

# Developing State Assessment Systems That Support Teaching and Learning: What Can the Federal Government Do?

**Aneesha Badrinarayan and Linda Darling-Hammond**

In collaboration with Michael DiNapoli, Tara Kini, Tiffany Miller, and Julie Woods

## Summary

The Every Student Succeeds Act (ESSA) invited states to use multiple measures of “higher-order thinking skills and understanding,” including “extended-performance tasks,” to create state assessment systems that support teaching for deeper learning. However, few states have been able to navigate federal assessment requirements in ways that result in tests with these features that can support high-quality instruction. This brief describes **three ways that federal executive action can help states realize their visions for more meaningful assessments:**

1. Better align technical expectations for assessment quality with ESSA’s intentions
2. Enable ESSA’s Innovative Assessment Demonstration Authority to better support innovation
3. Create additional pathways to higher-quality assessments through existing or new funding mechanisms

The report on which this brief is based can be found at <https://learningpolicyinstitute.org/product/developing-assessment-systems-federal-support>.

## Introduction

There is a growing call to reconsider current approaches to national and state assessment system policies and practices. State and local education agency leaders, educators, community leaders, and advocates have voiced concerns that our current state assessment systems—defined primarily by end-of-year multiple-choice tests—are unable to meet contemporary needs for information that supports teaching and learning. The need has grown more acute as schools seek to help students recover from the impacts of the pandemic on learning and achievement.

More than 20 states are involved in efforts to transform one or more aspects of their assessment systems; however, the process of securing federal assistance and approval to make transitions to substantially improved systems poses numerous challenges. Among them are the costs and time required to change systems, the management of trend disruptions when new assessments are introduced, and interpretations of how to meet federal approval criteria under business rules that often keep new tests looking very much like old ones.

This brief synthesizes policy analyses and findings from legal and research analyses, as well as consultations with national, state, and local leaders, to identify key ways in which the federal government could support reforms that enable thoughtful assessment of meaningful skills in ways that also better support teaching and learning.

## **The Every Student Succeeds Act: Opportunities and Barriers for Meaningful Assessment Systems**

In 2015, the Every Student Succeeds Act (ESSA) opened new possibilities, relative to the prior decade under No Child Left Behind, for how student and school success are defined and supported in American public education. The language in the law (see Table 1) deepened the concept of student learning to be more consistent with what students need to be successful in 21st-century society and careers, calling for measurement of “higher-order thinking skills and understanding” as part of “high-quality student academic assessments in mathematics, reading or language arts, and science.” ESSA intentionally created opportunities for assessment innovation by explicitly allowing the use of multiple types of assessments, including “portfolios, projects, or extended-performance tasks,” as part of state systems.

In addition to its statewide assessment provisions for innovation, ESSA explicitly allows a subset of states to pursue innovation through the Innovative Assessment Demonstration Authority (IADA). This provision invited up to seven states to implement new approaches to assessment and gradually scale them statewide. IADA defines innovation flexibly, allowing for state systems that may include competency-based assessments, curriculum-embedded performance assessments, and through-year assessment approaches. The primary promise of IADA is that it provides states a means to pilot new assessments by allowing a subset of districts to use the new assessments rather than the old ones, without double testing students. This is an important feature of reform in those jurisdictions, where states are not simply substituting one commercially administered standardized test for another at the end of the school year.

While many states were initially pleased to have the opportunity to explore the flexibilities in the law through IADA, the process of applying for and complying with the terms of the waiver have proved to be so onerous and constraining that few states have yet been able to use IADA to implement innovative assessment designs. Fewer still have been able to develop systems that provide insights into student learning in ways that are particularly meaningful to teaching, as originally envisioned by ESSA.

