AEI | CERN
Conservative Education Reform Network

Reflections on the Future of K–12 Assessment and Accountability

By Max Eden

July 2021

Key Points

- Education leaders in the Conservative Education Reform Network's working group on assessment and accountability believe that comparable statewide assessment data in reading and math are an important part of determining school performance in helping students learn but also that the structure and format of current state tests should be improved.
- Accountability systems should not only highlight low-performing schools but also recognize and reward high-performing schools, particularly for progress with student subgroups.
- Future assessment and accountability systems, especially at the high school level, should measure how well students are prepared for success after high school rather than focusing on only math and reading proficiency.

In March and April 2021, conservative education leaders representing state education agencies, charter networks, foundations, think tanks, research organizations, testing companies, universities, and teachers convened in a private, three-part working group to discuss the future of state assessment and accountability systems. Our guiding questions were on:

- Conservative principles on assessment and accountability: What should be our key principles when crafting assessment and accountability policies?
- Overview and evaluation of current problems with assessment and accountability systems: What federal and state laws, regulations, and requirements stifle desired innovation and change?
- Recommendations for improving assessment and accountability: Given these problems and principles, what recommendations do conservatives have for federal policy to support desired state and local innovations and changes related to assessment and accountability?

While these are evergreen education policy questions, they take on new urgency after COVID-19. When students return to school in fall 2021, most states will not have run their assessment and accountability systems for two full school years. No new schools will have been identified for improvement in the past two full school years. All states will have limited or no statewide assessment data for the past two school years. This pause has led many education leaders to question current testing and accountability regimes, their utility, and what changes, if any, should be made to these systems.

The conversation among participants was nuanced and not always straightforward. Participants discussed the difficulty of delineating between actual and perceived problems with accountability, a difficulty compounded by perceived problems' propensity to become actual. While accountability

skeptics routinely decry "high-stakes" testing, in the Every Student Succeeds Act era, these tests' stakes are minor. But if school leaders perceive the stakes to be high, and organize schooling around standardized test score growth, then the accountability system is, in practice, high stakes.

Many working group members insisted current policies weren't necessarily the problem as much as their implementation and opponents' messaging campaigns against them were, which discussants expressed difficulty in countering.

Conservatives should be clear about this message: Uniform statewide assessments are necessary to provide information about student performance.

As the conversations progressed, some contradictory viewpoints were put forward and not resolved to the point of consensus. But these meetings produced several conclusions about the future of accountability and assessment systems:

- Comparable, statewide assessment data in reading and math are an important part of determining school performance in helping students learn, but state tests' structure and format could be improved.
- The federal government should set light parameters for state assessment and accountability systems, as state and local bureaucracies add more complexities and requirements to these systems.
- Accountability systems should not only highlight low-performing schools but also recognize and reward high-performing schools, particularly for progress with student subgroups.
- Future assessment and accountability systems, especially at the high school level, should measure how well students are prepared for success after high school, including in the workforce, military, or postsecondary education, and if students are globally competitive, rather than limiting the focus to math and reading proficiency.

Conservative Principles on Assessment and Accountability

When articulating conservative principles to guide policies on assessment and accountability, the working group's views diverged. However, when looking at assessment policy, almost all participants valued statewide assessments that measure the extent to which all students, including student subgroups, master skills. Assessments play an important role in calling balls and strikes and providing an objective, comparable look at student achievement across a state and district, ensuring academic progress in the nation's schools.

While the national focus is on providing a "well-rounded" education to students, including prioritizing social, emotional, and mental health needs (which participants felt was important), there is no substitute for statewide objective information on students' academic achievement. Conservatives should be clear about this message: Uniform statewide assessments are necessary to provide information about student performance. There is no substitute for measuring how schools are serving all students academically. But state assessments cannot provide real-time data for teachers or solve all the problems of teaching and learning in daily classroom instruction. Therefore, formative assessments are also important to help teachers tailor instruction to students' needs.

