



ASSESSMENT AND ACCOUNTABILITY RECOMMENDATIONS FOR THE NEXT REAUTHORIZATION OF THE ELEMENTARY AND SECONDARY EDUCATION ACT

January 28, 2020

Scott Marion, Chris Domaleski, and Chris Brandt¹



**Center for
Assessment**

National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

CENTER FOR ASSESSMENT



INSIDE

INTRODUCTION.....	3
ASSESSMENT SYSTEMS	4
ACCOUNTABILITY SYSTEMS	7
INNOVATION	10
RESEARCH AND EVALUATION.....	11
CLOSING THOUGHTS.....	12

¹ We are grateful to our colleagues at the Center for Assessment, numerous clients, and other partners who have helped shape our ideas through many conversations and collaborative work. We are also thankful to Charlie DePascale, Paul Leather, and Ted Coladarci for very helpful contributions and editorial support. Of course, any errors or omissions are our own.



ASSESSMENT AND ACCOUNTABILITY RECOMMENDATIONS FOR THE NEXT REAUTHORIZATION OF THE ELEMENTARY AND SECONDARY EDUCATION ACTⁱ

INTRODUCTION

At the National Center for the **Improvement** of Educational Assessment, we think a lot about how state assessment and accountability systems can promote student learning. With an eye on promoting innovation, restoring balance, and bolstering impact, we propose a number of recommendations to consider in the next reauthorization of Elementary and Secondary Education Act (ESEA), which could happen as early as 2021.

While we have quite an appetite for innovation, we nonetheless are sensitive to practical limitations: Assessment and accountability design is always a case of operationalization under constraints. Importantly, policy requirements are a key constraint for the design and implementation of state-level assessment and accountability systems.

The introduction of the ESEA in 1965 signaled a sweeping federal initiative for improving the equality of educational opportunities in K-12 public schools. The law was a key part of Lyndon Johnson's Great Society programs and was intended to provide substantial resources to support educational opportunities for economically disadvantaged students. Given the considerable financial commitment in what had always been state responsibility, it is not surprising the federal government required states to hold its schools accountable for how they used these resources. Early accountability efforts addressed inputs, but subsequent reauthorizations shifted the focus to outputs, which, in turn, increased the requirements and associated constraints on state assessment and accountability systems. The 1994 reauthorization, the Improving America's Schools Act, ramped up the federal policy influence on assessment system design, but it was the 2001 reauthorization, the No Child Left Behind Act (NCLB), that exerted a massive influence on states' accountability systems.

The current authorization, the Every Student Succeeds Act (ESSA), offers more assessment and accountability flexibility than NCLB—a low bar, indeed—but ESSA arguably remains far too restrictive to spur the needed innovation in assessment and accountability systems for supporting student learning and organizational change in schools. For example, the basic assessment requirements of testing all students English language arts and mathematics performance in grades 3-8 and once in high school remain in place as does the requirement for testing science achievement once each in elementary, middle, and high school. True, ESSA allows states to replace their high school assessments with “nationally recognized” college readiness assessments, as long

With an eye on promoting innovation, restoring balance, and bolstering impact, we propose a number of recommendations to consider in the next reauthorization of Elementary and Secondary Education Act (ESEA), which could happen as early as 2021.

as such assessments meet the same technical requirements. The law also allows states to use multiple interim assessments in place of a single summative assessment. But no state has taken advantage of these provisions because of the difficulties of determining a valid summative score, mismatches with local curricular sequences, and several other technical and logistical challenges as our colleagues Nathan Dadey and Brian Gong [explain](#).

We acknowledge that the ESSA accountability system requirements, which entail outcomes beyond student achievement and graduation, are considerably broader than the Adequate Yearly Progress requirements of NCLB. We want to build on the flexibility of ESSA, but, in doing so, find ways to improve assessment and accountability for the ultimate purpose of increasing student learning opportunities. Our recommendations pertain to assessment systems, accountability systems, innovation, and research and evaluation.

