

A Policymaker's Guide to State Summative Assessment Systems

Ben Erwin and Shayna Levitan

Summative assessments measure students' mastery of grade-level academic standards and skills in specific content areas after learning. State summative assessment systems provide students and families, school and district leaders, and state policymakers with valuable data to help understand student progress and school quality.

Assessment data allow families to understand their student's academic needs, which can inform collaboration with educators to determine necessary supports and interventions. For educators, summative assessments offer year-to-year data on student learning and allow them to adjust their instruction accordingly. These data also provide school, district and state leaders with comparable data to better understand performance within and across schools and identify inequities between student groups.

Assessment data factors heavily into school ratings in state school accountability systems — this data is central to the identification of schools for support and improvement. Assessments provide a valuable data point when conducting school needs assessments and developing improvement plans. Understanding inequitable student outcomes and other differences in performance allows local leaders and state policymakers to target resources and interventions to ensure all students receive the supports they need to succeed academically and to evaluate the effectiveness of these efforts.

Summative assessments are a key component in a balanced assessment system that leverages other assessment types (e.g., diagnostic, formative and interim) to achieve different purposes. For example, summative assessment results are one source of many that help create a holistic snapshot of students' academic progress and their school experiences. As they are currently constructed, summative assessments do not necessarily provide data that supports instruction and academic interventions during the same school year.

When combined with other tools, such as classroom assignments and end-of-unit tests, a balanced assessment system can help create a comprehensive snapshot of student performance that informs students, families, educators, and local and state leaders. States are increasingly adopting approaches to summative assessments that address some of their limitations to increase their [quality and utility](#).

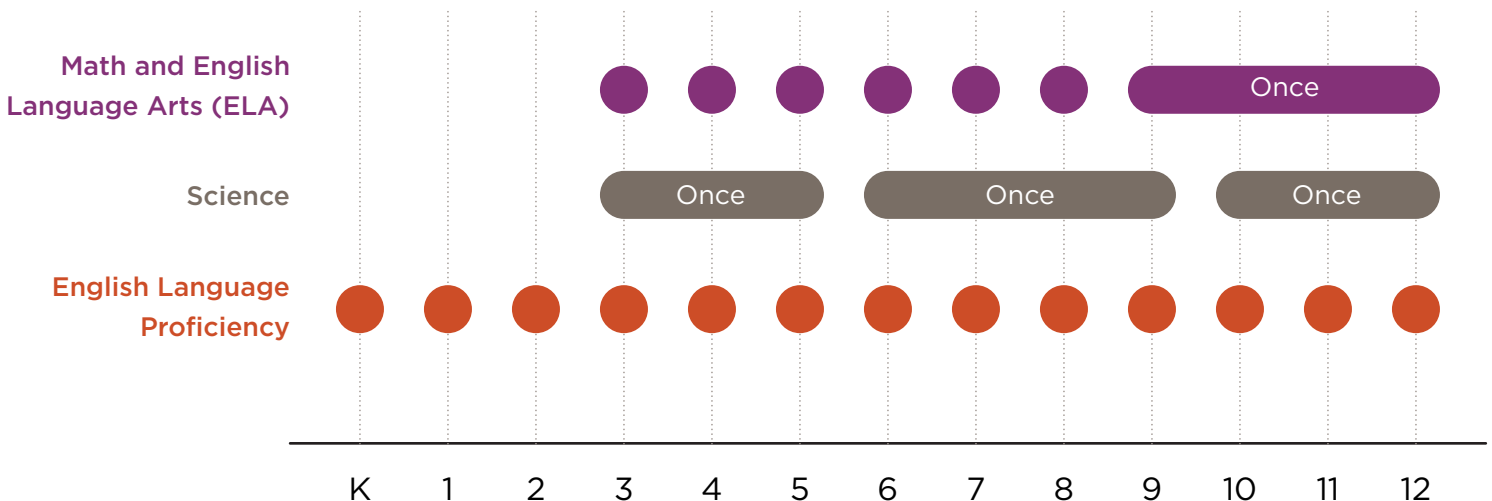
This Policy Guide provides comprehensive information about the federal policies governing state summative assessment systems, flexibilities available to state leaders in designing and administering summative assessments, and opportunities for innovation. It offers considerations for state leaders as they examine their own assessment systems and highlights state examples of various approaches to summative assessments.