**Table 1: Summary of Assessment Requirements Under Section 1111 of the Every Student Succeeds Act**

SEC. 1111. STATE PLANS. 1111(b)(2)(B) Assessment Requirements	
<b>Assessments generally</b>	<p>(i) except as provided in subparagraph (D), be—</p> <p>(I) the same academic assessments used to measure the achievement of all public elementary school and secondary school students in the State; and</p> <p>(II) administered to all public elementary school and secondary school students in the State;</p>
<b>Information produced</b>	<p>(ii) be aligned with the challenging State academic standards, and provide coherent and timely information about student attainment of such standards and whether the student is performing at the student’s grade level;</p> <p>(xi) enable results to be disaggregated within each State, LEA, and school by—</p> <p>(I) each major racial and ethnic group;</p> <p>(II) economically disadvantaged students as compared to students who are not economically disadvantaged;</p> <p>(III) children with disabilities as compared to children without disabilities;</p> <p>(IV) English proficiency status;</p> <p>(V) gender; and</p> <p>(VI) migrant status...</p> <p>(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;</p>
<b>Technical requirements</b>	<p>(iii) be used for purposes for which such assessments are valid and reliable, consistent with relevant, nationally recognized professional and technical testing standards, objectively measure academic achievement, knowledge, and skills, and be tests that do not evaluate or assess personal or family beliefs and attitudes, or publicly disclose personally identifiable information;</p> <p>(iv) be of adequate technical quality for each purpose required under this Act and consistent with the requirements of this section, the evidence of which shall be made public, including on the website of the State educational agency;</p> <p>(xiii) be developed, to the extent practicable, using the principles of universal design for learning.</p>
<b>Measurement methods</b>	<p>(vi) involve multiple up-to-date measures of student academic achievement, including measures that assess higher-order thinking skills and understanding, which may include measures of student academic growth and may be partially delivered in the form of portfolios, projects, or extended-performance tasks;</p>

SEC. 1111. STATE PLANS. 1111(b)(2)(B) Assessment Requirements	
<b>Administration flexibility</b>	(viii) at the State’s discretion— (I) be administered through a single summative assessment; or (II) be administered through multiple statewide interim assessments during the course of the academic year that result in a single summative score that provides valid, reliable, and transparent information on student achievement or growth;
<b>Assessment flexibility</b>	Locally selected, nationally recognized option for high school (H)
<b>Assessment design flexibility</b>	Computer adaptive assessments: (I) subparagraph (B)(i) shall not be interpreted to require that all students taking the computer adaptive assessment be administered the same assessment items; and (II) such assessment—“(aa) shall measure, at a minimum, each student’s academic proficiency based on the challenging State academic standards for the student’s grade level and growth toward such standards”; and “(bb) may measure the student’s level of academic proficiency and growth using items above or below the student’s grade level, including for use as part of a State’s accountability system under subsection (c).”
<b>Standards</b>	(A) IN GENERAL.—Each State, in the plan it files under subsection (a), shall provide an assurance that the State has adopted challenging academic content standards and aligned academic achievement standards (referred to in this Act as “challenging State academic standards”), which achievement standards shall include not less than 3 levels of achievement, that will be used by the State, its local educational agencies, and its schools to carry out this part.

Source: Every Student Succeeds Act. (2015).

## Calls for Assessment Systems That Better Support Teaching and Learning

Many state leaders want to transform their state assessment systems to take advantage of ESSA’s affordances—with or without IADA. Through a series of conversations with state and local leaders as well as teachers and partners in the education space, a common set of goals for assessments that can inform and improve teaching and learning in schools is emerging. These common goals are:

- **Assessment tasks should encourage applied learning and higher-order skills.** Statewide assessments should prioritize engaging, realistic tasks that promote and support better teaching and learning—and, ultimately, provide better information about student progress.

- **Assessments should be integrated into a system that supports high-quality teaching and learning.** To support meaningful, deeper learning for students, assessment systems should be both designed and used as part of a coherent, well-integrated system of curriculum, instruction, and professional learning for teachers.
- **Assessments should be part of accountability systems that support student access and success.** Assessments should be part of improved accountability systems designed to encourage behaviors and actions that lead to a more informed focus on school improvement, more equitable access to learning opportunities, and greater student success.

States are pursuing several promising approaches to address many of these needs. Some of these approaches strengthen the capacity for high-quality instruction informed by formative assessment in schools within integrated curriculum frameworks that also inform summative assessments. Other approaches position more innovative assessments to become part of the state summative assessment process itself when sufficient comparability safeguards are in place. In nearly all efforts, states seek to address the vision and challenges described above in ways that position the state assessment system to signal and incentivize what high-quality teaching, learning, and student performance should look like, while allowing for appropriate flexibility for local decisions.

## Possibilities for Federal Executive Action

States need time, support, and permission to innovate. Several possible federal executive actions could support these goals. Some actions can strengthen all statewide assessment systems, while other recommendations focus solely on strengthening IADA implementation. All actions discussed here are permissible under current federal law.

### Align Technical Expectations and Peer Review Processes With ESSA's Assessment Allowances and Requirements

The U.S. Department of Education's approval process for all state assessment systems is guided by an internally developed and moderated peer review process,<sup>1</sup> used to render judgments about state systems. While ESSA explicitly encourages more instructionally relevant assessment approaches, the peer review process often inadvertently disincentivizes the very kinds of assessments that ESSA encourages. The Department of Education's interpretation of ESSA's assessment provisions in the peer review guidance privileges assessments that are administered once to all students; have many short, quick, grade-level items to maximize standards coverage; and can be rapidly machine scored without needing expertise to evaluate.

