

Test Development and Validity

Michael Maher, Ph.D.

Deputy State Superintendent, Division of Standards, Accountability, and Research

Tammy Howard, Ph.D.

Senior Director, Office of Accountability and Testing

Kinge Mbella, Ph.D.

Assistant Director, Office of Accountability and Testing



Statewide Assessments

- Required by state and federal laws
 - End-of-Grade Tests in reading and mathematics grades 3–8, and in science grades 5 and 8.
 - End-of-Course Tests in reading, mathematics, and science
 - NC Math 1 and NC Math 3
 - English II
 - Biology



Purpose of State Assessments

- The purpose(s) of the statewide assessments include
 - providing data on what students know and are able to do, and
 - using the school performance data to help determine schools' strengths and weaknesses in the identification of those most in need of support.



What makes a good assessment

A good assessment...	Which means...
Meets all constraints	development timeline and budget, test administration time, student reading level, and student developmental level.
Meets all intended purposes	how the assessment score will be used and interpreted.
Is reliable	if a student retakes an assessment without more study, both scores should be similar.
Is valid	the assessment agrees/aligns with the content standards (adopted by the State Board of Education).
Is accessible and fair	all students have equal opportunity to show their knowledge.



Critical First Step

- Content standards are adopted by the State Board of Education.
 - The Elementary and Secondary Education Act (ESEA) requires that the academic content assessments be aligned with the State’s academic content standards and address the depth and breadth of those standards.



Peer Review Process

- After the first administration of the tests, evidence of the assessments' technical quality must be submitted to the U.S. Department of Education for peer review.
 - Assessment peer review is the process through which a State submits evidence to demonstrate the technical soundness of its assessment system
 - Sections 1111(a)(4) and 1111(b)(2)(B)(iii)-(iv) of the ESEA and 34 CFR § 200.2(b)(4) and (5) and (d) specifies the U.S. Department of Education has an obligation to conduct a peer review of the technical quality of State assessment systems implemented under section 1111(b)(2) of the ESEA.
- Peer review panels consists of assessment experts in the field of educational measurement and assessment professionals.
- Experts from other states, higher education, and researchers.

Peer Review Critical Elements

- SECTION 1: STATEWIDE SYSTEM OF STANDARDS AND ASSESSMENTS
 - The State formally adopted challenging academic content standards for all students in reading/language arts, mathematics and science and applies its academic content standards to all public schools and public school students in the State.
 - The State’s challenging academic content standards in reading/language arts, mathematics, and science are aligned with entrance requirements for credit-bearing coursework in the system of public higher education in the State and relevant State career and technical education standards



Peer Review Critical Elements

- SECTION 2: ASSESSMENT SYSTEM OPERATIONS: Alignment studies
 - The State’s test design and test development process is well-suited for the content, is technically sound, and aligns the assessments to the depth and breadth of the State’s academic content standards for the grade that is being assessed.
- Alignment studies, conducted by EdMetric and submitted as evidence in most recent peer review confirmed evidence of adequate alignment of North Carolina’s reading and mathematics assessments:
 - Overall, the alignment evaluation found substantial evidence to support alignment of the North Carolina EOG and EOC assessments to the North Carolina content standards in all grade levels. (Egan, Davidson, and Rabinowitz, 2023).



Peer Review Critical Elements

- SECTION 3: TECHNICAL QUALITY – VALIDITY
 - The State has documented adequate overall validity evidence for its assessments consistent with nationally recognized professional and technical testing standards.
 - Documentation of adequate alignment between the State’s assessments and the academic content standards the assessments are designed to measure in terms of content (i.e., knowledge and process), balance of content, and cognitive complexity.
 - The State has documented adequate validity evidence that the scoring and reporting structures of its assessments are consistent with the sub-domain structures of the State’s academic content standards.
 - The State has documented adequate validity evidence that the State’s assessment scores are related as expected with other variables.



Peer Review Critical Elements

- SECTION 4: TECHNICAL QUALITY – OTHER

- The State has documented adequate reliability evidence for its assessments for measures of reliability for the State’s overall student population and **each student group** consistent with nationally recognized professional and technical testing standards.
- The State has taken reasonable and appropriate steps to ensure that its assessments are accessible to all students and fair across student groups in their design, development and analysis.
- The State has ensured that each assessment provides an adequately precise estimate of student performance across the full performance continuum for academic assessments, including performance for high- and low-achieving students.
- If the State administers multiple forms of academic assessments within a content area and grade level, within or across school years, the State ensures that all forms adequately represent the State’s academic content standards and yield consistent score interpretations such that the forms are comparable within and across school years.

Peer Review Critical Elements

- SECTION 5: INCLUSION OF ALL STUDENTS
 - The State has in place procedures to ensure the inclusion of all public elementary and secondary school students with disabilities in the State’s assessment system. Decisions about how to assess students with disabilities must be made by a student’s IEP Team under IDEA, the placement team under Section 504, or the individual or team designated by a district to make that decision under Title II of the ADA, as applicable, based on each student’s individual abilities and needs.
- SECTION 6: ACADEMIC ACHIEVEMENT STANDARDS AND REPORTING
 - The State formally adopted challenging academic achievement standards in reading/language arts, mathematics, and science for all students. (Not Proficient, Level 3, Level 4, Level 5)



Peer Review Outcomes

- Peer review findings released November 2024 for the current edition of reading and mathematics end-of-grade and end-of-course tests were determined to substantially meet requirements of the ESEA.
 - This means that the components of North Carolina’s assessment system meet most of the requirements of the statute and regulations, but some additional information is required.
 - The process includes a plan (due within 30 days) of how to provide additional evidence within one year.

Technical Process

- To ensure the development of technically sound assessments that meet federal requirements and Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education), the NCDPI psychometricians collaborate with
 - North Carolina State University
 - University of North Carolina at Greensboro
 - Nationally recognized test measurement experts (Technical Advisors that meet twice a year to review processes and outcomes)



Test Development Process

- The [test development process](#) begins with the State Board of Education (SBE) adoption of content standards.
- **North Carolina educators** are consulted to specify what can be tested and how from the SBE adopted content standards.
- **North Carolina teachers** are contracted to develop test items aligned to specified grade content standards.
- After extensive review by content experts, students with disabilities experts, and English learner experts, test items are field tested, then reviewed to ensure they satisfy measurement goals.
- Psychometricians use statistical models to construct test forms that meet industry standards of reliability and fairness and produce scores that are valid for uses identified in state and federal statute.



Conclusion: Claims

- The end-of-grade and end-of-course tests, including the alternate assessments, in reading, mathematics, and science
 - Are an accurate estimation of what students know and are able to do
 - Are valid measures of the State Board of Education adopted content standards
- The assessments produce reliable test scores that are technically sound, supporting their use for school improvement planning and school accountability.





NORTH CAROLINA
State Board of Education
Department of Public Instruction