Federal Assessment Requirements

Assessment Administration

Although states take various approaches to summative assessments, they must comply with the federal requirements for summative assessment systems outlined in the Every Student Succeeds Act (ESSA). ESSA sets minimum requirements for annual assessments in math, English language arts (ELA) and science, including:

- A math and ELA assessment each year between third and eighth grade and one time between ninth and 12th grade.
- A science assessment one time in each of the following grade spans: third through fifth grade, sixth through ninth grade and 10th through 12th grade.



To meet the summative assessment requirement, ESSA offers states flexibility to select or adopt [various assessments](#) that are aligned with state academic standards and meet certain technical criteria. Federal statute and regulations highlight options for state leaders: computer-adaptive tests, multiple interim assessments (aka through-year assessments) and the use of a nationally recognized assessment for high school students. This also includes performance tasks or student portfolios.

Each state assessment administered in compliance with federal law must be submitted and approved through the U.S. Department of Education's peer review process that evaluates the technical quality of assessments. The department identifies [critical elements](#) of assessment systems used in peer review. States are required to submit evidence that their assessment system addresses each critical element sufficiently, which includes technical documentation such as:

- State academic standards.
- Test blueprints.
- Assessment items.
- Reports on stakeholder engagement.
- Contracts with vendors.
- Procurement procedures.
- Test administration manuals.
- Reporting guidelines and resources.
- Other relevant information.

Assessment Participation

Student Participation

Student participation in summative assessments is a key provision in federal law especially considering lower test participation rates and parent opt-out options in some states. ESSA outlines a minimum participation standard for statewide assessments to ensure states have sufficient information to measure school quality and disaggregate student achievement and growth data.

ESSA requires state accountability systems to penalize schools that have summative assessment participation rates below 95% for all students and each identified student group. If more than 5% of students are not tested, the lowest possible score will be assigned to non-test takers beyond 5%. States are also required to outline corrective actions for schools that do not meet the 95%

participation rate in their [ESSA plans](#). Based on a review of state ESSA plans, states take various approaches to corrective action. In some cases, states rely on the negative impact that low participation has on achievement and growth indicators as the sole corrective action. Others publicly report on schools below 95% participation and require those schools to develop a corrective action plan or provide state resources and supports to increase student participation.

Test participation is also important for ensuring states can disaggregate data by specific student groups, including students from major racial and ethnic groups, students from low-income backgrounds, students with IEPs and students classified as English learners (also known as multilingual learners). ESSA grants states flexibility to set the n-size for both reporting and accountability purposes. N-size refers to the minimum number of students required to report on student performance or include a student group in accountability determinations.

While setting a sufficient n-size is important to ensure [data privacy and reliability](#), setting the n-size too high results in the exclusion of student groups from accountability determinations and data reporting, which has [significant implications](#) for identifying and addressing inequities between student groups. For instance, [one analysis](#) finds that an n-size of 30 may exclude up to 40% of Black elementary-grade students nationwide from student group accountability. While states may set different n-sizes for accountability and reporting purposes, most use the same n-size for both. [A recent analysis](#) of state accountability systems found n-sizes ranging from 10-30 for accountability.

Students With IEPs

ESSA enables states to administer an alternative assessment for students with the “most significant cognitive disabilities,” which must be capped at no more than 1% of the total number of students assessed in the state. Alternative assessments must be based on alternative academic standards that are aligned with state academic standards, promote access to the general curriculum, and allow successful transitions to postsecondary education or employment. Schools must notify parents during the Individualized Education Plan (IEP) process that the student is being assessed on alternative standards and “how participation in such assessments may delay or otherwise affect the student from completing the requirements for a regular high school diploma.”

Individualized Education Programs (IEPs)

IEPs are plans or programs created by a student with a disability, their family and/or caregivers, and practitioners from their school to support their instruction needs. Read more from the [U.S. Department of Education](#).

The 1% cap provision is designed to protect against [unnecessarily assigning](#) students to alternative assessments based on alternative academic standards that could limit a student's exposure to the full curriculum. This may impede a student's ability to earn a high school diploma, which has implications for long-term outcomes. States can apply for a [waiver](#) from this provision. However, this recent [report](#) identified 33 states currently out of compliance with this provision, which raises concerns over inappropriate assignment and student success after high school.

To better enable participation, students with IEPs taking the statewide summative assessment must be provided with accommodations as identified in their IEP — such as the use of assistive technology — and the assessment itself must [incorporate](#) Universal Design for Learning principles.

English Language Proficiency Assessments

Federal law requires states to administer English language proficiency assessments for all identified English learners annually. Screening assessments and home language surveys are the most common approaches to identifying multilingual learners. Once students are identified, districts must determine if they are classified as English Learners through a valid and reliable English language proficiency assessment.