If the federal peer review process were updated to incorporate technically strong approaches to higher-quality assessments and data, state assessment programs could then be designed to enable instructionally useful innovations with confidence about their ultimate approval. Following are four recommendations to bring peer review into alignment with the opportunities for instructionally relevant assessment allowable within ESSA.

- **Recommendation 1: Highlight opportunities and update processes to support more instructionally relevant assessments that reflect student performance in relation to both grade-level standards and multiyear learning progressions.** ESSA explicitly allows assessments to measure both grade-level proficiency and learning along a broader continuum. When educators are given information about precisely what students know and can do—not only whether students are meeting grade-level expectations or not—they can accelerate learning by tailoring instruction to help students build the knowledge and skills needed to meet grade-level standards. This can be done efficiently with the advent of computer adaptive testing; however, current guidance requires states to focus almost exclusively on grade-level standards, thus preventing accurate assessment of skills for students above or below grade level and undermining efforts to measure growth. Instead, guidance can emphasize ways to leverage items that sample along multiyear learning progressions to yield results that provide more precise information about what students know and can do, while still maintaining high expectations for all learners and providing information about grade-level achievement to comply with ESSA’s requirements.
- **Recommendation 2: Highlight technically sound approaches to meeting federal peer review requirements that allow state assessments to assess the depth of state standards while ensuring sufficient coverage.** Current requirements for producing subscores and aligning to the breadth of states’ academic content standards yield assessments that are a mile wide and an inch deep. Because there are so many items that assess isolated knowledge and skills, it is nearly impossible for this approach to appropriately measure the sophisticated understanding and abilities that are central to current state standards (e.g., developing complex arguments, solving realistic problems). Rather than prioritizing extensive coverage of the easily tested aspects of standards with many superficial items, guidance can encourage states to sample strategically, allowing space for more sophisticated items and holistic performance tasks that evaluate the complex forms of thinking, disciplinary practices, and performance intended by the standards.
- **Recommendation 3: Update peer review guidance to emphasize requirements for test security that are appropriate to the design of the assessment.** Test security requirements are intended to ensure that assessment scores are trustworthy. However, current requirements for test security are based on tests featuring discrete items that can be easily “gamed” through memorization; thus, the requirements assume that high levels of secrecy about test content are a necessary condition for trustworthy scores. However, test secrecy becomes increasingly irrelevant as the tests themselves involve increasingly authentic performances that students must be able to demonstrate.

For example, the performance task (see Figure 1) in the [Advanced Placement \(AP\) Computer Science Principles](#) course asks students to invent, develop, test, submit, and explain a computer program as part of their final score on the AP exam, which is used to confer college credit to students. This task is widely known and is completed (in part collaboratively) as part of classroom activities, but because students must still have the knowledge and skills to design a computer program, knowing what the task

is does not compromise interpretation of students' performance. Indeed, knowing that students are expected to develop a program incentivizes teaching students how to do so—positioning the assessment as a driver of high-quality instruction while generating scores that are valid, trustworthy, and more representative of what students know and can do than most on-demand assessment designs.

Rather than assuming that the only way to achieve valid scores is to keep all test items secret, guidance could explicitly recognize that some test designs—such as those that use more authentic tasks requiring demonstrations of skills that cannot be memorized—should be subject to different expectations for test security than a multiple-choice test.

**Figure 1: Advanced Placement Computer Science Curriculum-Embedded Performance Task**

## Create Performance Task




Programming is a collaborative and creative process that brings ideas to life through the development of software. In the Create performance task, you will design and implement a program that might solve a problem, enable innovation, explore personal interests, or express creativity. Your submission must include the elements listed in the Submission Requirements section below.

You are allowed to collaborate with your partner(s) on the development of the program only. **The written response and the video that you submit for this performance task must be completed individually, without any collaboration with your partner(s) or anyone else.** You can develop the code segments used in the written responses (parts 3b and 3c) with your partner(s) or on your own during the administration of the performance task.

*Please note that once this performance task has been assigned as an assessment for submission to College Board, you are expected to complete the task without assistance from anyone except for your partner(s) and then only when developing the program code. You must follow the Guidelines for Completing the Create Performance Task section below.*

### General Requirements

You will be provided with a minimum of 12 hours of class time to complete and submit the following:

-  **Final program code** (created independently or collaboratively)
-  **A video that displays the running of your program and demonstrates functionality you developed** (created independently)
-  **Written responses to all the prompts in the performance task** (created independently)

Scoring guidelines and instructions for submitting your performance task are available on the [AP Computer Science Principles Exam page](#) on AP Central.

