accountability, but when used in a low-stakes reporting system, may provide information useful for improvement.

Compared to NCLB’s very prescriptive conjunctive approach for arriving at overall determinations for schools, ESSA appears to allow for more varied approaches for states to use to produce overall ratings. States must, starting in 2017-18 and **at least once every three years** thereafter, produce **a statewide category of schools** for comprehensive support and improvement for schools in the following categories:

- ✓ lowest performing 5% of Title I schools,
- ✓ HS with graduation rate less than 67%, and
- ✓ schools with low performing subgroups.

State systems can produce determinations more frequently or include more performance categories.

### Educator Evaluation

While not required, ESSA authorizes states to use funding to implement teacher and leader evaluation systems, reform teacher and school leader certification systems, improve equitable access to effective teachers and leaders for all students, and develop mechanisms for effectively recruiting and retaining teachers. Further, states are still required to disclose the steps they’re taking to evaluate and publicly report on the inequitable distribution of teachers and the qualifications of their teachers and school leaders, spelled out by high- and low- income schools and schools with high and low concentrations of students of color. Finally, ESSA enshrines into law, the Teacher and School Leader Incentive Fund Grants (TIF), with the goal of expanding performance-based compensation systems and human capital management systems for both teachers and principals.

While the absence of federally-required teacher evaluation systems under ESSA received a lot of press when the law was passed, it is clear that states can continue their efforts, with federal support, to improve the quality of state and district educator and leader effectiveness systems. Further, it is not clear how states could report on the “inequitable distribution of teachers” without some sort of systemic evaluation data. The Center for Assessment and others are currently working on some educator evaluation 2.0 approaches that may fit more coherently with a progressive school accountability system.

Finally, given ESSA’s relative silence on educator evaluation, a state could choose to include something like the “quality of educator evaluation decisions” as the fifth indicator as long as we can come to agreement about how to operationalize this indicator.

### A Sketch of a State Design Process

As noted above, the ESSA accountability system is required to produce operational results for the 2017-2018 school year, but must be piloted during the 2016-2017 school year. In other words, time is of the essence! It is important to get started on this work quickly, but since accountability systems are designed to instantiate stakeholder values, it is critical to avoid shortcutting opportunities for key stakeholders to provide meaningful input. However, accountability systems cannot be designed by hundreds of people, so what follows is a very high-level sketch of a process designed to both include all relevant stakeholders, but to do so efficiently.

1. There must be an internal state department of education (DOE) group, operating on behalf of the state chief, that can make critical policy decisions. Similarly, there must be a DOE person who is the responsible point person for this work. A technical consultant, if used, would work directly with this point person and the internal leadership group.
2. Early meetings should be convened with leaders of key stakeholder groups, such as the various associations, state board members, gubernatorial representation, and legislative leadership. These meetings will be designed to ensure that representatives understand the constraints, requirements, and opportunities available under ESSA and to ensure that the representatives understand and, to the extent possible, buy into the proposed design process. This document, for example, can serve as the foundation for these meetings. These groups should be informed of the progress on a regular basis (e.g., 2-3 months), with groups such as the district superintendents informed more regularly.
3. The DOE leadership and key stakeholders should be clear regarding the degree to which it wants to build on an existing accountability system in the state or start with a blank slate.
4. A working group—something like an “accountability task force”—should be charged with serving as advisors to the smaller group of system designers. The membership of this group may need to be expanded to ensure that key stakeholders are appropriately represented. This group will need to meet monthly, at a minimum, to reflect on design work and to help weigh in on key value and practical decisions. This group should be convened as soon as possible because of the need to get to work. There is no need to wait for all of the meetings described in #2 to begin working with the task force.
5. The advisory group, along with other key stakeholders, will first have to explicitly articulate goals for the system. This foundation will be an important touchstone for creating a theory of action to guide the design of the full system.
6. The lead DOE representative(s), along with the technical consultant if used, will be responsible for bringing design proposals to the advisory group and reflecting the advisory group’s input in subsequent meetings.

7. Once the goals are agreed upon, the advisory group will turn to identifying appropriate indicators and approaches for measuring the indicators that fit with the theory of action.
8. The DOE and technical consultant will model the various indicators and work with the advisory group to determine how best to aggregate and combine (or not) the various measures to make overall determinations.
9. The goal will be to have a system design produced by September 2016 that can be piloted through the 2016-2017 school year.
10. After the 2016-2017 pilot period, the technical consultant and the DOE representative will work with the advisory group to analyze the pilot results and propose a final design for 2017-2018.

Of course, these are just recommended and general steps that can be tailored to each state's needs. As noted above, it will be important for states to begin this work quickly even though they will need clarification from the U.S. Department of Education through the rule-making process and through guidance. The accountability systems called for under ESSA clearly offer more flexibility than under NCLB even with the waivers. Taking advantage of this flexibility to further the state's policy and educational aims will require considerable thought and collaboration with other states and other partners.