Communicating assessment results is also an important tool that policymakers rely on, from the school to federal levels, to show how systems serve students. As K–12 education is one of states' biggest expenditures, state and local policymakers often use test scores (in addition to accountability system results and school ratings) to advocate for (or against) increases in spending for education systems.

Many in the working group felt strongly that such application of test results was improper and that assessment scores should not be used to decide how to allocate state and local resources. It is not good policy to link poor performance with more

resources. We should not be rewarding poor performance. These schools often have ample resources yet use these resources poorly or wastefully.

When discussing conservative principles related to accountability systems, many working group members thought the current punitive nature of state accountability systems was problematic. People perceive these systems as overbearing and not reflective of happenings in daily classroom instruction—that you are punished if you do not receive the highest school rating. Accountability systems need to incentivize educators and schools to improve, and the current structures and models have limited rewards. The word "accountability" denotes a negative consequence for doing something wrong.

At the same time, the idea that schools are punished if they receive a low school rating is not accurate either. In the real world, many members of the working group thought, there are limited consequences for poor performance.

Some argued that changing the language could ameliorate this problem. Some suggested saying "incentive systems" instead of "accountability" could mitigate angst and combat misconceptions of the current systems' consequences.

Other participants insisted that transparency combined with choice—rather than policy-based "interventions"—was the ultimate form of accountability. A family's ability to select a school based on various factors it cares about is the ultimate form of accountability, in which individuals enroll in schools based on their priorities. If the system allowed parents to opt in and out of options—private, magnet, home, charter, and traditional schools—that would be true accountability, because funding would be driven by students' school choices. Those decisions would also be driven by a host of factors outside of just reading and math scores, incentivizing a well-rounded and excellent learning environment.

However, the group recognized that systems that provide information on a school's performance in improving student achievement and the quality of its educational experience are important. Even in a system in which school choice is not an option, the default choice needs to be a good one, and transparency of school performance must be available.

The working group discussed two final points on conservative principles related to assessment. First, working group participants felt it was important that conservatives correct negative rhetoric and lies about assessment and accountability's utility. Most conversations conflate these topics and inject false notions of what the federal government requires in assessment and accountability. It is important to correct false statements on the importance of and requirements around assessment results and accountability systems. These systems are valuable in illuminating achievement gaps among a state's student subgroups.

Second, the group discussed a potential accountability system in which states and school districts are rewarded with increased flexibility from state and federal mandates and requirements when results and progress in schools are shown. If a school, district, or state is producing results, it should earn autonomy to innovative more. This would complement an incentive-based system that rewards successes rather than punishing failure.

Current Problems with Assessment and Accountability Systems

The discussions highlighted the false narrative about problems in education or with assessment and accountability systems that contrasts with the reality of what is (or is not) required in these systems, particularly from the federal level. Special interest groups have often made up or overgeneralized problems with assessment and accountability systems to accomplish a policy goal or instill anger in people, with little evidence behind the claims.

Working group members felt that conservatives should work hard to correct false assumptions about tests. The most dangerous new line of attack, participants noted, is that tests are inherently racist. Given that some of the most prominent public intellectuals on the left are now promoting this claim, conservatives have a special duty to defend objective assessments from the nihilistic argument that all efforts at objectivity are inherently racist.

Education stakeholders are certainly split on the value of high-stakes tests, but the many reasons for this divide should not be overtaken by progressive language and views about testing.

Several state education chiefs felt strongly that bureaucracy at all levels—federal, state, and local—inhibits innovation. At each level of leadership, more constraints and requirements are added to education systems. For example, state legislatures can add parameters to constrain accountability systems and dictate rigid timelines about what and when assessments are given to students. Any new ideas related to assessment and accountability must be voted on and approved by statute in some states. Finally, state procurement processes for assessment vendors are also entangled in bureaucracy.

Some of the state chiefs described slow, outdated procurement processes to engage new vendors; antiquated budget approval procedures; a large lack of workforce capacity in procurement offices; and little ability to conduct oversight to ensure vendors follow through on agreements. A strong argument for minimizing the scope of federal statutory requirements around assessment is that implementation will inevitably lead to complications that can stifle innovation.

