ASSESSMENT OPTIONS FOR STATES:

Potential Designs With Increased Federal Flexibility

April 2026

Scott F. Marion & Kyla McClure

Center for Assessment

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

ACKNOWLEDGEMENTS:

Thanks to Chris Domaleski, Meagan Karvonen and Ajit Gopalakrishnan for providing extremely helpful comments and suggestions. The views expressed and any errors are our own.

SUGGESTED CITATION:

Marion, S.F., & McClure, K. (2026). *Assessment Options for States: Potential Designs With Increased Federal Flexibility.* The National Center for the Improvement of Educational Assessment.

PHOTO CREDIT:

Cover photo by Allison Shelley/The Verbatim Agency for EDUimages

TABLE OF CONTENTS

ASSESSMENT OPTIONS FOR STATES:

Potential Designs With Increased Federal Flexibility

OVERVIEW

The Center for Assessment has made a strong case for state testing. We fully support that stance. But as the Center acknowledged in the paper, current state achievement tests are not perfect. This is true because state tests, since at least 1994, have been designed to meet specific federal requirements and operate within established constraints.

The current moment presents an opportunity for improvement: The U.S. Department of Education has invited states to seek flexibility from certain legal and regulatory provisions that have constrained them. In this paper, we'll describe the potential opportunities and risks of several assessment options that might become available. We'll also discuss some types of flexibility that might unintentionally undermine student learning goals if states pursue them. We offer these options based on what we know now, as well as what research and practice teach us about high-quality student assessment.

The Case for State Testing laid out four purposes that current summative state assessments serve, including:

1. Monitoring statewide educational growth and achievement
2. Evaluation and continuous improvement
3. Transparency and public engagement
4. Signaling rich learning expectations

Existing state assessments typically serve the first two purposes well, but could improve in adequately supporting the last two. In particular, we believe there is an opportunity to redesign state assessments to signal even richer learning expectations than those currently in place.

While we are nervous about too much federal flexibility, we think some additional flexibility could expand assessment options to better serve students, educators, and school systems. However, with flexibility comes responsibility. State leaders are constitutionally responsible for ensuring that all students have access to fair educational opportunities. States define these opportunities differently, but they must be able to compare and evaluate the performance of all students in their jurisdictions. High-quality assessments help state leaders evaluate and document the degree to which students are, in fact, receiving opportunities to learn the expected content and skills. Many of the current federal requirements were instituted to help ensure that our state assessments can serve this purpose.

One of our reasons for writing this paper is to help state leaders engage in scenario planning. This is a technique used by businesses, the military, and other organizations to imagine and prepare for an

uncertain future. The U.S. Department of Education's functions are currently being outsourced to various agencies, and we are operating in a landscape of great uncertainty.

We think there is a high likelihood that certain types of assessment and accountability flexibility might be granted to states. We are concerned that if state assessment and accountability do not engage in scenario planning, they might be caught flat-footed when educators, leaders, and policymakers demand changes to their state assessment systems.

Requirements and Constraints

When federal lawmakers approved the Improving America's Schools Act in 1994, and again when they passed the No Child Left Behind Act (NCLB) in 2002, they required states to produce annual summative determinations for each student that are comparable across students, subgroups, schools, and time. The law requires statewide achievement tests to be aligned, to the extent possible, with the state's learning standards. Additionally, all students must participate in the state assessment, and states must make the assessment accessible and fair to all students. The limited number of students unable to show what they know on the general assessment are permitted to participate in the state's alternate assessment on alternate academic achievement standards.

It is hard to argue with any of these requirements—together, they create a set of guardrails that ensure assessment systems can serve their primary purposes effectively. But loosening comparability and alignment requirements (just a little) might open up some possibilities for innovation. States can impose additional requirements based on their values and goals. For example, if one of a state's goals is to signal the importance of having students analyze and write about original historical texts, it could include tasks that require students to do such analyses as part of its English/language arts assessment.

> *Loosening comparability and alignment requirements (just a little) might open up some possibilities for innovation.*

States face various constraints on the design of their assessments, but in general, these stem from budgets, testing time, politics, and the capacity of state and district personnel. For example, the Center for Assessment staff served as the lead technical advisors for New Hampshire's innovative Performance Assessment of Competency Education (PACE) system. We learned early in the initiative that the local capacity required to make the system work exceeded that of most participating districts.

This is an early step in any design exercise. We recently described the design process we use at the Center for Assessment, and we strongly suggest that any state or district interested in reenvisioning its assessment system use a broad-based, deliberative design process like the one described in this paper.