We want to build on the flexibility of ESSA, but, in doing so, find ways to improve assessment and accountability for the ultimate purpose of increasing student learning opportunities.

ASSESSMENT SYSTEMS

Balanced Assessment Systems

A *balanced* system of assessment is important because it more meaningfully serves the multiple stakeholders of the educational system, particularly those closest to the classroom. ESSA focuses solely on the state accountability assessment, which might be appropriate if the purposes ascribed to the summative tests were more limited. However, this law places far too many—and sometimes conflicting—demands on the summative assessment. For example, ESSA requires diagnostic uses for summative assessment reports but also couples the summative assessment with intense accountability pressures. This is an unreasonable expectation; if we ask one assessment to do everything well, it will do nothing well. A better approach is to restrict the summative assessment's requirements while encouraging the development of balanced assessment systems. To be sure, ESSA is not the only impediment to creating balanced assessment systems (see our [Tricky Balance](#) paper). But by relaxing the assessment requirements (including regulations and peer review requirements), providing incentives to states, and supporting research on balanced assessment system design and implementation, the federal government can partner with state and district leaders to improve the use of balanced systems of assessment. We discuss these recommendations below.

If we ask one assessment to do everything well, it will do nothing well.

Assessment Sampling and Maintaining Equity

Our ability to optimize assessment design is constrained by ESSA's continuation of the NCLB requirements for comparable annual student-level determinations (i.e. scores and proficiency levels that are comparable across students). We understand the equity rationale for having every student participate in the same assessment in every grade (3-8 and HS), every year. However, this requirement is not necessary for monitoring equity, as called for by the law. In fact, it may be counterproductive. Since NCLB's inception, for example, we have seen increased achievement gaps

among key student groups—an observation that doubtless would have surfaced had we tested only a *sample* of students in the targeted grade levels. Nobody should view the concept of sampling as backing away from equity. Perhaps if states had been encouraged to create assessments and assessment systems with research backing to support increases in learning (e.g., formative assessments), then we might have made more progress in closing achievement gaps than we have up until now.

Nobody should view the concept of sampling as backing away from equity.

Thus, we recommend allowing states to incorporate sampling—of both students and content—in the design of their assessment systems. Reduced testing time frees up time and resources for instruction, assessment literacy development, and especially for creating deeper and more relevant local assessment systems. Such flexibility will reduce costs, of course, but it also can improve system validity by assessing content and skills more meaningfully and deeply. Sampling methodology is the foundation of most scientific disciplines, as it is for educational measurement. We outline two major sampling classes below.

Reduced testing time frees up time and resources for instruction, assessment literacy development, and especially for creating deeper and more relevant local assessment systems.

Sampling students. Rather than test every student in all required grades every year, we can select from various sampling approaches depending on the desired inferences. We can return to the Improving America’s Schools Act model by testing all students only once in the elementary grades, the middle grades, and high school (instead of grades 3-8 and HS), or we can sample students within the current tested grades. The latter carries certain logistical challenges for monitoring subgroup performance, so it should be used only if certain safeguards (e.g., subgroups might need to be “over-sampled” to ensure they are not missed in reporting) are in place.

We acknowledge that certain sampling approaches may disrupt current approaches for tracking academic growth annually. But we are confident that it will still be possible to provide multiple measures for producing growth estimates. In fact, we believe producing student and aggregate measures of growth may be an important aspect of a comprehensive assessment and accountability system. We urge state leaders to be clear in their theory of action about the extent to which measuring student longitudinal growth on a common summative assessment is essential for realizing their intended outcomes. The resulting assessment system design should support this theory of action.

Sampling items. Matrix sampling divides the entire pool of test questions into a specific number of test forms so that each student completes only a portion of the total test. Consequently, school-level achievement results can reflect a much broader and deeper content domain than when all students take the same or parallel test. It also is possible to adopt a hybrid approach that yields student-level scores. Matrix sampling for large-scale monitoring assessments (used for accountability or large-scale monitoring of educational trends) was endorsed by a National Research Council committee responsible for [recommendations related to statewide science assessment](#).