State and local education agencies must ensure that annual English language proficiency assessments are aligned with state English language proficiency standards and assess the proficiency of students in all four domains of English (i.e., speaking, listening, reading and writing). These assessments are held to the same standards of peer review as required academic assessments. While states are permitted to develop their own English language proficiency assessments, most are [members of a consortium](#).

The results of these assessments are a required metric in a school accountability rating; however, flexibilities in federal law impact how this is measured and how students are counted. [Most states](#) rely solely on English learner growth for calculating the indicator, but a handful of states have incorporated proficiency into the calculation as well. Growth can be measured in several ways, but states primarily rely on growth-to-standard measures. Although the data and methods used to calculate the indicator have important implications for accountability systems, the [n-size](#) that states use to determine whether or not English language proficiency is included in accountability calculations significantly impacts the number of schools evaluated on student progress toward English language proficiency. The [weight](#) of the [indicator](#) may also impact the significance placed on English language proficiency at the school or district level.

Multilingual Learners

ESSA also requires that multilingual students (formally identified as English learners) participate in the statewide summative assessment. States may choose to exempt multilingual students who just started at a school from one year of the ELA assessment. ESSA requires states to make every effort to develop [native language assessments](#) in languages other than English that are present to “a significant extent” within the test-taking population. [A 2020 analysis](#) indicates that 31 states and the **District of Columbia** offer native language assessments, most commonly in Spanish, for math or science. Other accommodations — such as language supports, glossaries and extended time — [must be made available](#) to multilingual students if native language assessments are not available for a tested grade or subject.

Student Achievement and Growth Measures

States are required to develop a system of “annual meaningful differentiation,” which refers to the indicators, formulas and determinations that distinguish the performance of schools in their state accountability system. This federal requirement helps states identify schools with the lowest academic outcomes for students and provide these schools with the resources and supports they need to improve. State systems of annual meaningful differentiation benefit all schools by providing them with data that can inform resource allocation and continuous improvement efforts. This information, which must be publicly reported, also helps ensure parents, communities, system leaders and policymakers understand which schools and districts need support to meet the needs of all students and students from underserved groups.

ESSA further stipulates that a state’s accountability system uses summative assessment data to create a measure of **student achievement** in math and ELA and to use this as a significant factor in determining school performance ratings. ESSA requires these student achievement measures to be disaggregated by race/ethnicity, economic status and by specific population such as for students with IEPs and multilingual learners. [Nearly all states](#) also leverage summative assessment data to calculate a measure of **student growth** to fulfill ESSA’s requirement of using another “academic indicator” to evaluate the performance of schools.

Student achievement and growth measures [work together](#) to provide an important snapshot of how schools are meeting the needs of students. This is because as compared to student achievement measures, growth measures can

provide a [better look](#) at school quality: Unlike academic achievement which is [strongly correlated](#) with student demographics and family income, growth is less influenced by how much access a student has to opportunities outside of school and more reflective of what a student has actually learned over the course of the year.

State leaders can select the type of statistical model their state uses to calculate growth measures. This decision is important as [different growth models](#) often produce distinct outputs that answer fundamentally different questions about how schools are contributing to student success. Some states measure growth using normative measures; however, because normative approaches highlight relative performance, the results are not able to indicate how far students are from reaching grade-level proficiency. One common type of normative measure is a student-growth percentiles, which compare a student's growth to the growth of other students who performed similarly on the prior year's assessment.

Other models, such as value tables, growth-to-standard and gain-scores, are criterion-based measures that measure growth against grade-level proficiency. A focus on progress toward a predetermined proficiency benchmark can contribute to a better understanding of students' grade-level performance, their growth trajectory and help maintain high academic standards for all students. However, since criterion-based measures assess growth relative to proficiency, these measures are more sensitive to how and where states set their [proficiency cut scores](#). These types of measures are often considered the most straightforward and potentially easier to communicate to stakeholders and members of the public.

School Identification and Continuous Improvement

States are federally required to identify schools for improvement by considering how schools are serving the needs of all students and, separately, the needs of specific student groups. ESSA outlines three types of identification status:

- **Comprehensive Support and Improvement (CSI):** represents the lowest performing 5% of Title I schools across accountability measures and/or with graduation rates below 67%.
- **Targeted Support and Improvement (TSI):** represents schools “consistently underperforming” for any group of students as defined by the state.
- **Additional Targeted Support and Improvement (A-TSI):** represents schools where a specific student group would be identified for CSI.