As good as state assessments are now for the specific purposes they serve, many assessment experts and advocates would like to see greater flexibility in federal requirements to allow states to improve the quality and usefulness of these assessments. We've been frustrated by calls for innovation in assessment that lack specificity about how people want to innovate and why. In fact, few even define what innovation means.

For example, many researchers and advocates are pushing to incorporate generative artificial intelligence into assessment to better measure what students know and help users make better

sense of the results. This would constitute an innovation, but it differs from other types of innovation, such as advancements in computer-adaptive testing (CAT). Therefore, it is critical to be clear and specific about the rationale, risks, and benefits of any proposed innovations.

We start our discussion at the high school level because we think this grade span is most in need of assessment reform. Many of the same principles apply to earlier grades, but the context and structures of high schools warrant separate treatment.

HIGH SCHOOL ASSESSMENT

At a national assessment conference a few years ago, several leading state assessment directors outlined their top recommendations for revising the Every Student Succeeds Act. One of them said he'd eliminate the high school testing requirements. We agree with this sentiment. In the context of current accountability requirements, the high school testing system is broken, as this paper describes. That's why the widespread use of the SAT and ACT as high school achievement tests may be seen as the least harmful of many bad options. Selecting the least-harmful option is not a way to inspire confidence in the statewide assessment system. On the other hand, state leaders could pursue many other options, especially if states had more flexibility than they currently do.

End-of-Course Assessments

End-of-course (EOC) assessments evaluate students' mastery of course-specific material at the end of the academic year, whereas general or survey assessments sample knowledge across multiple domains. For example, an EOC assessment could be designed for students taking Algebra I, aligning directly with the subject's state content standards and ensuring that all students who take the Algebra I assessment have had the opportunity to learn the content. Several states use EOCs, so it is hard to call it an innovation, but accountability and cost constraints have limited their widespread use.

Possible Federal Flexibility

Current federal accountability requires that the same assessments be used to measure the academic achievement of all secondary school students in the state (Section 1111 of the Every Student Succeeds Act, 2015). However, high school often marks a point where students' academic trajectories diverge. Some students take classes earlier or later, or choose different classes from their peers. Limited flexibility that allows a state to offer a menu of challenging EOC assessments, without requiring all students to take the same assessments, could help states serve students with diverse interests and needs.

Opportunities

There are numerous educational advantages to a robust end-of-course testing system, including closely aligning the content tested with the content students have had documented opportunities to learn and establishing a shared set of high expectations for key content domains. EOC systems can support shared and generally higher expectations across the state, and state leaders can be confident that students have had an opportunity to learn the tested content. With greater flexibility, states could limit accountability to one or two required courses per content area. If EOCs were used in a state accountability system, there are several ways to incorporate test results without the rigidity of requiring all students to complete every course and exam. We encourage states pursuing this type of flexibility to document that all students are provided meaningful opportunities to learn. In addition, we would encourage a state to ensure that all assessments represent equally challenging standards to avoid the potential negative consequences associated with "tracking."

Challenges

Some states continue to employ an EOC approach, but many others have shifted away or significantly reduced it due to cost and accountability concerns. There is no question that adding assessments costs more money. For example, administering four EOCs might not be exactly twice as expensive as administering two, but it could be close. The current accountability requirements to test students on the full set of content standards they are expected to learn have made EOC testing unwieldy, especially since many students complete high school requirements while still in middle school. We understand why schools want middle school EOC scores to count for high school accountability. The students who take EOCs in middle school are generally the highest-achieving students. However, it's hard for us to imagine why states do this, except perhaps as a contorted bid to meet federal accountability requirements. If such requirements remain in place, these same accountability challenges would persist.

Creating a limited system of EOC courses could also signal to schools that only some subject areas are valued or "count" towards federal accountability. Of course, states could turn this challenge into an opportunity by offering a broad array of EOCs. States and districts would need to ensure that resources are allocated to subject areas that may not be included in accountability determinations. In addition, states would need to decide whether students who do not pass their EOC assessment must enter remediation or another intervention to ensure they possess the knowledge and skills necessary to graduate from high school, or whether they may demonstrate their knowledge in some other way (e.g., a portfolio).

Portrait of a Graduate

A Portrait of a Graduate represents a state or district's agreed-upon set of skills, attributes, and/or competencies that students should demonstrate upon exiting K-12 education. Such competencies can encompass traditional skills, such as literacy, or broader competencies, such as adaptability, collaboration, communication, and critical thinking. States and districts can choose to tie different skills and competencies to state or local accountability systems, or provide guidance to schools and districts for how to determine whether their students are meeting the specified goals.

