

SUPPORTING HIGH-QUALITY STATEWIDE ASSESSMENT: WHAT EVERY STATE POLICY LEADER NEEDS TO KNOW



Produced by the Assessment Literacy Working Group of Council of Chief State School Officers' Technical Issues in Large Scale Assessment State Collaborative & The National Center for the Improvement of Educational Assessment

October 2020

INTRODUCTION

Educational assessments are powerful educational equity tools for guiding and informing teaching, learning, and school and district improvement efforts. They are one of multiple measures used to improve the performance of schools and districts.

Classroom and other local assessments are critical components of the teaching and learning system.

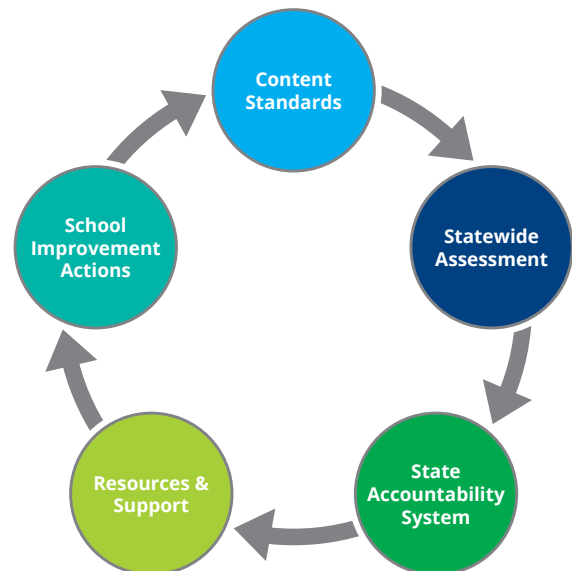
State policy leaders, particularly state education chiefs, legislators, and state board members, play a critical role in defining state and even some local assessments. In addition to relying on state and national assessment experts, state policy leaders will find it helpful to understand key assessment concepts to effectively support state assessment programs that yield meaningful, actionable results. Understanding certain foundational concepts of student assessment can help policy makers positively influence the quality of state assessments and, in turn, the quality of education.

The State Collaborative on Technical Issues in Large-Scale Assessment (TILSA Collaborative), a group of state assessment leaders from more than 40 states, industry partners, and other technical experts identified several important concepts to help state policy makers support high-quality assessment.

COMMUNICATION

Statewide educational assessment is a highly visible aspect of educational policy. As such, it receives considerable attention from the media and stakeholders. State agencies

Figure 1. Assessment as part of a school improvement system.



are responsible for clear, credible, and coherent communication about its assessment program. **Policy leaders should ensure that coherent and credible messages regarding assessments are in concert with the agency's communication strategy.**

PURPOSES AND USES

State assessments are designed to fulfill a specific and limited number of uses and purposes in a valid, fair, and reliable way. **A balanced assessment system includes multiple measures to serve the purposes and stakeholders for which they best fit.**

- ✓ Statewide assessments are complex initiatives. Unlike teacher-made tests, state assessments require considerable planning and attention to create and administer. **State education agencies must allocate the appropriate time and resources in the design, development, and maintenance of state assessments.**
- ✓ Stable statewide summative assessment programs allow educators and school leaders to focus on teaching and learning instead of having to constantly adapt to new state assessments. Stable assessment systems help policy makers and educational leaders more confidently monitor trends.
- ✓ Policy makers should codify in law only those testing requirements that are absolutely necessary. This allows for appropriate flexibility during design and development to yield the most effective and efficient use of the state's resources.
- ✓ Statewide achievement tests provide an estimate of students' knowledge and skills at a single point in time. Multiple assessments of achievement and growth allow for a more robust understanding of student and school performance.

DESIGN AND DEVELOPMENT

Policy makers should have a general understanding of the options available for creating and/or selecting assessments and a basic understanding of important design considerations.

- ✓ Policy makers, educational leaders, and key stakeholders must clearly articulate how and why they expect to use the assessment results. Engaging in this activity helps prioritize design choices.
- ✓ Assessments must give all students a meaningful opportunity to demonstrate what they know and can do. Assessment designers should employ Universal Design for Learning procedures to ensure assessments are as fair and accessible as possible.

Figure 2. Assessment as part of a learning system.

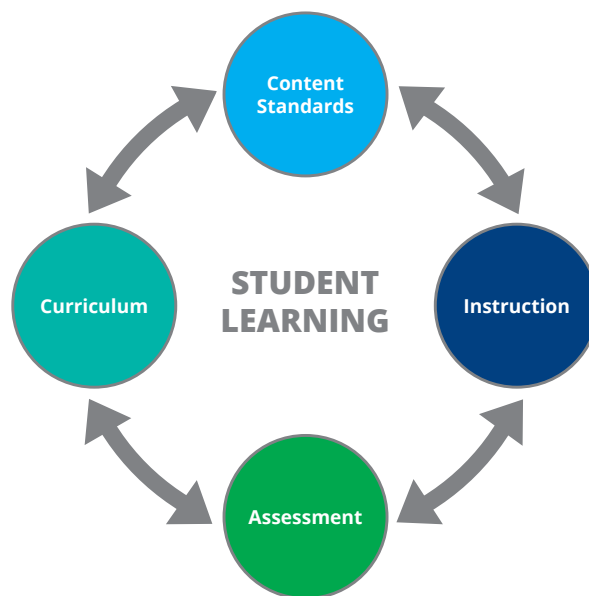
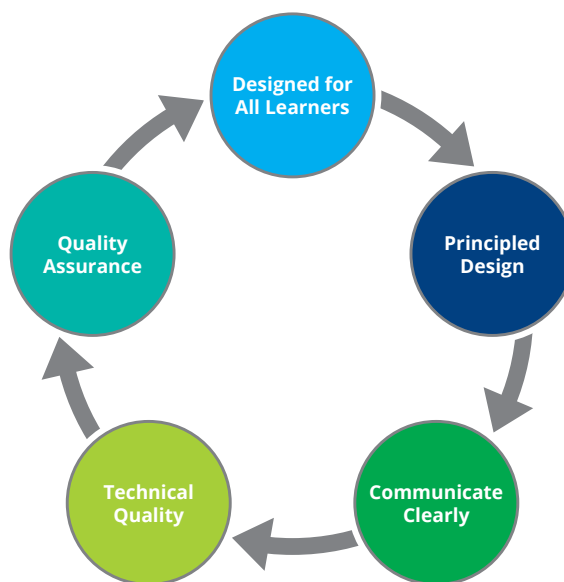


