Innovation in Assessment and Accountability Systems

April 2026

To provide state departments of education with information on the current state of innovations in assessment, the Central Comprehensive Center conducted a review of the policy and technical literature on the design and uses of state and district assessments. This brief focuses on the design of assessment systems, including the roles of within-year assessments, to enable measurement of student achievement and growth and support of instruction. The appendix lists assessment innovations currently being considered or implemented in selected states. Additional information on the design and expanded uses of assessment and accountability systems and state innovations in these systems can be provided.

The Purpose and Design of Assessment Systems

State and district assessment systems are expected to serve multiple purposes. They must: (1) provide valid and reliable measures of student achievement and growth for accountability, (2) support educators with timely information for instruction, (3) offer diagnostic insight for struggling learners, and (4) minimize burden on schools. These goals often conflict. Assessments designed for accountability generally require standardized administration, broad content coverage, and psychometric precision. Instructionally focused assessments generally require flexibility, targeted standards coverage, and rapid feedback. Fortunately, it is possible to design an assessment system that serves multiple purposes by including different types of assessments, some primarily designed to support accountability and some primarily designed to support instruction. States are increasingly considering balanced assessment systems that address this need.

Components of Comprehensive Assessment Systems

A comprehensive assessment system could include some or all of the following assessments, differentiated by their primary use.

End-of-year summative assessments. Required, comprehensive, aligned to grade-level standards, and administered in narrow windows to ensure comparability. Summative assessments are used to construct end-of-year student and school-level measures of proficiency and growth.

Within-year assessments: interim assessments. Administered multiple times per year (typically Fall, Winter, and Spring), they provide updated estimates of achievement and progress. Interim assessments are generally comprehensive, aligned to grade-level standards, and measured on the same scale as summative assessments (if provided by the summative assessment vendor), but may or may not be administered in narrow testing windows. Interim assessments generally contain fewer items than summative assessments. We discuss below whether interim assessments can contribute to measurement of summative achievement and/or growth in achievement.

Within-year assessments: focused assessments. Flexible, short assessments focused on subsets of standards, administered many times during the school year, and not administered in narrow testing windows. These assessments are useful for classroom instruction but not for accountability.

Within-year assessments: through-course assessments. Distributed assessments that test subsets of standards across several administrations and typically are not administered in narrow testing windows. Through-course assessments may be administered 2-3 times during the school year or many times. Most through-course assessment systems currently administer an end-of-year summative assessment but may also include an end-of-year assessment that covers a subset of standards. Through-course assessments that measure subsets of standards operate much like focused assessments, particularly if administered many times during the year. As a result, these assessments contain many fewer items than summative and interim assessments. We discuss below whether through-course assessments could contribute to measurement of summative achievement and/or growth in achievement.

Diagnostic assessments. Designed to identify specific skill gaps among low-performing students. Diagnostic assessments are typically measured at specified times during the school year to inform selection into programs for low-performing students.

Computer Adaptive Testing (CAT). Enables tests to match item difficulty to student ability, improving precision while potentially reducing total testing time. Summative, interim, through-course, and diagnostic assessments are often administered using CAT. Focused assessments may or may not be administered using CAT. Most assessments can be administered online using fixed form or adaptive tests.

Uses of Summative Assessments and Within-Year Assessments

Summative assessments remain the backbone of state accountability systems, including systems that incorporate within-year assessments, for several reasons. They comprehensively measure all grade-level standards, test students within narrow test windows, produce comparable results across students and schools, and enable measurement of student and school-level growth. Because they assess learning at the end of the instructional cycle, they capture the cumulative effect of teaching and learning on all learning standards. This approach reduces the incentive for students and educators to focus on narrow parts of the curriculum and short-term content mastery when scores are used for student or school accountability.

Summative assessments from prior years also provide important information about student status at the beginning of the school year and thus can be used in early warning systems and to help educators prepare to support diverse student needs. However, summative assessments are not designed to inform students' progress during the school year.

Interim assessments fulfill that function, particularly if they are administered in a narrow testing window. The Fall interim assessment also provides information on beginning-of-year performance for students new to a district or state.

Within-year assessments (through-course and focused assessments), particularly those that assess small subsets of standards many times during the school year, are explicitly designed to inform instruction. They are intended to operate somewhat like high-frequency classroom assessments but differ from classroom assessments in that they are created by test vendors whereas classroom assessments are always created by teachers based on classroom-specific curriculum pacing and student instructional needs. One of the possible advantages of using vendor-

provided through-course or focused assessments is that these assessments are designed using the same test blueprint as the summative assessment and thus may emphasize teaching to the state standards embedded in the state summative assessment.

Can Interim Assessments Be Used to Create Accountability-Based Metrics: Summative Achievement and/or Growth in Achievement?

In order to answer this question, we need to consider the four conditions described below.

- ***Condition 1:*** Do the assessments lead to teaching to the test? One of the concerns with using student test scores for accountability purposes is that doing so increases the "stakes" on these assessments, thus potentially amplifying the incentive to "teach narrowly to the test." This incentive is much reduced when summative and interim assessments are designed to comprehensively cover all learning standards.
- ***Condition 2:*** Do the assessments provide information comparable across students and schools? For accountability purposes, test scores need to be comparable across students and schools, and thus need to be administered within specified, narrow testing windows. Summative assessments satisfy this requirement, and interim assessment administration timing can be restricted to satisfy this requirement.
- ***Condition 3:*** Are the assessments informative about current end-of-year proficiency and growth? End-of-year measures are intended to measure what students know at the end of the year and how much they have gained. In contrast, within-year assessments capture prior year achievement plus growth in achievement from the prior year to the point in time of the within-year assessment (which can vary across students and schools).
- ***Condition 4:*** Does the augmented assessment or accountability measure exhibit substantially increased reliability? An augmented metric is inevitably more complex than an original metric. The value of using a more complex metric depends on whether and how much the added complexity improves the quality of the metric.