When a school is identified for support under any of these three designations, federal law requires state, district and school leaders to engage various stakeholders to develop and implement a plan for improving these schools, and for schools to partake in [improvement activities](#) according to a school's identification status. For instance, schools identified for CSI must develop a school improvement plan, conduct a [resource allocation review](#) and select evidence-based practices to implement. Recent analyses suggest that many CSI schools need additional support and [stronger state oversight](#) to carry out these activities with fidelity.

To help support the improvement efforts in identified schools, states are required to set aside 7% of their Title I funds for school improvement efforts. However, federal law grants states significant flexibility in how these funds are allocated to identified schools. States can allocate these funds via a formula, a competitive process or a combination of both and can attach requirements for obtaining or using these resources. For example, in **Texas**, grant recipients are required to participate in leadership development and coaching that aligns with their improvement plan.

Regardless of how states choose to allocate these funds, it's unlikely the amount will be [sufficient](#) to support large-scale programs or interventions — in the 2020-21 school year, identified schools received [an average of \\$96,000](#). Some advocacy organizations [have argued](#) federal policymakers should increase funding for the set-aside amount altogether. However, in the absence of this increase, these funds can still help build the capacity needed to thoughtfully identify needs, plan for improvements and/or set-up pilot programs.

ESSA's stipulation that states must create a system to meaningfully differentiate the performance of all schools helps identify the ways every school can continuously improve by pinpointing areas to better meet students' academic needs. **Missouri** goes as far as prioritizing improvement efforts in all schools by adding a measure of continuous improvement in their revised accountability system, which considers improvements in effective teaching and learning, data-based decision making, and equity and access.

Public Reporting

ESSA requires states to publicly report student achievement data that includes overall performance and specific student groups performance each year. Most states choose to report this information via a state report card [platform](#). These platforms [offer transparency](#) to allow stakeholders to understand how states, districts and schools are serving the academic needs of all students. They

also further contextualize achievement data with other measures the state is required to report such as access to [experienced and qualified educators](#), per-pupil funding expenditures, graduation rates, and any other measures the state chooses to report.

ESSA does not outline a uniform format for this reporting; However, the law does require that report cards be concise, presented in an understandable and uniform format, and accessible to the public, including disabled people, and to the extent practicable, provided in a language that parents and other stakeholders can understand. The U.S. Department of Education released [non-regulatory guidance](#) to assist states with meeting these conditions.

Beyond these parameters, states have flexibility to design report cards that best meet the needs of their communities. Despite the flexibility, several recent research reports from the [Data Quality Campaign](#), the [Center on Reinventing Public Education](#) and [EdTrust](#) have showcased the significant limitations in states' current reporting. Additionally, some states maintain separate reporting platforms outside of their report card used for federal accountability purposes, which can prevent stakeholders from getting a consistent and accessible picture of student opportunities and experiences in school.

ESSA also requires that each student receives their individual summative assessments results through an individual score report and that educators and administrators receive aggregated score reports across a classroom, school or district.

When score reports are clear, accessible, comparable and timely, stakeholders can accurately understand what the results represent. However, as it stands, stakeholders often have to wait [several months](#) before receiving summative assessment results back, at which point, the data may be [outdated and less useful](#). **Ohio** addressed this by [legislating a requirement](#) for districts to inform families of their child's summative assessment results no later than June 30 of the same school year.

States can consider how to report on scores in an asset-framed manner, which means that data is presented in a way that highlights the responsibility of the education system toward students rather than the onus or responsibility being placed on students themselves. For instance, **Massachusetts'** [individual score report](#) notes it is the role of the school system (alongside the family) to ensure that students scoring below the "meeting expectations" performance level receive the additional supports they need to succeed. Asset-framed data also highlights student and/or school strengths without masking areas for growth and contextualizes the data by providing information on students' access to important resources.

Assessment Quality

A high-quality assessment supports an accurate, meaningful understanding of student performance and provides opportunities for stakeholders to make data-informed decisions. To ensure that state assessment systems comply with the necessary quality standards outlined in federal law, the U.S. Department of Education outlines and maintains a system of peer review. This peer review process checks for:

- **Validity:** ensures the test accurately captures its intended measurement. If a test is valid, it accurately measures the skills and knowledge it is intended to measure.
- **Reliability:** ensures the assessment results are consistent for the test-taker across multiple attempts in similar conditions.
- **Comparability:** ensures assessment results — both individual student results and student group results — can be accurately compared to other student or student group results. An assessment's comparability is directly connected to its validity and reliability.
- **Alignment:** ensures the assessment evaluates state grade-level academic standards.
- **Accessibility:** ensures the assessment integrates proper accommodations into the assessment experience. This includes accommodations for students with IEPs with disabilities as outlined in their IEP such as assistive devices and technology, Braille materials, large print or speech-to-text services. Assessments must also integrate additional supports for English learners. Assessments should aim to adhere to Universal Design for Learning and Universal Design for Assessment properties.