*Note: Students in nontraditional classroom environments should consult a school-based AP Coordinator for instructions.*

Source: College Board. (2020). *AP computer science principles: Course and exam description*. <https://apcentral.collegeboard.org/media/pdf/ap-computer-science-principles-course-and-exam-description.pdf>



- **Recommendation 4: Revitalize foundational elements of the peer review process—peer reviewer selection and moderation of the process—to ensure that states can take full advantage of the opportunities provided by federal law and the Department of Education’s technical guidance.** Many of the recommendations outlined represent high-leverage changes to the technical guidance provided by the Department of Education that could serve to encourage more innovative assessments. However, two important elements of the federal peer review process must be addressed globally for the prior recommendations to have impact:
  - a. updating the psychometric standards that underlie many notions of quality and sufficient evidence that are embedded within the peer review criteria, and
  - b. expanding the people and processes involved in operationalizing updated guidance.

Without attention to these foundational components of the peer review process, even the most profound changes run the risk of being just words on a page. As assessment technologies have evolved to incorporate innovative measures across many fields, it will be essential that the peer review process be updated. In addition to considering what psychometric conventions are held up as the appropriate standard, the Department of Education could also consider (1) ensuring that the selection of peer reviewers includes experts familiar with these methods, (2) revisiting the moderation and calibration of peer review to ensure application of these standards, and (3) increasing state and partner engagement during the process.

## Enable IADA to Better Support Innovation in Assessment

While the inclusion of IADA within ESSA was first met with excitement by states, this optimism has waned. IADA does not currently offer states enough opportunity and flexibility to make the tremendous effort needed to create new assessment systems worthwhile. In fact, many of IADA’s requirements are viewed as onerous and may actually limit efforts to develop innovative systems. Three additional recommendations, which address how executive action could shift the cost-benefit trade-offs to open opportunities for innovation and remove barriers to state participation, follow.

- **Recommendation 5: Update the interpretation of comparability of results within current IADA regulations to better enable high-quality innovative assessment approaches.** It is essential to ensure that a given assessment provides comparable standards-aligned tasks that generate comparable student scores across students, schools, and districts. However, IADA constrains innovation by requiring comparability of results across the innovative and traditional tests, limiting how much an innovative assessment can differ from the current test, even when the new assessment seeks



to better surface student understanding of state standards (e.g., measuring more complex skills and abilities, addressing standards that are not well represented on current tests).

If comparability is defined as generating the same results across the innovative and current state assessments, the innovative assessment will automatically be constrained by the same design and reporting decisions of the current summative test, including any limitations in surfacing useful student data. This approach positions the current state assessment as producing the “right” results, even as many efforts for innovation seek to produce more meaningful results that better reflect student understanding of state standards and that can be used to support teaching and learning more effectively. For example, many states pursuing assessment flexibilities or innovative assessment systems are seeking to measure deeper learning and other aspects of standards not well represented in their current tests (such as problem formulation; investigation; data set development; and analysis, writing, and speaking). These innovations should be expected, by design, to produce *different* student scores than the tests they would be replacing. Thus, the requirement for comparability of results with lower-level assessments may inadvertently serve to prevent higher-quality assessments from being developed.

There are at least three other ways to conceptualize comparability with relation to state assessments:

- a. Comparability of assessments with respect to the standards they measure. That is, to what degree do different assessments measure the same learning goals?
- b. Comparability of tasks and scores across students and schools. That is, to what degree can students’ assessment scores be interpreted in similar ways across students and groups of students (e.g., across different administration contexts and disaggregated subgroups)?
- c. Comparability of task scoring. That is, for a single assessment (e.g., a performance-based assessment), how likely is it that a student’s essay would receive the same score from two different raters?