Aligning Technical Assessment Requirements With the Purposes of the Assessment

ESSA's technical requirements for state assessments, spelled out in Section 1111 (b)(2)(B), are too detailed and restrictive. Each requirement should be appraised in light of the core purpose of these state summative assessments: school accountability. Two requirements are particularly troublesome in this regard:

(x) produce individual student interpretive, descriptive, and diagnostic reports, consistent with clause (iii), regarding achievement on such assessments that allow parents, teachers, principals, and other school leaders to understand and address the specific academic needs of students, and that are provided to parents, teachers, and school leaders, as soon as is practicable after the assessment is given, in an understandable and uniform format, and to the extent practicable, in a language that parents can understand;

(xii) enable itemized score analyses to be produced and reported, consistent with clause (iii), to local educational agencies and schools, so that parents, teachers, principals, other school leaders, and administrators can interpret and address the specific academic needs of students as indicated by the students' achievement on assessment items;

Even as well-trained measurement professionals, we puzzle over how best to meet these requirements on an end-of-year summative assessment designed for accountability uses. End-of-year accountability tests are just not designed to provide fine-grained information and feedback necessary to improve learning. These requirements are purportedly addressed through subscore reporting, which, unfortunately, is rarely done in a meaningful and [technically defensible manner](#) and only modestly addresses the ESSA requirements.

The two ESSA requirements above speak to the need for a system of assessments, as does the requirement for “multiple up-to-date measures of student academic achievement, including measures that assess higher-order thinking skills and understanding” (subsection vi).” We advocate supporting states and districts in creating balanced assessment systems, which would meet the multiple measures criterion, but for now ESSA requires assessments to exclusively serve an accountability function.

There are other problematic requirements as well, but many of the associated issues involve how the law has been translated through the corresponding regulations and the U.S. Department of Education's peer review process. Consider the matter of alignment. The alignment requirement has reified the criteria delineated by Norman Webb and others in ways that generally force states to privilege breadth over depth. This has become increasingly problematic with complex content standards such as the Common Core State Standards and the Next Generation Science Standards.

Rightsizing State Summative Assessment

We do not recommend the revisiting of ESEA sampling and technical requirements simply to reduce the corresponding burden on state assessment professionals (although that is a worthy goal). More importantly, these requirements stifle innovation (discussed below). By allowing more flexible implementation of state assessments—with appropriate safeguards (such as making sure no students or student groups are “hidden” from assessments and reporting)—ESEA will encourage the development and implementation of balanced assessment systems that can address more successfully the goals we believe motivated some of the law's requirements.

We do not advocate a free-for-all when we call for relaxing the assessment requirements. Indeed, we envision the “default” requirements in a new law being similar to the current ESSA assessment requirements. States should be encouraged, however, to propose plans that are more coherent with their aims of improving school quality and student learning for all students. Such plans should be grounded in a defensible theory of action for how and why the state wants to modify the ESSA/NCLB testing schedule while maintaining or enhancing the focus on equity.

States should be encouraged, however, to propose plans that are more coherent with their aims of improving school quality and student learning for all students.

ACCOUNTABILITY SYSTEMS

Accountability policies determine how statewide assessments are used and the consequences associated with assessment results. In short, we find that ESSA accountability is insufficiently flexible to support the hard work that state educational reform requires. Further, and in keeping with our push for balanced systems of assessment, we fear that ESSA's [focus on school accountability is out of balance](#). We offer suggestions below for creating more balanced accountability systems.

In short, we find that ESSA accountability is insufficiently flexible to support the hard work that state educational reform requires.

The Big Picture: School Accountability

The accountability system, defined in law and rule, is the result of political negotiations like all laws. Unfortunately, this also means the defined accountability system is not informed by a defensible theory of action for supporting the state’s educational efforts.