Possible Federal Flexibility

Currently, Portrait of a Graduate profiles operate outside of the federal accountability system because they address state and local graduation requirements and are not used to assess academic achievement. However, if the federal comparability requirements were significantly revised, states and districts might be able to use their Portrait of a Graduate profiles to evaluate students' knowledge of key high school content requirements. States' accountability systems could shift to consider the proportion of students who meet defined Portrait outcomes and the extent to which schools and districts support all students. Further, reducing comparability requirements could allow states or local districts to set unique criteria for evaluating student work related to the Portrait of a Graduate goals. Such scores likely require a qualitative evaluation of student quality, but could still be technically sound with sufficient scoring guidance.

Opportunities

Many states have adopted policies, typically called Portrait of a Graduate or Profile of a Learner, that articulate the critical aspirations state leaders and policymakers have for their students. Creating an assessment system, likely a state-local hybrid system, to support and document the degree to which students demonstrate competence relative to these crucial outcomes would represent an ambitious approach to high school assessment.

One opportunity embedded in a Portrait of a Graduate system is the ability to think broadly about the skills and competencies that schools should foster and support. Tying a state or district's Portrait of a Graduate to an accountability system would require careful thinking about how to operationalize such competencies in a fair and useful way.

For example, to demonstrate critical thinking, a state could ask students to complete an extended project, complete a relevant performance task, or engage in community action related to a problem they identify. Providing students with choice and allowing them to demonstrate competence in multiple ways may improve student motivation and learning outcomes. Choice and self-determination are important skills critical for postsecondary success. In addition, creating opportunities for students to synthesize knowledge across content domains more accurately represents the authentic learning and knowledge required in 21st-century jobs.

The International Baccalaureate (IB), an internationally renowned curriculum and assessment provider, created the IB learner profile to describe the key "capacities and responsibilities" that go beyond academic success. Their ten aspirational capacities, though not explicitly assessed, are related to the tasks required to earn an IB diploma. For example, the IB internal assessments ask students to pose their own questions, thus demonstrating students' ability to be an "inquirer," one of the ten IB profile aims. States could look to the IB framework as an example of how to develop an assessment system that embeds their specific Portrait of a Graduate competencies. IB will not be a model for everyone, but similar constructs will have to be defined for students working toward alternate achievement expectations.

Challenges

Connecting a Portrait of a Graduate system to accountability presents a host of challenges and considerations. For example, how many competencies would students need to demonstrate? What "counts" for demonstrating progress towards any one competency? What level of standardization would be desired and possible at the state or district level? Who would grade student work? How could the state or district ensure that all students were held to challenging standards?

Grading and scoring student work against the specified Portrait of a Graduate competencies is likely to be one of the biggest challenges in implementing this approach. Developing rubrics to score student work across multiple submission formats and then scoring and grading it with sufficient reliability can be costly and likely require significant support from local districts. Further, building local capacity to support this work faithfully would require time and resources.

As noted above, it would be challenging to meet current high school testing requirements in the Every Student Succeeds Act (ESSA) if the high school assessment were used to certify students' competence relative to the Portrait of a Graduate's knowledge, skills, and dispositions. However, with newfound flexibility, some interesting local-state hybrid approaches could be available to states.

Certification and Specialization

If state leaders and policymakers are interested in serious high school reform along the lines advocated by Mark Tucker and the National Center on Education and the Economy (NCEE) for many years, the high school assessment system could be reformed considerably. NCEE advocated that students demonstrate, by the end of 10th grade, sufficient knowledge and skills to pursue credible postsecondary options. The last two years of high school would be spent specializing in a particular realm, through advanced coursework and/or internships, so they have the knowledge and experience necessary to leave high school ready to pursue their chosen path.

This approach is reflected in the Blueprint for Maryland's Future, which requires students to meet a basic level of certification by the end of 10th grade and then pursue a specialized and/or deeper learning pathway in their last two years of high school. The initial certification can be documented using fairly common assessment approaches. However, assessing the flexible pathways students pursue after their 10th-grade certification may require a local-state hybrid, similar to what might be necessary for a Portrait of a Graduate-type system discussed above.

Possible Federal Flexibility

Currently, any program that requires or supports certification is implemented in addition to federal accountability requirements. Allowing states to substitute their proposed certification programs for the current high school testing requirements could reduce the burden of running two programs concurrently. Therefore, federal flexibility to replace ESSA's achievement testing requirements with certification opportunities could support the robust application of a certification approach to high school assessment.