However, these quality provisions featured in federal peer review requirements represent the minimum requirements for state policymakers to ensure that summative assessments produce meaningful data. Moreover, while not explicitly required in federal law, there are several additional elements that impact the quality of a summative assessment.

For instance, state education leaders can consider the coherence of state assessment systems. For a balanced assessment system, summative assessments should be aligned to and complement other assessments to paint a [clear and comprehensive picture](#) of student learning. When an assessment system is not coherent, schools may administer conflicting or redundant assessments that produce fragmented — sometimes conflicting — data and

limit the ability of practitioners to provide the targeted supports students need despite their best efforts. State leaders can work to improve the coherence of their assessment system by issuing guidance, offering technical assistance opportunities to districts and schools, and developing high-quality interim assessments (if not pursuing a through-year assessment model).

Additionally, state leaders can consider the relevance of summative assessments to students to ensure they are able to fairly demonstrate what they know and are able to do. For example, assessments featuring a wide range of [students' identities and interests](#) allow all students to see themselves in the assessment and draw upon their existing knowledge to demonstrate their learning. For instance, the [2028 NAEP Science Assessment Framework](#) provides an example (Exhibit 3.20) of how assessment items can incorporate “the use of non-traditional evidence sources,” such as multi-generational/elder accounts, which hold cultural validity in non-white cultures, especially within Indigenous cultures. This is important because [research](#) shows that students are [more engaged](#) when they like themselves in school materials. This increased engagement may produce summative assessment scores that paint a more accurate picture of students' abilities. States may consider revisiting assessment vendor processes and [bias controls](#) to ensure they align with the state's vision for inclusive assessments.

State Flexibility in Assessment Design and Selection

Federal law grants states flexibility to adopt various assessment approaches, including computer-adaptive tests, through-year assessments, performance assessments and nationally recognized assessments as long as each meets specific technical criteria and passes the U.S. Department of Education's peer review process. Several states are pursuing innovations using funds from the [Competitive Grants for State Assessment \(CGSA\)](#) and flexibilities provided through the [Innovative Assessment Demonstration Authority \(IADA\)](#) (see more below).

Computer-Adaptive Tests

While some states adopted computer-adaptive testing over a decade ago, it is still an uncommon approach used in summative assessment systems. A computer-adaptive test adjusts the difficulty of questions based on student responses throughout to provide more nuanced and precise results. Students are

scored on both the number of correct answers and the difficulty of the questions. This approach can reduce testing time by ensuring students are not spending time on questions that will not yield useful information. Targeted questions may help better identify standards on which students need additional supports.

Despite its potential benefits, shifting to computer-adaptive testing may [pose challenges](#). Students and test administrators may need support to effectively transition to the new assessment system and understand or communicate the results. Additionally, developing a precise understanding of student achievement may make it difficult to cover the full depth and breadth of state standards without increasing testing time. Finally, computer-adaptive assessments require a larger bank of questions, which is more resource intensive to develop. It may also limit the use of different question formats or open-ended responses to allow for real-time scoring of student work.

Hawai'i was an [early adopter](#) of computer adaptive testing. The state joined an [assessment consortium](#) that offers summative assessments in third through eighth grade and in high school. The assessments include an adaptive component and a performance task. The state education agency developed a [FAQ resource](#) and [other resources for families](#) to address common concerns around computer-adaptive assessments and ensure families understand how the state academic standards and summative assessments are aligned.

Nationally Recognized Assessments

ESSA permits states to adopt a nationally recognized assessment as the high school summative assessment or allows districts to adopt an approved locally selected, nationally recognized [summative assessment](#). States opting to use a nationally recognized assessment for high school math and ELA or those interested in permitting districts to implement the locally selected option must submit selected assessments for peer review, satisfy federal technical criteria and demonstrate alignment with state standards.

Federal law defines a nationally recognized assessment as an “assessment of high school students’ knowledge and skills that is administered in multiple states and is accepted by institutions of higher education in those or other states for the purposes of entrance or placement into courses in post-secondary education or training programs.” The ACT and SAT are explicitly named as nationally recognized assessments, but advanced placement, international baccalaureate, and assessments offered by one of the major

consortia have been [highlighted](#) as options that meet federal criteria. In a 50-state comparison of state accountability systems, Education Commission of the States [identified](#) 17 states using the ACT or SAT as the primary high school assessment.