Why should the accountability system be defined so tightly in federal law in the first place? For example, ESSA specifies five accountability indicators—achievement (ELA and math), other academic measures (commonly growth or gap measures), English language proficiency, a measure of school quality or student success (SQSS), and high school graduation rate. Why these five indicators and not others?

Moreover, why are some indicators rigidly proscribed? For example, academic indicators are restricted to English/ language arts and mathematics. Consequently, when states want to add additional academic indicators to their accountability system, such as science or social studies, these indicators are classified as SQSS. Some may dismiss this as a distinction without a difference, but we regard such absurd work-arounds as evidence of overreach.

To be sure, the federal government should hold states to account for providing all students with sufficient opportunities to learn. But states can be held accountable and can hold their schools accountable without the federal government dictating the specific means of accountability.

There are better ways to design school accountability systems than having every state follow the same script. To be sure, the federal government should hold states to account for providing all students with sufficient opportunities to learn. But states can be held accountable and can hold their schools accountable without the federal government dictating the specific means of accountability.

A Broader Picture of School Quality/Student Success

With the passage of ESSA, few components generated as much buzz as the SQSS indicator. There was a flurry of interest and [writing](#) about this so called “5th [indicator](#)” when ESSA was first passed. State leaders had high hopes for using this “5th indicator” to incentivize desired outcomes by including measures of school climate, teacher engagement, and other school-wide indicators that more broadly capture a school’s quality. Unfortunately, the regulatory requirements for the SQSS indicator were particularly rigid and, in the end, this indicator did little to promote innovation.

The regulations allow only student-level indicators that can be disaggregated by student group. Actually, some examples in ESSA, such as teacher engagement, are at higher levels of aggregation, even though ED prohibited states from implementing such indicators. This constraint forced states to rely on readily available student-level information (e.g., chronic absenteeism), rather than using school-wide information or even student-level information that is best collected anonymously (e.g., school climate surveys). On a positive note, many states are using a postsecondary readiness indicator at the high school level. But even in this case, state accountability leaders have been unable to incorporate some very important reforms such the [Portrait of Graduate](#) initiatives taking hold in states and school districts because the data do not fit neatly in the accountability requirements.

We recommend relaxing the SQSS indicator requirements if it is retained in the next reauthorization to allow indicators beyond the student level such as at the classroom or school. Given its low weight, furthermore, we suggest considering whether the SQSS indicator needs to be the same statewide or allow it to be tailored to the needs of districts and states. On the other hand, we would not object to increasing the allowable weight of this indicator if it fits with the state’s theory of action.

Accountability Systems for Alternative Schools

Many of a state’s lowest performing schools are dedicated to serving students who have not been successful in traditional schools. These alternative schools often are successful in their mission to keep students in school and to provide them with the necessary skills for life after. Unfortunately, ESSA does not recognize alternative school accountability systems; rather, these schools are evaluated by an accountability model that does not align with their mission and context. In recognition of this incongruity, [several states have designed and implemented alternative school accountability systems](#) that do align with the mission and context of alternative schools—systems that recognize the successes of these schools, but holding them accountable when they fall short.

Therefore, we recommend that a reauthorized Elementary and Secondary Education Act allow states to develop accountability models for alternative schools, and that this provision be informed by the fine work of the aforementioned states in this regard.

Therefore, we recommend that a reauthorized Elementary and Secondary Education Act allow states to develop accountability models for alternative schools, and that this provision be informed by the fine work of the aforementioned states in this regard. This modification of the law would bring more balance to state accountability systems.

Bring Back District Accountability

Schools have been the accountability focus for 20 years, due to an implicit theory of action holding that, as the most proximal locus of change, schools should be directly held to account. But schools operate within districts, which are the legal entities controlling finances and most other governing policies affecting school quality and student outcomes. Yes, district leaders are involved when a school is identified for additional support, but that is different from districts being the direct focus of accountability. Balancing accountability incentives and consequences among schools, districts, and the state will ensure that each entity does its part to realize equitable and excellent student performance.

