

ASSESSMENTS 101: A policymaker's guide to K-12 assessments

JULIE WOODS

Assessments serve a variety of purposes for stakeholders at all levels of the state education system. Because assessments play such an integral role in learning, teaching and accountability, policymakers can benefit from having a working knowledge of the assessment landscape and common terms used for discussing assessments. This brief supports state leaders' understanding of assessments by first classifying and categorizing assessments and then providing an overview of common terms used when choosing and utilizing assessments.



**EDUCATION
COMMISSION
OF THE STATES**
Your education policy team.

The Many Purposes of Assessments

Assessments come in many forms in part because they serve many purposes, and those purposes often vary by the stakeholders they support. Students, parents, teachers, and school, district and state leaders may all be end users of the information provided by various assessments. For example, assessments can support the needs of:

Students and Parents:

- By informing students and parents about the student's progress in learning content based on the state academic standards.
- By informing students and parents — as well as teachers and schools — about the student's readiness: for grade advancement, graduation, college and careers.

Teachers and School Leaders:

- By allowing teachers to better plan and tailor instruction to student and classroom needs.
- By supporting teachers and school leaders in identifying where students need intervention, remediation or acceleration.
- By holding teachers and schools accountable, and identifying opportunities for their growth through teacher evaluations and school report cards.

LOOKING FOR MORE INFORMATION ABOUT TESTING?

- [Thinking About Tests and Testing: A Short Primer in "Assessment Literacy."](#)
- [Using Balanced Assessment Systems To Improve Student Learning and School Capacity: An Introduction.](#)
- [Designing a Comprehensive Assessment System.](#)

TESTING AND FEDERAL LAW

The Every Student Succeeds Act (ESSA) requires state education agencies to implement statewide assessments in:

- Mathematics and English-language arts (ELA) in third-eighth grade and once in ninth-12th grade.
- Science once in each of the following grade spans: third-fifth grade, sixth-ninth grade and 10-12th grade.

Many states exceed the minimum federal testing requirements by mandating, for example, a social studies or college and career readiness assessment statewide. Education Commission of the States' [summative assessments database](#) provides information on the required statewide assessments in all 50 states plus the District of Columbia (D.C.).

Districts and States:

- By informing districts and the state about school performance, allowing them to determine the appropriate interventions for low-performing schools and to recognize high-performing schools.
- By allowing for comparisons of student subgroups, schools, districts and, when possible, states.
- By informing district leaders' and state policymakers' education policy decisions.

State and local leaders often want state tests to accomplish as many of these purposes as possible while simultaneously:

- Limiting financial, time and operational burdens.
- Providing information-rich and timely results.
- Measuring deep content knowledge and relevant skills.

Yet research cautions that tests should only be used for the purposes for which they were designed, which means that multiple tests may be necessary to accomplish all the purposes needed for a state's education system. Given this tension, state leaders must balance efficiency and limited testing with the need for information that can best support student success.

Research cautions that tests should only be used for the purposes for which they were designed, which means that multiple tests may be necessary to accomplish all the purposes needed for a state's education system.

Classifying Assessments

Assessments come in many shapes and sizes depending on the purpose(s) they serve. The following questions can serve as a guide when mapping the landscape of different assessments.

When does assessment occur?

- Before learning — diagnostic tests (identify gaps).
- During learning — formative tests (inform instruction).
- At key points in learning — interim tests (identify specific gaps).
- After learning — summative tests (determine mastery).

Which transition does the assessment support?

- Preschool to kindergarten — kindergarten entrance exams.
- Grade to grade — summative assessments.
- Course to course — end of course assessments.
- High school to college or career — exit exams and college entrance exams.

What is assessed?

- Mastery of core academic content standards.
- Proficiency in areas of a well-rounded education:
 - Arts — **example:** short answer responses to works of art.
 - Civics — **example:** U.S. Citizenship test.
 - Health — **example:** physical fitness assessment measuring strength, endurance and flexibility.
- Social-emotional knowledge and skills — example: student surveys.
- Readiness for college and careers - example: the SAT or ACT WorkKeys.

How is the information assessed?

- Multiple choice questions.
- Constructed responses, essays.
- Performance tasks.
- Portfolio of student work.

ASSESSING SOCIAL-EMOTIONAL LEARNING: WHAT DO I NEED TO KNOW?

While still a developing field, student and teacher surveys can assess social-emotional learning (SEL).

- Currently, leading researchers recommend against using social-emotional surveys for high-stakes decisions.
- Few states have articulated learning goals for SEL beyond preschool. All states incorporate SEL into their preschool standards.