Opportunities

Certification opportunities provide students with the chance to earn credit or a credential that demonstrates their readiness to enter future career pathways. For example, Florida's Career and Professional Education (CAPE) Act directs the Florida Department of Education to partner with business and professional organizations to identify high-demand career and professional courses and credentials that could be made available to high-school students. In 2020, the program reported that at least 74,000 students earned more than 88,000 certifications, with a pass rate of 70.1 percent.

Certification or credentialing can occur through different pathways and be awarded either through course participation or through a summative assessment. If the certificate is used in industry, offering access to such certification at the high school level could provide a financial incentive and encourage parents and students to participate in the certification or credential program. Many certification opportunities reflect the real-world, authentic learning needed to participate in various 21st-century employment opportunities.

Challenges

A certification approach presents several logistical challenges. A state must determine who administers the certification programs and how certifications are awarded. Will the certifications be externally administered industry or university certificates, or will they be offered by a state education department? What level of oversight will a state provide to ensure that all offered certificates represent rigorous and ambitious standards for students? In addition, a state must determine how to approach offering certificates that may lead to differential career outcomes, while ensuring that students are not tracked into particular fields. Adopting an initial certification, like the Blueprint for Maryland's Future, is one way to ensure that all students have mastered key knowledge and skills.

High School Assessment Summary

The approaches described above are options states could pursue with additional federal flexibility. States could also pursue a combination of approaches. For example, it would be hard to imagine assessing students on the kinds of competencies defined by most Portrait of a Graduate systems, such as complex communication, global citizenship, and quantitative reasoning, without using performance tasks, which we outline as an option for all grade levels below. If states move toward an EOC system, incorporating performance tasks into EOC tests could send strong signals about the types of learning and assessment the state wants to see in classrooms.

K-12 ASSESSMENT REFORM

Implement Statewide Performance Tasks

Performance assessments ask students to demonstrate knowledge, skills, or competencies by performing a task or producing a product. The term "performance assessment" encompasses a broad range of assessment tasks. They can be designed to offer students multiple ways to demonstrate their knowledge, or to more narrowly guide them toward a smaller set of possible responses. Typically, a performance assessment asks students to draw on multiple skills in an authentic, real-world context to demonstrate their knowledge and learning. Performance tasks are typically rubric-scored according to predetermined criteria.

Possible Federal Flexibility

States have employed performance-based assessments as part of their statewide assessment systems on and off over the past 30 years, but most of these efforts predate the stringent assessment requirements of the No Child Left Behind Act (NCLB), enacted in 2002. If some strict interpretations of technical quality embodied in the peer review guidance were relaxed (e.g., alignment), state leaders could use more performance tasks in their state systems. Performance assessments should be designed to measure deeper learning outcomes. While some alignment criteria evaluate the degree to which the assessments tap the appropriate depth of knowledge, most criteria assess the extent to which the assessments cover the breadth of the required content (e.g., the state content standards). It is hard to go both broad and deep without having a very long assessment. Educators and school leaders already express concern about the time required for state testing, so we think that a performance-based assessment system could benefit from a relaxation of requirements to measure almost every standard every year.

> *Performance assessments should be designed to measure deeper learning outcomes.*

Opportunities: Incorporating Performance Tasks Into K-12 Assessment Systems

If our goal is to align assessment systems with a rich vision of student learning, as Lorrie Shepard discussed 25 years ago, then we need to employ assessment approaches that call for students to synthesize and integrate content and skills within and across multiple content domains. Performance tasks ask students to do exactly that. Developing high-quality performance assessments that require students to integrate knowledge across domains signals to teachers and schools the kinds of rich learning experiences that should constitute the bulk of classroom practice.

Challenges

There are several ways to implement performance tasks to address some of the concerns (many of them misplaced) that were raised when performance assessments were used more widely in the 1990s. These concerns included low levels of individual-level reliability/generalizability because students completed only a few tasks. However, if more flexibility were available, performance tasks could, for example, be matrix-sampled at the school level, allowing schools to gain a richer and more generalizable picture of student performance. In this case, each student might be required to complete only one or two performance tasks in each subject area.

You might ask how we could obtain a generalizable student score with only one or two performance assessments. We are not suggesting using performance assessments in isolation. Rather, they could

be included in a curriculum-embedded approach (see below) or in another approach that supplements information generated by performance assessments. This way, the performance assessment wouldn't have to carry most of the generalizability weight.

> *Developing high-quality performance assessments that require students to integrate knowledge across domains signals to teachers and schools the kinds of rich learning experiences that should constitute the bulk of classroom practice.*

As with the Portrait of a Graduate model in high school, grading and scoring performance tasks pose a significant barrier to implementation. Performance tasks have traditionally been hand-scored by multiple raters, which increases scoring costs and delays the return of scores. With advances in Large Language Models and newer generative artificial intelligence models, rigorously scoring performance assessments cost-effectively is already possible.