The district accountability provisions in NCLB, for example, treated districts like oversized schools—using the same indicators, and the same rules for combining these indicators as an overall rating. If district leaders are to be accountable for their schools' student achievement and equity status, the corresponding accountability system should not be limited to school-related indicators. For example, we would include such district-level indicators as the distribution of key resources across schools: high-quality teachers and leaders, curriculum materials, and professional learning programs, to mention a few. The performance of a district's portfolio of schools should certainly count in a district accountability system, but district-level indicators would put the focus directly on the equitable distribution of resources and opportunities across these schools.

Balancing accountability incentives and consequences among schools, districts, and the state will ensure that each entity does its part to realize equitable and excellent student performance.

Holding States Accountable to Support a Balanced Accountability System

States have the constitutional responsibility to ensure that students receive a proper or adequate education (the language differs across states). Consistent with our theme of balanced accountability, a state accountability system should share some of the indicators we would expect to see in a district accountability system, such as the distribution of key resources, teacher quality, and other factors related to school functioning and student performance. For example, a recent [report](#) from Education Week found that 26 states received F grades on school spending. Shouldn't a state be held to account if, other things equal, certain districts have twice the per-pupil expenditures as other districts?

Moving Accountability Beyond Performance on the State Assessment

A balanced accountability system comprises additional indicators of school quality and holds states and districts accountable for their unique responsibilities in the K-12 education system. This evokes the early days of [standards-based reform](#), where multiple components of a system—content standards, performance standards, and assessments—ideally worked in concert to improve student learning at scale. Early [advocates and researchers](#) also pushed for “school-delivery standards,” later

referred to as opportunity-to-learn standards, to ensure that schools had the necessary resources to succeed. Unfortunately, school delivery standards fell by the wayside prior to NCLB, and schools were left with the entire burden to meet students' needs. We recommend bringing opportunity-to-learn indicators back into a balanced accountability system.

INNOVATION

Innovative Assessment

The Innovative Assessment Demonstration Authority (IADA) was a good first attempt to introduce innovative assessment into federal education law. But like most first attempts at innovation, the IADA fell a bit short (see this [post](#)). Thinking ahead to the next reauthorization of the Elementary and Secondary Education Act (or even U.S. Department of Education waivers), we build on our suggestions in a previous [post](#) for better supporting innovation.

There is an inherent tension between innovation and scaling a reform statewide. The more innovative, the harder it will be to scale in a short timeline. We [wrote](#) about some considerations for scaling, based on the work of [Coburn and colleagues](#), but spreading educational reforms is complex and challenging. Therefore, we recommend allowing states to conduct a limited number of innovations so the state can see what is working and what is not, without the pressure of bringing on more schools and districts. Similarly, states should not be required to select a single innovation for scaling. This way, state and district leaders can tailor innovations to specific contexts and not worry about arbitrary timelines for scaling to a single statewide system.

We were pleased to see that the U.S. Department of Education (ED) proposed a competitive grant program to support states that were approved for an IADA as well as for those states planning to apply. To its credit, ED could pool unused funds for this purpose. We recommend fully authorizing this type of support for successful IADA applicants, and increasing the funding. States, consequently, could focus on the reform without worrying about raising money to support their work.

Finally, we recommend that ED consult with assessment experts to obtain technical advice on critical reform-related issues but in a way that recognizes how the innovative assessments differ from traditional standardized tests. For example, such experts could weigh in topics such as what is “comparable enough” to meet the general assessment and accountability requirements of ESSA while allowing for innovation? As an aside, some of us at the Center have wondered why ED lacks a standing technical advisory committee, given their charge to develop technical regulations and guidance (e.g. assessment regulations, peer review guidance) and evaluate states with respect to these criteria.

A balanced accountability system comprises additional indicators of school quality and holds states and districts accountable for their unique responsibilities in the K-12 education system.

There is an inherent tension between innovation and scaling a reform statewide. The more innovative, the harder it will be to scale in a short timeline.