For **growth measures**, the answer can be "yes" for all four conditions. How? There are new methods for using interim test scores to increase the reliability of growth measures. These approaches also report a growth measure at each interim test date that is optimally predictive of the end-of-year growth measure. To minimize the number of assessments used for high stakes purposes, it may be prudent to incorporate only a single interim assessment in the augmented growth measure, preferably, the mid-year/winter interim assessment. One major advantage of this new approach is that it can produce substantial increases in precision of growth measures for small schools. As a result, implementation of this approach may allow the state and test vendors to reduce the number of items on the summative and interim assessments. The increase in precision is likely higher for interim assessments designed using the same test blueprint (and the same test vender) as the summative assessment, compared to interim assessments provided by different vendors.

For **end-of-year summative test scores,** the answer is "no" for all four conditions. Summative test scores, based on end-of-year summative assessments, are typically estimated with high reliability, particularly if the assessments utilize computerized adaptive testing (CAT), as is the case with most summative assessments. End-of-year interim assessments (if administered) could potentially increase the precision of summative test scores, but they create test burden without appreciably improving the reliability of summative test scores. Furthermore, mid-year assessments and especially beginning-of-year assessments are problematic because they are more reflective of prior-year proficiency than current-year proficiency, a violation of Condition 3.

Can Through-Course Assessments or Focused Assessments Contribute to Accountability-Based Metrics: Summative Achievement and/or Growth in Achievement?

No, for both summative achievement and growth measures. Through-course and focused assessments fail to satisfy Conditions 1, 2, and 3, presented above. This finding clarifies the appropriate roles of through-course or focused assessments. They are designed to and best used to inform instruction.

Example Assessment System with High Validity and Reliability but with Reduced Testing Time

The example system below is a system that states could realistically implement but without requiring substantial increases in mandated assessments. Districts could have the freedom to select one or more of the three options or none of them. The system could also be expanded to make some or all of the options required. A state assessment and accountability system can be designed to accommodate the different choices made by school districts. Districts that value more reliable estimates of growth would have the incentive to adopt the Winter interim assessment. Districts or schools with large numbers of new teachers might elect to use state and vendor provided focused assessments, rather than relying only on educator-created classroom assessments. Districts or schools with high rates of mobility might elect to require the Fall interim assessment.

- Required: summative assessment with comprehensive, full standards coverage.
- Optional: Winter interim assessment with comprehensive, full standards coverage and a narrow testing window. This assessment will enable implementation of an assessment and accountability innovation to strengthen growth measurement.
- Optional: Fall interim assessment, only for students entering the state system.
- Optional: focused assessments as flexible instructional tools.

This model reduces testing time by avoiding multiple mandated interims while still enhancing the reliability of summative achievement and growth measures through the use of a winter interim assessment.

Prepared by: Rob Meyer and Claudia Gentile
The contents of this brief were developed under a grant from the U.S. Department of Education (Department). The Department does not mandate or prescribe practices, models, or other activities described or discussed in this document. The contents of this brief may contain examples of, adaptations of, and links to resources created and maintained by another public or private organization. The Department does not control or guarantee the accuracy, relevance, timeliness, or completeness of this outside information. The content of this report does not necessarily represent the policy of the Department. This publication is not intended to represent the views or policy of or be an endorsement of any views expressed, or materials provided by any Federal agency. [2026, NORC at the University of Chicago]

Appendix: State Assessment Innovations

The table below lists some of the assessment innovations being piloted and/or implemented by several states.

State	Topic Area
Georgia	Innovation around revisions to ELA standards and ELA assessments.
Michigan	Innovation around M-STEP literacy assessments to fully locate assessments in the Michigan context, and engage Michigan authors, educators and students in test development, and the use of audio voice.
Pennsylvania	Innovation around migrating all PSSA and Keystone testing online for the 2026 administration; Adopting Technology-Enhanced Items (TEIs) in all possible subject/grade levels; Developing a benchmark assessment tool to be provided to schools; Enhancing reporting and messaging within the State's assessment systems.
Missouri	Innovation around Missouri Approach to Interim Assessments: Focus on foundational skills, learning progressions and the spiraling of content within curricula.
Wisconsin	Innovation around improving educators' assessment and data literacy, A Statewide Initiative for Assessment and Data Literacy to Enhance Student Outcomes Across Wisconsin's School Districts.
Arkansas	Innovations around improving the Arkansas Teaching And Learning Assessment System (ATLAS) enhancing fairness, usability, and students' long-term success. Also, how to best use alternative assessment data to better measure year to year progress of students with complex learning needs.
Colorado	Innovation around integrating authentic student voice into large-scale assessment development through student engagement sessions for the Colorado Measures of Academic Success (CMAS).
Illinois	Innovation around establishing meaningful and coherent performance standards in Illinois: aligning k-12 assessments with transparent, data-driven accountability systems.
North Dakota	Innovations around state-provided interim assessments. Interim assessments have been an area of focus since 2009, and more recently studies in 2021-22.
Iowa	Focus on summative assessment system but strong support for formative assessments.
Montana	Innovations in MAST assessments (ELA and math, grades 3-8) to use testlets administered throughout the year to provide timely and relevant feedback for teachers and students, as well as comparable summative data for education leaders and federal accountability.