Carla Evans wisely cautioned us about the risks of using performance tasks for high-stakes accountability, given the potential for negative consequences. But what if the accountability consequences, strict alignment, and comparability requirements were tamped down? The value proposition of performance tasks is well established, particularly for supporting deeper learning and related initiatives.

Assessments Based on Learning Progressions

Typically, summative assessment designers begin with a grade-level blueprint that specifies the relative emphasis of content domains for a given grade. Assessment items are then developed to ensure an appropriate sampling of those domains. However, by focusing explicitly on grade-level content, we may be missing the opportunity to assess what students know across a continuum of knowledge. Learning progression-based assessments could provide an opportunity to more accurately assess student thinking across content and grade levels.

We suggest an alternative approach to assessment design could begin with the question "What does my student know about [insert content domain]?" rather than "Does my student know 4th grade [insert content domain]?" This shift in approach would change the distribution of content students took in the assessment and adjust reporting accordingly. For example, rather than receiving a score report that identifies a student as "Proficient" in 4th-grade math, a report might share that a student is thinking about fractions as a "Fair Share" (a specified level on the learning progression), which is consistent with 4th-grade expectations. One shift is that items might be targeted at levels above or below grade level to identify how students think about the content.

Possible Federal Flexibility

Implementing a learning progression-based assessment approach would require relaxing the federal requirement to assess students on grade-level content. Learning progressions span multiple grade levels, so identifying how a student is thinking in relation to the progression may require assigning the student items above or below grade level.

Opportunities

Most summative assessments are not currently designed to provide student-level feedback to support instructional next steps. One exception, however, is the Dynamic Learning Maps (DLM) assessment system, which is designed to provide instructionally-relevant information built on underlying student learning trajectories, or "learning pathways." DLM is an alternate assessment for students with the most significant cognitive disabilities. In a recent study, 84% of surveyed teachers agreed that the DLM instructionally embedded assessments were useful for their instructional practice.

Similarly, the ATLAS group recently piloted the Pathways for Instructionally Embedded (PIE) Assessment, which used a format similar to DLM but was designed for general education students in 5th-grade mathematics. Teachers reported using the PIE learning pathways to understand how standards connect to each other. A shift to a learning progression-based assessment system could provide more information to students, teachers, and parents about student progress through school. In addition, clearly defining learning progressions could help create a common language and professional development opportunities for teachers to better understand trajectories of student thinking across grade levels. In each content area, there are concepts that span multiple grade levels. Using these content areas as the foundation for assessment design and development across all grade levels could support a more connected educational experience for students.

Challenges

Learning progressions are difficult to develop and validate. They describe typical patterns in how students think and learn within a domain as they move from fragile to deeper levels of understanding. As such, they represent a common pathway for learning, but not necessarily the pathway taken by all students. Therefore, it is important to use a learning progression as a guideline without assuming that all students will progress linearly.

Using a learning progression approach would require a fundamental shift in how assessments are designed and developed. States would likely have to start from scratch to develop item banks that align with a fundamentally different goal from ensuring full blueprint coverage. DLM is, to our knowledge, the only large-scale assessment system that has successfully applied this approach operationally.

Current federal guidelines require that students be assessed on grade-level content to ensure that all students are provided access to appropriate content. Relaxing the requirement that all students are assessed primarily on grade-level content could encourage some schools or districts to provide content below grade level to select groups of students. Appropriate guardrails would need to be put in place to ensure that did not happen, so we do not return to the pre-IDEA (Individual with Disabilities Education Act) days when students with disabilities were not even exposed to grade-level content. One guardrail could require that any assessment provide valid evidence of a student's knowledge and skills relative to their current grade level before providing additional evidence of their location along a learning progression.

ADDITIONAL FLEXIBILITY

The flexibility options outlined above do not stray far from current assessment laws and regulations. In this section, we offer options that clearly "color outside the lines." We offer three examples and encourage state leaders to consider other options that match their values and goals.

Reduce Every Grade Testing for Every Student

One concern with the current system is that testing at every grade level, from 3 to 8, and once in high school, has a chilling effect on the development of more robust local assessment systems, leading to choices that conflict with deeper learning goals. Some have proposed that reducing the state assessment footprint could free up resources and space, enabling more ambitious initiatives, such as increasing the use of the performance tasks discussed in this paper. In particular, matrix sampling across grade levels and subject areas has been proposed as an option that would continue to allow schools and districts to monitor school-level student growth, without requiring every student to participate for every subject, every year.