Proponents of the nationally recognized assessment option, particularly the adoption of college entrance exams, point to the ACT and SAT as assessments that can serve as a measure of achievement and college readiness. Adopting a college entrance exam also ensures students have access to the assessment, which removes a potential barrier to postsecondary education. However, [research](#) raises some concerns around the alignment of these assessments with state standards, which has implications as measures of student achievement based on learning and for use in state accountability systems. This is especially important in states that permit the locally selected option by creating comparability challenges between districts using a nationally recognized option and those using the state summative assessment.

North Dakota is an example of a state that leveraged the locally selected, nationally recognized assessment flexibility, while also ensuring they have comparable data for students in high school. The state education agency partnered with a vendor to develop a [math and ELA assessment](#) aligned to state standards that schools and districts may opt to administer to students in 10th grade. The state completed the peer review process and received approval for districts to administer the ACT for accountability purposes in 2019. The [state education agency](#) set its own ACT cut scores for achievement levels to align with state grade-level performance expectations, which has implications for student performance reporting and school accountability ratings. North Dakota is currently transitioning to a [new assessment system](#), including formative, interim, and summative assessments for all students and will no longer permit schools and districts to use the locally selected option.

Performance Assessments

[Performance assessments](#) allow students to demonstrate their knowledge by directly exhibiting a skill, conducting an investigation, developing a product or giving a performance. Examples of performance assessments range from short written answers and essays to experiments, portfolios, and projects or capstone assignments. States are increasingly [exploring](#) the use of performance assessments during instruction to meet [graduation requirements](#) or for inclusion in statewide summative assessment systems. The use of performance assessments for accountability purposes would represent a shift away from common approaches to summative assessments.

Some advocates argue that [performance assessments](#) can increase instructional relevance, provide applied measures of student achievement and increase the value of assessment systems for educators. When employing performance assessments as a part of summative assessment systems, there are [some challenges](#) that may impact their value and intended purpose. Experts have [pointed](#) to the logistical burden of scaling performance assessments, teacher capacity constraints, increased testing time, lower levels of reliability and a longer timeline for reporting results.

Notably, New Hampshire [cited](#) some of these concerns with their performance assessment when they withdrew from the IADA. There are also validity concerns due to variations in implementation and scoring across districts. These challenges may make it difficult to uncover and address inequities and require substantial time and resources to effectively implement — particularly when attempting to scale across multiple schools and districts.

Some states are working to scale performance assessments and address some of these challenges. Massachusetts offers one example of a state developing a performance assessment that could comply with federal summative assessment requirements. Massachusetts received approval under IADA to develop and pilot a new science assessment that includes an abbreviated version of their existing summative assessment paired with technology-enhanced performance tasks that are aligned to content standards and principles of deeper learning. According to the 2022-23 school year progress report, the pilot included 12,000 students with plans to field test the performance assessment for 100% of fifth and eighth graders in 2024-25 and implement the performance assessment for all students in 2025-26 and beyond. To support scaling of the assessment and continued implementation of deeper learning, the state education agency developed performance tasks for use in instruction and reviews curriculum materials for quality and standards alignment.

Through-Year Assessments

Many state leaders are [exploring opportunities](#) to move from a single summative assessment for accountability purposes to a collection of interim assessments that are sometimes paired with a final summative assessment. Through-year assessments adjust the frequency of a state's summative assessment so that students take smaller, more frequent tests during the year, and results are returned sooner than typical summative assessment results.

Many argue that this structure provides timelier and more meaningful assessment results for educators and administrators and is a potential cost-savings to districts and schools. This is because through-year assessments can allow districts to skip spending limited resources administering costly commercial interim assessments. Others point to this structure as a promising way to afford students multiple opportunities to demonstrate proficiency during the year.

However, many experts have underscored that designing and implementing a coherent, equitable through-year assessment system is a complex endeavor that presents states with difficult choices. Most fundamentally, states must determine which standards students will be assessed on (and when), how the assessment will connect to the instructional pacing and/or curriculum (if at all), and how to score the assessment. Others have also pointed out the heightened need to support educators, administrators and families with understanding the unique components of through-year assessment data.