The working group also discussed specific problems related to assessment and accountability in the Elementary and Secondary Education Act (ESEA)—the nation's signature K–12 education law—that inhibit new ideas in these areas. Several members mentioned that the Innovative Assessment Demonstration Authority (IADA) in ESEA was completely unworkable for states that wanted to try something new in assessment and accountability. The IADA aimed to allow for "innovation" in assessment and accountability, particularly to give states the opportunity to use new types of tests. In reality, the prescriptive nature and magnitude of the IADA's requirements inhibited innovation. The requirement for states to demonstrate comparability between any new test and the traditional state test and the requirement to scale the new tests statewide within five years (with a potential two-year extension) pose serious threats to real innovation.

State chiefs in the working group stated that the options in the IADA regulations from which states could choose to prove comparability forced the new tests to be almost exactly the same as the current state tests. The requirement to scale new tests statewide within five to seven years also was too short a timeline to prove any new test was better. Many state chiefs felt the risk and high stakes of investing time and state funds into a new assessment system with rigid requirements that the US Department of Education (ED) could take away authority for at any time were not worth it. Members expressed that asking for waivers from specific requirements of ESEA rather than pursuing the IADA seemed like a more beneficial approach to assessment and accountability innovation. Relatedly, many state chiefs are pursuing new types of testing on their own while still running current assessment systems just to avoid burdensome and often unhelpful oversight from ED and still comply with ESEA.

Some participants also raised how ESEA's requirements to test all students at grade level might inhibit new assessment systems in which students progress along a continuum as they develop certain skills. Grade-level testing might not enable students above or below "grade-level standards" to appropriately advance in their coursework and skill development. Thus, limiting content to a grade level might limit students' abilities to grow. Admittedly, the group was not clear how anything other than a grade-based system would create accountability in the current structure.

Also related to ESEA requirements, discussion arose about whether annual testing for every student in the state was necessary. There was no group consensus on this question. The group's researchers and a few state chiefs felt annual assessment data were important to collect, but others shared that testing every other year to measure student growth toward proficiency would be sufficient for an accountability system.

All participants mentioned that perhaps the main problem with state assessments was not ESEA's requirements but the peer-review process ED engaged in with states to approve each state's test. Many felt the ED peer-review process stifled any ability to try something new and that the arguments with ED staff over any slight changes or deviations to state tests were not worth the effort to change them. Most members felt the ED staff that reviewed state tests overreached in their back-and-forth with states on test approval, which greatly inhibited innovation.

Consensus seemed to emerge that current assessment systems and the federal and state requirements around them are more problematic than current accountability systems are. Related to assessments, participants expressed that state tests do not measure the things we say we value. We say we value using data and assessment results to drive improvements in teaching and learning, but summative assessments do not provide any helpful information to move toward that goal. We say we value personalized learning approaches in which curriculum is adjusted based on student skills and teachers can teach students based on their current needs, but state summative tests do not help advance these approaches.

Instead, state summative tests are expensive, time-consuming, and feel "high stakes" for school districts and schools. Perhaps de-prioritizing state test results in accountability systems could help state chiefs move to more useful tests, although participants largely felt that standardized test scores were an important aspect of accountability systems—just not the only or most important one. Some members expressed that assessment systems could be built that serve dual purposes.

A system with which many states are experimenting would administer shorter tests over the course of a school year that embed curriculum items the students learned in daily instruction. The tests' results would become a summative score for accountability purposes. Each test would not be high stakes but would measure the breadth and depth of standards over a school year. In this system, assessment results could help inform and adjust instructional practices in real time and inform how well a school is doing at increasing student achievement.

Participants also discussed significantly shortening the end-of-year summative tests and focusing more on formative assessment throughout the year. Some felt the state summative tests could be much shorter and still evaluate how well a school academically serves students.