 *Want to learn more?*

- [Outcomes Beyond Test Scores – What Is Social-Emotional Learning?](#)
- [Transforming Education Resources for Policymakers.](#)
- [Collaborative for Academic, Social and Emotional Learning \(CASEL\).](#)

Glossary of Terms for Choosing and Using Assessments

Given the wide variety of assessment purposes and uses, many of the terms commonly used to describe assessments can have different meanings when used in different contexts. While not exhaustive, a brief glossary of assessment terms that frequently lead to policymaker questions, and accompanying ESSA implications where applicable, follows. For ease of reference, terms are divided into two categories: terms likely to arise when choosing appropriate assessments for a specific purpose, and terms likely to arise when implementing chosen assessments. Terms are presented alphabetically in each category.

Choosing Assessments

Assessment Audit

In an assessment audit, states inventory which and how many assessments are administered at the state and/or local levels. These audits can help states alleviate the testing burden on students and teachers by eliminating unnecessary or redundant testing. ESSA provides funds states can use to conduct an audit.

Competency-Based Assessments

In a competency-based education system, students progress through a unit of study at their own pace based on their demonstrated mastery of knowledge and skills. Because competency-based education is relatively uncharted, assessments aligned to such systems are challenging to define and can vary significantly. Existing competency-based assessments are typically locally-developed and incorporate performance tasks. **New Hampshire's** Performance Assessment of Competency Education (PACE) is the most well-known and developed example of this type of learning and assessment.

 *Want to learn more?*

- [Assessment to Support Competency-Based Pathways.](#)
- [Two Sides of the Same Coin: Competency-Based Education and Student Learning Objectives.](#)
- [New Hampshire Performance Assessment of Competency Education \(PACE\).](#)

Computer-Adaptive Assessments

A computer-adaptive assessment (CAT) adjusts the difficulty of questions during an exam — based on a student's response — and is distinct from computer-based assessments that replicate traditional tests on a computer. ESSA explicitly permits states to develop and administer CATs for math, ELA and science, and does not prohibit states or districts from using CATs for other tested subjects. Under No Child Left Behind (NCLB), states could only use CATs following approval granted through the U.S. Department of Education's peer review process and federal waivers.

Innovative Assessment Pilot

ESSA provides an opportunity for a limited number of states to pilot an innovative assessment system in some or all of their districts. The law provides an open-ended list of possible innovative assessment options: performance-based, instructionally-embedded, competency-based, portfolios or several interim tests — rather than a single summative test — among other options that the law leaves open. For states, the long-term goal is to implement a high-quality innovative assessment system statewide for accountability purposes.

 *Want to learn more?*

- [15 Assessment Designs for the Innovative Assessment Pilot.](#)
- [Innovative Assessment Demonstration Authority Pilot under ESSA: Frequently Asked Questions.](#)
- [ESSA: Quick guides on top issues.](#)
- [Deeper learning: A primer for state legislators.](#)
- [Curriculum-Embedded Performance Assessments \(CEPAs\): Policy considerations for meaningful accountability.](#)

Interim Assessments

Unlike summative assessments, which measure student achievement at the end of a course of study, interim or benchmark assessment are administered at intervals throughout a course of study. Interim assessments allow for predictions of how well students will perform on subsequent assessments, including summative assessments. Like diagnostic and formative assessments, these tests can help teachers tailor instruction to students' needs, and like summative assessments, these tests can demonstrate how well students have mastered the content from a sub-unit of study. New flexibility under ESSA permits states to administer math, ELA and science tests as statewide interim assessments — rather than summative assessments — and combine the scores into a single summative score used for accountability purposes.

 *Want to learn more?*

- [Interim Assessment Resources.](#)
- [Distinguishing Formative Assessment from Other Educational Assessment Labels.](#)
- [“Interim” Assessments and ESSA: A Great Opportunity.](#)

Nationally Recognized Assessment

New provisions in ESSA authorize districts to administer a locally selected, nationally recognized assessment in high school in place of the state-determined, statewide assessments required for math, ELA and science. While the law does not define “nationally recognized,” proposed, but rejected, regulations described it as “an assessment of high school students’ knowledge and skills that is administered in multiple states and is recognized by institutions of higher education in those or other states for the purposes of entrance or placement into credit-bearing courses in postsecondary education or training programs.”¹ Experts identify the SAT, ACT, PARCC and Smarter Balanced assessments as likely candidates for this assessment.²

 *Want to learn more?*

- [ESSA: Quick guides on top issues.](#)

Next Generation Science Standards — Aligned Assessments

The Next Generation Science Standards (NGSS) are content standards for physical science, life science, earth/space science and engineering. Based on the idea that science is not only a subject area but also an activity, these standards were developed with scientific and engineering practices folded into the descriptions of what students should know and be able to do. As a result, NGSS-aligned assessments may use performance tasks or other, more interactive formats to allow students to demonstrate their scientific knowledge and skills simultaneously, such as by conducting an experiment.