Several states are currently piloting a through-year assessment model. For example, **Montana** secured funding through the CGSA program in 2022 and 2024 to develop and scale their Montana Aligned to Standards Through-Year (MAST) assessment in ELA and math. The math assessment is unique in that it is designed with strategic, but flexible clustering of standards in “testlets,” which can be configured in the order that best aligns with a district’s local scope and sequence of instructional material. The state recently received a field test waiver approval from the U.S. Department of Education granting them permission to forgo “double testing” for a subset of districts. In these districts, students only take the through-year assessment — as opposed to taking both the pilot and the legacy summative assessment.

Opportunities for Innovation

While state summative assessment systems are governed by federal requirements, there are some opportunities for states to develop innovative assessment systems. The IADA offers states exemptions from some federal requirements while they pilot innovative options with the aim of scaling for use as the statewide summative assessment. The CGSA program provides funding to states developing innovative approaches to assessment, including innovations in summative assessment systems. There are also additional opportunities for state leaders to champion innovations to advance elements of quality highlighted above.

Innovative Assessment Demonstration Authority (IADA)

The IADA [permits](#) approved state education agencies to establish, operate and evaluate an innovative assessment system. The program was designed so that states successfully completing the demonstration period could receive approval from the department to scale the assessment for statewide use in compliance with federal assessment requirements. States piloting innovative summative assessments may be temporarily exempt from specific reliability requirements, so long as the assessment yields valid outcomes. Thus far, states have used the flexibility to develop through-year, performance and instructionally embedded assessments. As of publication, no states have completed the demonstration period and received approval to scale statewide.

State leaders raised concerns over the policies governing IADA and highlighted the [limitations of the policy](#). Both **Georgia** and **New Hampshire** withdrew from IADA citing capacity constraints and issues meeting IADA's standard for comparability between the legacy assessment and the innovative assessment. States also suggested that the lack of funding was a limiting factor. Following a request for information and stakeholder engagement process, the U.S. Department of Education issued [clarifications and regulatory updates](#) to support state participation in November 2023. Notably, the update:

- Lifted the cap on the number of states permitted to participate.
- Clarified methods for demonstrating comparability to ensure they wouldn't stifle innovative assessment design.
- Created standardized review windows.
- Created a planning status phase for states seeking early feedback from the department.
- Elevated funding opportunities that help to address capacity and resource challenges for interested states — namely the CGSA program.

Louisiana, Massachusetts and **North Carolina** are the only states identified as participants in IADA at the time of publication. Each state is working toward scaling the assessment statewide in the 2024-25 school year. **Massachusetts'** performance assessment was outlined previously and **North Carolina** has developed through-year assessments. In North Carolina, the [Personalized Assessment Tool](#) and [NC Check-Ins 2.0](#) are being piloted as the math and ELA assessment for third through eighth grade.

Competitive Grants for State Assessments (CGSA)

The CGSA program [provides](#) states with funding designed to increase the quality of assessment systems. Specifically, federal law outlines allowable uses for grant applicants that were [last updated in 2020](#), including:

- Assessments for English learners.
- Improving student growth models.
- Assessments for students with IEPs.
- Collaboration between postsecondary institutions, research organizations and state education agencies in assessment development.
- Developing multiple measures of student achievement.
- Creating competency-based assessment instruments.

In 2024, the most recent grant cycle, the U.S. Department of Education identified two [absolute priorities](#) for grant applicants and additional competitive or supplemental priorities. Each absolute priority is considered its own funding category and grants may be awarded under either. The absolute priorities include:

- 1| Developing an assessment instrument that relies on multiple measures or is aligned to a competency-based model.
- 2| A commitment to applying to IADA.

Competitive and supplemental priorities address educator diversity, professional learning and improved score reporting.

In the last two rounds of funding awards (2022 and 2024), 21 states have received funding through the program with award amounts typically ranging from \$2 million to \$5 million. States have leveraged these funds to develop, pilot and scale through-year assessments, [competency-based assessments](#) and instructionally embedded assessments that are aligned to high-quality curriculum materials. Many awardees demonstrate state efforts to develop balanced assessment systems beyond summative assessment requirements.

In addition to Massachusetts and Montana, Hawai'i, Kentucky, Illinois, Indiana and Nebraska are leveraging CGSA funding to improve their summative assessments. For example, **Indiana** is using federal funds to expand on their [through-year summative assessment](#) pilot. In 2024-25, 73% of schools will pilot the new assessment with the goal of launching the assessment statewide in

2025-26. CGSA funding will support the development of a predictive model, professional development for educators, family resources and connections between the assessment system and high-quality instructional materials.