Lastly, although there was a general recognition that current assessments were more problematic than current accountability structures are, the working group discussed that in the future, if new ideas or best practices emerged in either area, new policy should not be limited to fitting new ideas into current constructs. For example, one problem with current assessment and accountability systems is their heavy emphasis on math and reading test scores. More and more education reformers are interested in the broader measurement of whether schools are preparing students for success after high school. Perhaps more emphasis should be placed on figuring out how to adequately measure student outcomes after graduation.

Post-COVID-19, if effective initiatives emerge in measuring and closing achievement gaps or if there are new indicators that can better evaluate a school's ability to bring students to proficiency or prepare them for success beyond high school, those efforts should be seriously considered as part of a next-generation assessment and accountability system. No members thought the current system is perfect. Should better ideas emerge, a complete system redesign should be permitted. Participants asked, "How do we create a system that ensures disadvantaged students are not receiving a subpar education while allowing for innovation that could help all students?" Both are important.

Recommendations for Improving Assessment and Accountability

Just as the working group often had difficulty pinpointing specific (versus perceived) problematic requirements in federal and state laws related to assessment and accountability, so, too, did it struggle with articulating specific solutions about what comes next.

As long as teachers and parents perceive assessments as high stakes with accountability consequences, they will never embrace testing. Focusing on solutions, working group members expressed that the federal government should not overregulate assessment and accountability or get into the business of deciding what is "good enough" progress for school districts and schools. Federal law should have minimal parameters because bureaucracy at the state level will complicate system designs when multiple actors, such as governors or state boards of education, add ideas to assessment and accountability systems beyond federal requirements. The federal government should let states design their own assessment and accountability systems with minimal parameters, allowing states

to decide what is acceptable progress, and let them design an entirely new, different system if need be.

The federal government should recognize that it cannot adequately oversee solutions in every US school. Participants differed on whether federal requirements and the federal government should focus mostly on the lowest-performing schools and students. A few participants proposed having two distinct federal and state accountability systems focused on different purposes. The federal accountability system would be focused on ensuring low-income students and student subgroups, including students with disabilities and students of color, are progressing toward established goals. All federally required consequences would target students who are not hitting their goals (rather than whole schools).

The state accountability system would then be more nuanced and differentiated, focusing on all schools in the state. States could reward high performers and institute a series of differentiated interventions, designed with school districts, to improve schools across a range of performance levels. States would set standards and determine what scores would be deemed successful on a range of indicators and treat schools according to their performance.

The conversation also touched on how hard it is to have differentiated intervention strategies when K–12 education is not a free market. If families have no school choice for their children, there is no consequence (or incentive) for schools to perform, since students will enroll regardless of performance. Participants discussed that a differentiated state accountability system could treat schools of choice differently with different interventions and consequences based on performance. Additional consequences and intervention from the state for lackluster performance should not be necessary for schools that parents chose to send their children to.

Participants again discussed the purpose of federal accountability. The current systems focus heavily on math and reading proficiency but not on the larger question of how students are prepared for post–high school opportunities and to be globally competitive. Participants felt that accountability systems should be totally revamped at the high school level. Traditional standardized state tests are not that useful in high school, but serious work needs to be undertaken to figure out how to measure future student success. The federal government can help by facilitating or seeding funding for better data systems and data sharing among the workforce, the military, and the education institutons, but it should not set goals of what post–high school success looks like.

Relatedly, participants discussed the important federal role of providing transparency to parents and families about student and school performance. A school should ultimately be accountable to its students' families, and the federal government can take a larger role in helping schools and districts be transparent about performance data. Participants discussed how the federal government could provide more guidance on how to design school, district, and state report cards and present school performance information in a uniform way across the country so everyone can understand it. Much more work needs to be done in this area. This transparent, comparable information for families across the country could also bolster school choice because parents could easily and adequately compare schools across a state (or even beyond). Improving transparency around school performance is an important federal role that needs more attention.

Improving transparency around school performance is an important federal role that needs more attention.

The conversation focused on changing the perception that accountability is negative or something "being done" to a school. While the word "accountability" denotes consequences, few exist for poor-performing schools. Participants felt strongly that a federal system, while perhaps focusing requirements for interventions on the lowest-performing schools or students, should create incentives and rewards for high-performing schools, particularly ones accelerating growth for the students most behind. It is easy to demagogue rewarding success because of the notion that these schools "don't need help."