Since their publication in 2013, 21³ states have adopted the NGSS as their statewide science standards. At least three states — [Illinois](#), [Kansas](#) and [Nevada](#) — plus [D.C.](#) currently administer a statewide NGSS-aligned assessment, and at least five states — [California](#), [Connecticut](#), [Delaware](#), [Kentucky](#) and [Oregon](#) — are piloting or transitioning to a statewide NGSS-aligned assessment.

 *Want to learn more?*

- [Next Generation Science Standards.](#)
- [Developing Assessments for the NGSS.](#)

Performance Assessments

Performance assessments are a broader category of a variety of non-traditional assessment methods. These can include performance items, curriculum/classroom-embedded tasks, portfolios and student-designed projects. Performance assessments differ from traditional tests in that they typically involve an activity in which students demonstrate their knowledge and skills. Additionally, these tests may better measure students' problem-solving and critical-thinking skills.

 *Want to learn more?*

- [Developing and Measuring Higher Order Skills: Models for State Performance Assessment Systems.](#)

Using Assessments

Alignment

Alignment refers to how well tests align with other key aspects of a student's educational experience. For tests to be useful for the purposes for which they are intended, tests should align with the state standards, curricula, instructional materials, teacher training and teacher content delivery. In other words, students should be taught content that aligns to state standards, and tests should assess what students have actually been taught.

Cut or Threshold Scores

Cut or threshold scores are the scores students must achieve to reach a particular performance level; for example, scoring proficient on an assessment might require a score between 400-450. Where cut scores are set can determine the difficulty of an assessment. High-stakes decisions, such as teacher performance on evaluations, are often attached to students' ability to reach a certain performance level, making the process of setting these thresholds an important one for state leaders.

Participation Requirements

ESSA, similar to NCLB, requires states to annually test at least 95 percent of all public-school students and each subgroup in math and ELA. In 2014, dissatisfaction with testing time and quantity, as well as other concerns, gave rise to a movement in which students and parents opted out of mandatory assessments. Many states responded to these concerns by, for example: 1) eliminating statewide assessments not required for federal accountability, such as [social studies assessments](#) or additional high school assessments; 2) replacing high school assessments with a [college entrance exam](#) to minimize testing in high school; 3) limiting administration time of [state](#) or [local](#) assessments; and 4) [increasing transparency and reporting around testing requirements.](#)

 *Want to learn more?*

- [State Legislatures Opting in to Opting Out.](#)
- [Who opts out and why? Results from a national survey on opting out of standardized tests.](#)
- [ESSA: Quick guides on top issues.](#)
- [Assessment Opt-Out Policies: State responses to parent pushback.](#)

Performance Level Descriptors

Performance level descriptors identify what students know and are able to do at each level. For example, in a system with four different performance levels (1-4), a score at or above level 3 in 11th grade could demonstrate

the knowledge and skills necessary to be ready for college coursework.

Reliability

Reliability of a test refers to the degree to which results are consistent for the test-taker across multiple attempts in similar conditions. In other words, if a test is reliable, the results for an individual test-taker wouldn't change if the test is taken this week and again next week.

Validity

Validity of a test refers to the degree to which a test accurately measures what it is intended to measure. If a test is valid, it measures reliably across contexts. For example, students living across a state can expect to consistently receive accurate results.

For a more nuanced discussion of validity, see: [Thinking About Tests and Testing: A Short Primer in “Assessment Literacy.”](#)

ENDNOTES

1. Elementary and Secondary Education Act of 1965, As Amended by the Every Student Succeeds Act – Accountability and State Plans, 34 Fed. Reg. 200 and 299 (May 31, 2016). Federal Register: The Daily Journal of the United States.
2. Erin O'Hara, “Choices and Trade-offs: Key Questions for State Policymakers when Selecting High School Assessments,” High-Quality Assessment Project, <http://education-first.com/wp-content/uploads/2016/05/Education-First-HQAP-High-School-Assessment-Guide-May-2016.pdf> (accessed September 18, 2018).
3. State of the States, Center on Standards & Assessment Implementation at WestEd, <http://www.csai-online.org/sos> (accessed September 18, 2018).

AUTHOR

Julie Rowland Woods is a policy analyst in the K-12 Institute at Education Commission of the States. She holds a J.D. and M.A. in Education Policy from the Pennsylvania State University. When she's not busy working with the K-12 team, Julie is usually trying to find ways to be more like Leslie Knope. Contact Julie at jwoods@ecs.org or 303.299.3672 or tweet @JulieRoWoods.

© 2017 by Education Commission of the States. All rights reserved. Education Commission of the States encourages its readers to share our information with others. To request permission to reprint or excerpt some of our material, please contact us at 303.299.3609 or email askinner@ecs.org.