Possible Federal Flexibility

ESSA requires that students be tested in math and English/language arts every year in grades 3-8. The yearly testing requirement would need to be lifted or adjusted to implement a testing approach that deviates from this basic requirement.

Opportunities

State summative assessments are not designed to inform individual-level instructional decisions. Rather, in its current form, state assessment is optimally suited for school-, district-, and state-level evaluation and monitoring of student performance. Assessing every student, every year, can send the message that students, parents, and teachers should use the assessment information to support individual student interventions.

> *The ESSA provisions requiring "diagnostic information" have contributed to misconceptions about the limitations of state test scores.*

The ESSA provisions requiring "diagnostic information" have contributed to misconceptions about the limitations of state test scores. Moving to a matrix-sampled accountability system would limit the use of summative assessment for student-level decision making. Matrix sampling involves splitting the full assessment into sections and administering them to different sets of students. For example, if the assessment contained 100 questions, a matrix-sampled approach could involve cutting the assessment into 10 sections and administering each set of 10 questions to one-tenth of the student population.

Reducing required testing creates opportunities for states and districts to invest in defining and implementing their own high-quality local assessment systems. The Center for Assessment led a project examining various reduced-testing scenarios to determine how each performed in a "standard accountability model." Theoretically, states and districts could use the time saved by administering fewer tests to implement measures such as performance assessments or other high-quality local assessments. Developing state and local assessment systems can also provide

valuable professional development for teachers and school leaders, helping them better understand and become invested in the system.

Challenges

Reduced testing should save money and cut state testing time, but it also carries risks, particularly given that the time spent on state tests is relatively limited compared to that on district-required assessments. One important risk is that documenting students' longitudinal growth can be achieved with reduced testing, but measuring growth well without yearly testing is much harder, especially for students with disabilities and multilingual learners. We should promote systems that measure student growth because schools can more directly influence student growth than achievement scores, which are highly correlated with students' and schools' background factors.

A second risk is that, depending on the severity of the accountability system, an every-other-year system could result in assigning the best teachers to tested grades, thus unintentionally skewing who has access to quality instruction across subjects or grade levels. Finally, matrix-sampled assessments—NAEP being the most well-known—do not produce individual student scores. It could be a tough sell to have students participate in an assessment if they do not receive a score. While we are hopeful that reducing state testing could give districts more freedom to create higher-quality local assessments, we have not seen much evidence that this will, in fact, occur.

While we are hopeful that reducing state testing could give districts more freedom to create higher-quality local assessments, we have not seen much evidence that this will, in fact, occur.

Curriculum-Embedded Assessment System

If states were interested in a more adventurous approach, they could consider incorporating performance tasks in a curriculum-embedded assessment system. A curriculum-embedded assessment system is an approach in which measures of student learning are intentionally integrated into the curriculum. This differs from instructionally-embedded assessments, which are administered during instruction but are not explicitly linked to a particular curriculum product or sequence.

Using a curriculum-embedded approach, assessments are designed alongside the curriculum and are embedded within lessons, units, or instructional routines so that evidence of student understanding is generated as students engage with core content and tasks. Specifically, we propose a curriculum-embedded system that uses information from early in the year to document students' opportunity to learn, but does not count towards a summative score. Then, students are assessed at the end of the year using a curriculum-aligned assessment or performance tasks to probe their ability to engage deeply with the learning expectations.

Possible Flexibility

Depending on the assessment design, very little additional flexibility would be required to implement a curriculum-embedded assessment. The biggest challenge would be moving to a centralized state curriculum. Some federal flexibility could be needed if a state wanted to

incorporate strong performance tasks. Specifically, relaxing some of ESSA's strict alignment and comparability requirements could encourage more states to embed rich performance tasks into the proposed system.

Opportunities

Louisiana's Innovative Assessment Program (IAP) was an example of a through-year assessment model on known curricular units. Such a model ensures that all students have access to the same curriculum and the opportunity to learn the relevant content. In addition, linking assessment to the curriculum helps students build background knowledge, so any attached assessment can draw on ideas, concepts, and content from the specific curriculum rather than being curriculum-agnostic.

End-of-year performance tasks can be used to evaluate how well students generalize their knowledge to more complex and less familiar contexts. The assessments administered throughout the year carry much of the alignment and reliability work, but the performance tasks also serve an important signaling function and document the depth to which students have learned the required knowledge and skills.