Participants felt strongly that we should move away from this perception. They acknowledged that, from the federal level, rewarding "success" while focusing on the lowest-performing schools might be hard to do, but it should be considered. Proposed examples of rewards and incentives included providing more funds to successful schools, particularly when they showed growth in the lowest-performing students. Additionally, schools could get "time and space" from requirements, oversight, and bureaucracy when they made progress over time for all students, including disadvantaged students.

A conversation emerged around whether flexibility is really a reward for success. Some participants felt that states, school districts, and schools say they want freedom and flexibility from requirements, but when given the opportunity to do something different, they always do the same thing. However, many participants felt strongly that incentives, bonuses, and flexibility work, can seed the ground for future system changes, and can create examples of best practices for others to emulate.

Some additional ideas around rewards emerged, including allowing all schools to maintain their current Title I funding levels but requiring any new funds to go to high performers. Another idea was to associate rewards with high-performing National Assessment of Educational Progress scores that align closely with state scores. It is not easy to design a system that focuses on the lowest performers but recognizes high-achieving schools and those making large gains for students. However, many participants felt future accountability systems should try to do this, even if there were differences between federal and state accountability systems.

Lastly, participants cautioned that the federal government should not determine what school success looks like but leave states to set these metrics while perhaps prompting some areas to examine (such as how quickly achievement gaps are closing).

The final discussion focused on current summative assessments and the future of state testing. Participants felt strongly that the heavy reliance and focus on state math and reading tests are because they are the only fair, valid, and objective metrics to compare achievement across a state. Working group members felt this should shift to emphasizing content mastery and personalized learning. If so, assessments would likely have to change from testing students by grade level (discussed previously as a problem with the current system) to testing students along a continuum of skill instruction, regardless of grade level. Parents and teachers might embrace testing if assessment systems changed in this way, because they would see the instructional value for their students.

Accountability is still needed, however, so measuring student achievement is needed too. Participants felt that, as described earlier, tests could be redesigned to be shorter, less frequent high-stakes summative tests of student progress over the course of a year. Alternatively, the summative test could have lower stakes in accountability systems, allowing more incentives and resources to be used to improve classroom instruction and develop high-quality formative tests to inform daily activities.

Time and resources could also be focused on improving formative tests while removing the requirement to test every student in math and reading annually. A sampling of student test scores in math and reading across every school district (or even school) annually or biannually might be sufficient for "system check" accountability purposes. Either way, assessment and accountability systems need to rebalance formative assessment to inform daily instruction and need summative assessments to see how a school system is improving student performance.

Conclusion

The intricacies and complexities around conservative views on assessment and accountability were elicited over these three working group sessions. No easy answers are readily apparent, and perhaps new ideas should not have to fit into today's structures. Starting over completely might be the best bet in certain circumstances, learning from and building on past experiences.

A few common themes emerged.

- Uniform, statewide assessment data are vitally important to determine if schools are serving students well, but it might be time to rethink state tests' structures and formats.

- The federal government should recognize and respect that state-level bureaucracies add multiple layers of complexity to assessment and accountability systems and be minimalist in its approach to system requirements.
- Recognizing, rewarding, and incentivizing high-performing schools, particularly for progress with student subgroups, should be part of future accountability systems.
- Future accountability and assessment systems, especially at the high school level, should focus more on measuring if students are globally competitive and how well students are prepared for success after high school in whatever path they take.

The final point is vital to the US economy's long-term success. Conservatives should act and speak out when misperceptions regarding assessment and accountability are spread to mainstream media. It is important to correct the record on what conservative values exist in the current policy environment and correct falsehoods about assessment and accountability systems.

It will be interesting to see how states embrace change and new ideas coming out of a global pandemic. The opportunities for innovation are there if the federal government steps back and allows originality to flourish.

About the Author

Max Eden is a research fellow at the American Enterprise Institute and director of the AEI Conservative Education Reform Network.