Innovative Accountability

States should be able to experiment with alternative accountability systems that hold schools accountable, and taking full advantage of IADA-specific data. Innovation is impeded by subjecting pilot schools to the same accountability system as non-pilot schools.

However, going beyond states participating in the IADA for assessment, we encourage authorizing a more general “accountability IADA” (assuming our first major accountability recommendation is not adopted) to allow states to experiment with different types of accountability systems among schools and districts to better support continuous improvement efforts.

RESEARCH AND EVALUATION

We need more disciplined inquiry into standards-based accountability to steepen the shallow slope of improvement regarding student outcomes and the achievement gap since NCLB’s passage. Toward that end, we recommend that the next reauthorization of the Elementary and Secondary Education Act include a robust research-to-practice agenda—and the funding to support it. Ideally, some of this work should be done now to inform the next reauthorization; more realistically, the next federal education law should allocate significant funding for research and evaluation.

We offer examples of high-priority research and evaluation needs below. To frame a more comprehensive agenda, however, we recommend convening major research and state practice organizations, perhaps through the National Research Council, the National Academy of Education, and/or CASSO.

1. There is little research on the types of indicators and measures that best support the school improvement we hope to see, and this research should be conducted in a variety of contexts and under a variety of conditions.
2. Evaluations of adopted policies and practices, both large-scale and small, are needed to understand what is working and what is not, and for what stakeholders.
3. Users of assessment and accountability results struggle to make sense of the information; they particularly struggle with the “what next?” While the quality and interpretability of score reports, both at the individual and aggregate level, have improved considerably, we often hear that users do not understand what these reports mean and, therefore, what actions they should take. We recommend that both states and the federal government support intensive research and development into the most effective score reporting practices for the multiple stakeholder audiences.
4. We noted above that the field still struggles with the design, and especially implementation, of balanced assessment systems. We outlined the beginnings of a research and practical agenda in our [balanced assessment systems paper](#). We recommend using this agenda as a starting point for a larger and more comprehensive research agenda, funded through the ESEA reauthorization, into balanced assessment systems.

We recommend that the next reauthorization of the Elementary and Secondary Education Act include a robust research-to-practice agenda—and the funding to support it.

5. Finally, the U.S government, generally through the National Science Foundation, used to fund the creation of high-quality mathematics and science curricula. Now that we better understand the central role of curriculum in balanced assessment systems, a potent line of research and development would work toward the creation of high-quality curriculum, with embedded assessments, to form the crux of a balanced assessment system (at least at the district level).

CLOSING THOUGHTS

Assessment and accountability can play an important role in promoting better outcomes for students. But they are not a singular prescription for education reform. Moreover, the value of assessment and accountability in this regard is far broader than the limited aspects that should be specified and regulated at the federal level. For this reason, we have argued for the next reauthorization of ESEA to address some issues that stifle innovation, but also to help “rightsize” the role of summative assessment and school accountability—within a more comprehensive and balanced system, emphasizing the centrality of teaching and learning in American public schools. Our recommendations are intended to better help support the original ESEA mandate of improving educational opportunities for students who traditionally have been shortchanged by our nation’s educational system.

In this piece, we offered some initial thoughts related to assessment and accountability recommendations. We invite our colleagues and partners to join us in continuing to expand and refine these recommendations. We remain optimistic about the opportunities offered by the next authorization of ESEA.

Our recommendations are intended to better help support the original ESEA mandate of improving educational opportunities for students who traditionally have been shortchanged by our nation’s educational system.

ⁱ This paper was produced under a [Creative Commons Attribution 4.0 International \(CC by 4.0\)](https://creativecommons.org/licenses/by/4.0/) that allows the user and all subsequent users:

- ✓ To copy and redistribute the material in any medium or format
- ✓ To remix, transform, and build upon the materials for any purposes, even commercially

These permissions are granted as long as the user gives appropriate credit (i.e., attribution) to the original work, provides a link to the license, and indicates if changes were made. For more details on the legal foundation of the license, see: <https://creativecommons.org/licenses/by/4.0/legalcode>.



National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

www.nciea.org