Challenges

Currently, most school districts have autonomy to select curriculum. Moving to a state-wide curriculum would mark a significant shift in the locus of curricular control. A state-led curriculum and assessment system would likely need to be designed slowly, with buy-in and input from districts, to build investment across districts. Basing curriculum design on a solid, research-based learning model describing how students learn concepts across content standards might help with buy-in.

Additionally, to avoid distorting incentives, teachers, schools, and districts should be aware that curriculum-embedded assessments do not affect a student's summative score. This could help avoid creating incentives to cheat, thus better messaging that the assessment results should be used to support instructional decision-making. However, producing large-scale assessment results that actually support instructional decision-making is challenging. Therefore, such a program would need to thoughtfully design reports that students and teachers can use to inform next steps in instruction.

Finally, any curriculum-embedded assessment system would also need to intentionally consider how to assess transfer and application of learning beyond the curriculum. We hope that students not only learn exactly what is taught in any given curriculum but also apply their skills to unique contexts and problems. Therefore, a curriculum-embedded system should purposefully include tasks or items that assess skills in novel contexts.

Allow Local District Flexibility

Few people were aware of the "local assessment option" under NCLB. This provision permitted states to work with one or more school districts to implement an assessment system that differed from the main state system. As far as we know, only two states—Utah and Mississippi—took advantage of this provision. Utah briefly enabled two districts to use NWEA's MAP in place of the state assessment system. However, when it came time to submit the evidence to meet the U.S. Department of Education's peer review process, Utah and the districts abandoned this option. Mississippi supported two districts in using the Cambridge International Exams in place of the state high school assessments. This program operated for several years before being discontinued.

In some sense, the Innovative Assessment Demonstration Authority (IADA) under ESSA was a vehicle that allowed states to operate two (or more) assessment systems simultaneously. However, the requirement to scale the innovative assessment system statewide in seven years has proven to be the most challenging aspect of the IADA. Innovation and scaling statewide are opposing forces. The more innovative the system, like New Hampshire's PACE and Louisiana's IAP, the more challenging it is to convince all districts to participate in the innovation.

In both cases, the Center established comparability between the state legacy system and the innovative pilot. Given this comparability, we did not think it made sense to force all districts to adopt a single system. Therefore, if states were interested in promoting richer assessment approaches, but worried that not all districts are ready to move in that direction, we see no reason why states couldn't run two or maybe three assessment systems tailored to the different needs of districts. These assessment approaches could be organized by curricular preferences or other means. Obviously, there would need to be some guardrails built for this type of flexibility. Further, state leaders interested in such an approach would need to significantly increase the capacity of state and local assessment offices to support this work.

SOME INNOVATION CAUTIONS

Through-Year Assessment Program

A through-year assessment program is one in which assessments are administered at multiple points during a school year, and the responses are aggregated to produce a summative determination. Despite significant enthusiasm about the possibility of through-year assessments, states have struggled to scale through-year programs, with only two states (Montana and Delaware for social studies) operationally implementing a full through-year model for the general student population (DLM's Instructionally Embedded Model is used in several state alternate assessment programs). In particular, we have been concerned that several of the newer assessment initiatives have inadvertently promoted more atomized approaches to learning, where students are instructed on specific content, assessed on that content, and then move on without returning to demonstrate that they can use that content or skills again. This appears to be more of a problem in math than in reading/language arts.

Possible Flexibility

No additional federal flexibility is needed to implement a through-year model. Several states are currently pursuing through-year assessment models. Federal law requires states to produce a summative determination; however, it does not specify that the determination be derived from a single test administration.

However, additional flexibility along the lines mentioned previously could help make through-year assessment systems more robust and better represent student learning over time. In particular, loosening ESSA's comparability requirements slightly could help states integrate performance-based assessments into their through-year model.

Opportunities

A proposed benefit of through-year assessments is that they can provide information to students and teachers at multiple points throughout the year. If assessment results are to support instruction, they must be timely, foster a range of thinking, focus on actionable sets of standards, and reveal student thinking.

However, when through-year assessments include rich learning opportunities, such as those provided by performance tasks, they can offer a robust alternative to current summative assessment models. However, care is needed to avoid the challenges listed below.

Challenges

A major concern with implementing through-year models in practice is the potential for such systems to atomize learning into discrete, disconnected assessment experiences, thereby negatively shaping teaching and learning. This may be a particular concern if the assessments primarily contain items that ask students to recall knowledge or apply basic skills, but do not engage more complex reasoning. In certain modular approaches in which only a small subset of the domain is tested and not revisited, schools and teachers may be encouraged to teach their subjects as disconnected sets of knowledge. Such atomized approaches to teaching do not support enduring student learning. This is not a defining characteristic of through-year models, as a class of assessments, but it has been common to many of the designs we have seen in practice, except for Louisiana's IAP and New Hampshire's PACE.