Figure 3. Components of High-Quality Statewide Assessment Process



- ✓ Policy makers should be clear whether the state assessment will be part of a system of assessments or whether it be the sole focus of statewide assessment efforts¹.
 - Statewide summative assessments used for monitoring educational performance or for accountability purposes are the primary focus of state policy makers.
 - Other assessment approaches, such as formative and interim assessments, are used to address different purposes such as supporting instructional decisions, developing student understanding of their own learning, informing curriculum revision, or guiding remediation efforts. Decisions about these assessment approaches are typically the focus of local educational leaders and teachers.
- ✓ State content standards are the starting point for guiding assessment design. Various types of activities and documents are necessary to further specify the standards into assessment targets.

There are many other important aspects of assessment design, such as deciding between adaptive or fixed form testing, the types of items and tasks to include, and types of scores to report, all of which are dependent on purposes, uses, and constraints. However, the major considerations outlined above must come first.

TECHNICAL QUALITY

State assessment and accountability leaders are responsible—with their commercial assessment providers and other external partners—for evaluating, documenting, and ensuring the technical quality of the state assessment system. **Addressing the key dimensions of technical quality depicted below requires considerable evidence, logic, and documentation, and often requires the support of the state’s Technical Advisory Committee (TAC).** These time and resource intensive activities are necessary for meeting the U.S. Department of Education Standards and Assessment requirements through the peer review process and for assuring credibility with stakeholders.

QUALITY ASSURANCE

Producing trustworthy individual and group test scores requires extensive time, planning, procedures, and resources. The following questions are useful guides to help state leaders and their assessment professionals ensure the quality of their testing programs.

- ✓ Are all procedures and processes thoroughly specified and documented?
- ✓ Has the state established clear criteria for evaluating the success of all procedures and processes?
- ✓ Does the state have a system for replicating or otherwise verifying certain critical processes (e.g., equating)?
- ✓ Is the state supporting continuous improvement activities for all aspects of the testing program?

Figure 4. Dimensions of Assessment Technical Quality



CONTACT

Please contact your state assessment director to learn more about how your state is addressing the issues discussed in this brief.

¹ See this policy brief for more information on the challenges and opportunities associated with balanced systems of assessment: <https://www.nciea.org/articles/challenges-and-opportunities-balanced-systems-assessment-policy-brief>

GLOSSARY OF TERMS

Common Types/Uses of Assessments

Summative Assessment: Summative assessments are designed to support various types of determinations (e.g., proficiency, competency) given at the end of a defined instructional period, such as a unit of instruction or a school year, to evaluate students' performance against a set of learning targets for that period.

Formative Assessment: Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes.²

Interim Assessment: Interim assessments (sometimes call benchmark) are assessments administered during instruction to evaluate students' knowledge and skills relative to a specific set of academic goals, in order to inform educator and policymaker decisions at the classroom, school, and district levels. The specific designs of the interim assessment are driven by the purpose and intended uses, but the results of any interim assessment must be aggregable for reporting across students, occasions, and concepts.

Key Aspects of Assessment Quality

Fairness: Fairness refers to the expectations and evidence that all test takers have legitimate opportunities to demonstrate their knowledge and skills called for on the test. A fair test is impartial, accessible, and appropriate for all individuals in the intended population for the intended use of that test.

Peer Review: Peer review is a legally-required process used by the US Department of Education to evaluate the degree to which state assessment systems meet the technical and inclusion requirements spelled out in law and regulations. Peers are individuals with technical and/or operational expertise and experience with state assessment systems.

Quality Assurance: Quality assurance is a way of preventing mistakes and shortcomings in all testing products and processes from testing creation, administration, and reporting. Detailed and replicable procedures must be followed and documented to ensure a high quality testing experience.

Reliability: Generally, reliability refers to the consistency of test scores across real or hypothetical replications of a testing procedure. For example, if we chose different test questions thought to tap the same knowledge and skills as the original test, we would expect similar levels of student performance. Reliability helps quantify this consistency.

Universal Design for Learning (UDL): When applied to assessment design and administration, UDL provides flexibility in the ways information is presented, in the ways students respond or demonstrate knowledge and skills, and in the ways students are engaged. UDL reduces barriers in instruction, provides appropriate accommodations, supports, and challenges, and maintains high achievement expectations for all students, including students with disabilities and students who are limited English proficient.

Validity: Validity refers to the degree to which evidence and theory support the interpretations of test scores for the intended use of that test. Validity is the most important criterion for developing and evaluating tests and test scores. To boil it down, validity asks, "do the test scores mean what we think they mean and what is the evidence to support such claims?"

² See: <https://ccsso.org/sites/default/files/2018-06/Revising%20the%20Definition%20of%20Formative%20Assessment.pdf>