Under these conceptualizations, a state seeking to design an innovative, instructionally relevant system of assessments can explore the use of high-quality curriculum-embedded performance tasks in an assessment design that assesses the same standards as the current summative assessment, but does so in a way that allows for students to demonstrate aspects of the standards (e.g., more evaluative and critical thinking, writing ability, and problem-solving capacity) that were not assessable on the current test. Rather than requiring that new assessments produce the same scores as existing tests, guidance could encourage that states submit compelling evidence that their innovative test is of equal or *higher* quality than the

existing assessment, and that it produces comparable scores among students taking the innovative assessment. Doing so will enable states to develop assessments that can better support teachers, leaders, and families in supporting students and their learning. Language for guidance on the fifth option for demonstrating comparability under IADA might be updated to state: “(5) an alternative method for demonstrating comparability that an SEA can demonstrate will provide for a rigorous and valid comparison between the innovative assessment and the statewide assessment, such that:

- a. the innovative assessment is demonstrated to be of equal or higher quality than the state assessment in terms of measurement of academic standards, and
  - b. all tasks and scoring processes on the innovative assessment generate valid and reliable measures of student performance that are comparable across students, schools, and districts engaged in the innovative assessment ....”
- **Recommendation 6: Utilize existing flexibilities and promulgate new regulations to allow for additional time to scale innovative assessment systems statewide.** Innovative systems need space and time to ensure quality of the instruments and appropriate supports for users, as well as flexibility to course-correct during the scaling process. The limited timeline in IADA effectively requires that states have not only a predetermined plan for their innovative assessment, but also evidence and confidence in the functioning of the new system before entering the 5-year pilot rather than allowing for true innovation as part of the pilot. Acquiring this evidence and confidence requires launching a system that states have no guarantee will be approved under federal law, which means that doing so before securing IADA approval (itself no certain guarantee) is not generally practicable. The Department of Education could clarify and update regulations to provide states with additional time for planning, implementing, and scaling innovative systems.
  - **Recommendation 7: Lift the cap on the number of states able to participate in IADA and allow for states to collaborate on assessment designs.** Should IADA become more attractive to states, only three additional states could currently participate. The Department of Education could prioritize completing the required Institute of Education Sciences (IES) report and eliminating the seven-state cap, allowing more states to take advantage of the opportunity.

## Create Additional Pathways to Innovation

While IADA represents one major effort to create opportunities for assessment innovation, there are other ways the Department of Education can signal, incentivize, and support change. For example, the Competitive Grants for State Assessments (CGSA) program has been used to support states and multistate collaboratives in improving their state

assessment systems. This grant program both provides funding and has fewer constraints than IADA, and it may be an effective avenue to support innovative state assessment efforts. The primary way to accomplish this is explained in the recommendation below.

- **Recommendation 8: Use the CGSA program to stimulate individual or multistate efforts to develop and pilot new approaches that are instructionally useful and responsive to the broader view of assessment in ESSA.** The Department of Education could consider further leveraging CGSA—both through larger funding requests and through strategic allocation of funding—to support innovations that specifically target assessment designs that seek to advance better teaching and learning. Strategic use of CGSA could also be used as an on-ramp to IADA, providing financial support to states to develop the systems and technologies needed for their new assessment, as well as support for piloting prior to entering the demonstration period.

## Conclusion

State and local education agency leaders, educators, community leaders, and advocates have voiced concerns that current state assessment systems—defined primarily by end-of-year multiple-choice tests—are unable to meet contemporary needs for information that supports teaching and learning. Federal executive action could focus on short-term strategies to encourage more innovative state assessment systems that better support teaching and learning. These include updating technical guidance provided to states to better align with the Every Student Succeeds Act, enabling the Innovative Assessment Demonstration Authority to better support the innovation it seeks to incentivize, and leveraging additional programs, such as the Competitive Grants for State Assessments program, to foster assessment innovations that lead to assessments that advance teaching and learning.

## Endnotes

1. For a complete description of the process, including complete peer review criteria and examples of evidence, see: U.S. Department of Education Office of Elementary and Secondary Education. (2018). *A state's guide to the U.S. Department of Education's assessment peer review process*. <https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf>

## Acknowledgments

The report on which this brief is based benefited from the insights and expertise of two external reviewers: Randy Bennett, Educational Testing Service, and James Pellegrino, University of Illinois Chicago. We thank them for the care and attention they gave the report. This research was supported by the Carnegie Corporation of New York, Chan Zuckerberg Initiative, and Walton Family Foundation. Core operating support for the Learning Policy Institute is provided by the Heising-Simons Foundation, William and Flora Hewlett Foundation, Raikes Foundation, Sandler Foundation, and MacKenzie Scott. We are grateful to them for their generous support. The ideas voiced here are those of the authors and not those of our funders.

Badrinarayan, A., & Darling-Hammond, L. (with DiNapoli, M., Kini, T., Miller, T., & Woods, J.). (2023). *Developing state assessment systems that support teaching and learning: What can the federal government do?* [Brief]. Learning Policy Institute. <https://learningpolicyinstitute.org/product/developing-assessment-systems-federal-support-brief>

This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.