As with curriculum-based assessments, when assessments are given throughout the year with the intent to inform instruction *and* to count toward a summative determination, the summative use of the score is likely to distort its instructional usefulness. If students and teachers feel that their responses to an assessment are high-stakes, it can lead to increased anxiety and a reduced willingness to try and make mistakes, both of which can distort student performance. Again, this is not necessarily an assessment design concern but an assessment use issue due to the accountability requirements.

A final challenge of implementing through-year assessments is that combining scores over time changes the interpretation of a summative score. Most through-year systems are intended to make claims about what students have learned by the end of the school year (i.e., a summative determination). The score no longer reflects student skills and knowledge at a single point in time; instead, it includes student performance across multiple points during the year. If a student grows over time, including assessments from early in the year may underestimate a student's true ability by the end of the year. Conversely, if the type of "teach it and forget it" instruction that we are concerned about does take place in practice, student scores may be inaccurately inflated relative to what students know and retain at the end of the year, raising questions of whether the student did, in fact, "learn" the related content.

"Choose Your Own Adventure"

The federal Standards and Assessment Regulations, as instantiated through the peer review process, are meant to ensure that all state assessments meet rigorous technical quality standards. As we've discussed, federal law requires that all students in a state take the same assessments. While some of the above options would require additional flexibility, we do not endorse removing all requirements to meet those standards. Summative assessments cannot serve their evaluative and monitoring purposes if all students in the state take different assessments, or if those assessments cannot be compared across local education systems.

Thus, we caution states against falling into a "choose your own adventure" approach. Such an approach might arise if states allowed districts to identify their own assessments without meeting any technical or comparability requirements. In such a case, a state would no longer be able to use the assessment results to identify schools and districts whose students were not meeting expectations and thus needed additional support.

In addition, states may be tempted to rely on existing commercial interim products to fulfill accountability needs. We caution states against doing so. Instead, states should begin with a strong vision for teaching and learning, and build their assessment around that vision. Existing commercial and interim assessments were designed to provide more frequent updates on student progress towards state academic standards. The assessment design and content often mirror state summative assessments, without necessarily being tied specifically to each state's standards. A commercial assessment might meet an accountability need, likely with some modification, but choosing one should not be done out of convenience. Rather, the assessment should provide the best way to determine whether students are meeting the state's vision for teaching and learning.

States may be tempted to rely on existing commercial interim products to fulfill accountability needs. We caution states against doing so.

A Proliferation of Purposes

We issue a final warning: avoid falling into the trap of trying to do too much with a single assessment. Typically, a single assessment only serves one or two purposes well. Most state summative assessments are designed primarily for use by state, district, and school leaders. Asking teachers and students to use that information to adjust teacher practice or enhance student learning is likely asking them to use a hammer when they need a screwdriver. Many of the flexibility options suggest attending to the assessment *system* as a whole, in which different assessments can be used for particular purposes. Simply saying that an assessment serves multiple purposes does not make it so.

CONCLUSION

We wrote this paper to help state assessment leaders begin planning for anticipated federal flexibility. We provide several examples of assessment changes we could envision and would welcome. We approached this exercise analytically by describing the potential assessment change, the flexibility required to implement it, and the trade-offs associated with the change. We encourage others to approach any calls for assessment flexibility in a similar manner.

Of course, this would not be a Center for Assessment paper if we did not mention the use of a theory of action to guide any design activity. One of the key features of any principled design process is clear problem identification. Engaging in such an activity in response to calls to change the state assessment system might reveal that the real problem is the accountability uses of the test scores. In that case, almost any assessment change will raise concerns if the uses have not changed. Therefore, we encourage accountability pilots in conjunction with assessment innovations to better help the state realize its goals. This does not mean waiving accountability. The state is responsible for ensuring that all of its students have legitimate learning opportunities and outcomes.

Similarly, we often hear complaints about "too much testing." However, as Education First recently documented, much of the "test pile" is due to district- and school-selected assessments, not the limited time devoted to state testing. We think that here, too, the accountability use case distorts the

image of time. In any case, the important point here is that state leaders must be prepared to engage in a deliberative design process to ensure that any changes to the state assessment system support the intended purposes and uses while minimizing potential unintended negative consequences.

A final warning: avoid falling into the trap of trying to do too much with a single assessment.

Center for Assessment

National Center for the Improvement of Educational Assessment, Inc.

Dover, New Hampshire

www.nciea.org