Nebraska is also leveraging federal funding to develop their [through-year assessment](#) program. They plan to use grant funds to develop assessment items that enhance instructional utility for educators, develop new score reports that prioritize utility for key stakeholder groups, and build capacity of school personnel and awareness of students and families through professional development and stakeholder engagement.

State-Level Initiatives

State policymakers can pursue innovations in summative assessments through various actions with or without federal support via CGSA funding or IADA participation. Primary options include directing improvements through legislation or regulation, leveraging the request for proposal process to advance innovative approaches, engaging stakeholders in assessment development, and providing professional development to support implementation and use of assessment data.

State Legislation and Regulation

State legislators may choose to introduce and monitor legislation that seeks to innovate their summative assessment system. State legislators can also consider how much additional funding is needed to effectively manage statewide efforts to innovate assessments. For example, in 2019, legislators charged the Texas Education Agency to pursue a redesign of their summative assessment, including developing and piloting an innovative through-year assessment model as a possible replacement. Recently, the Texas Education Committee released a [2024 Interim Charge](#) to monitor the status of the effort.

Request for Proposals Processes

State agency leadership may also leverage their consumer power when procuring a summative assessment. The contract renewal or request for proposals process offers an opportunity for states to communicate their priorities and push for advances in assessment design and implementation. For instance, **Vermont** [recently transitioned](#) to a new statewide assessment and cited that this new selection better aligns with the state's values of diversity, equity and inclusion and offers an easier user experience for students, families and educators.

Task Forces and Advisory Groups

States can also establish standing task forces, advisory councils and other feedback channels to understand the experiences of teachers, families and administrators — particularly those serving students of color, students from low-income backgrounds, English learners and students with IEPs — with administering and using summative assessment results. These perspectives can help identify if, and how, future innovations could address the needs of the state.

There are several ways states can establish such feedback channels: **Kentucky** secured [CGSA funds](#) to [establish](#) and engage the [United We Learn Council](#), which consists of students, educators, community and business leaders, and families making [recommendations](#) on the future of assessment and accountability in the state. Meanwhile, **Virginia** lawmakers passed [legislation](#) to commission [a working group](#) to develop recommendations for revising their summative assessment.

Capacity Building

Finally, as states develop and implement innovative assessment systems, professional development and local capacity building play a vital role in maximizing the value of these improvements. Local capacity is especially important when considering alternatives to summative assessments since educators are less familiar with administration, scoring and analyzing results from these sorts of assessments. In 2024, **Connecticut** was awarded [CGSA funds](#) to develop competency-based interim assessments that focus on student choice and locally constructed response tasks and items, including the development of specific resources to help educators implement and administer the assessment.

State leaders are uniquely positioned to support the development of innovative assessments. Through legislation and policy directives, as well as investments in community engagement and local capacity building, state leaders have several levers to increase assessment quality and utility for key stakeholders.

Final Thoughts

While federal law outlines clear summative assessment and school accountability requirements, state policymakers have flexibility in the design, selection and use of summative assessments and the data they generate. Despite concerns over the value of existing summative assessment systems, they provide vital data for students, families, educators, districts, and state and federal leaders to better understand student progress and school quality. Without this data, comparing performance between schools and districts and identifying inequities in performance across student groups would be more challenging. Summative assessment data shines a light on student and school performance and should be leveraged as a tool for school improvement planning, implementation and evaluation.

State policymakers are working to not only create innovative assessment options, but also maximize the quality and value of existing assessment systems. States have used federal flexibilities, pilot programs and funding opportunities in addition to their own policy levers to advance innovations and increase the value of summative assessment systems for education stakeholders.



About the Authors

Ben Erwin



As a senior policy analyst, Ben works on several topics including assessment and accountability, school choice, student pathways, and teaching, among others. Prior to joining Education Commission of the States, he taught high school social studies in Kentucky and worked in education policy at the National Conference of State Legislatures. He earned a master's degree in education policy from the University of Colorado Boulder and a bachelor's degree in history and education from Transylvania University. Contact Ben at berwin@ecs.org.

Shayna Levitan



Shayna is a policy analyst on the P-12 team at EdTrust. In this role, Shayna supports work related to the Alliance for Resource Equity and conducts research and analysis on innovative assessments, accountability systems and equitable literacy policies. Prior to EdTrust, Shayna worked as a policy analyst and development associate at the National Council on Teacher Quality. Shayna holds a B.A. from American University, where she studied law and education and earned a certificate in ethical leadership studies. Contact Shayna at slevitan@edtrust.org.