

# RACE TO THE TOP ASSESSMENT PROGRAM TECHNICAL REVIEW PROCESS



U.S. Department of Education  
Washington, DC 20202

April 2013

## RACE TO THE TOP ASSESSMENT TECHNICAL REVIEW

### Table of Contents

- I. INTRODUCTION
  - A. Background
  - B. Purpose
  - C. Role of Reviewers
  - D. Review Process
  - E. How to Use This Document
  
- II. GUIDING QUESTIONS FOR THE TECHNICAL EXPERTS
  - A. Validity Framework
    - 1. The summative assessment measures student knowledge and skills against a common set of college- and career-ready standards (as defined in the NIA) in mathematics and English language arts
    - 2. The assessment system consists of assessment components in mathematics and in English language arts that include, for each subject, one or more summative assessment components that produce student achievement data and student growth data (both as defined in the NIA) that can be used to determine whether individual students are prepared for success, without remediation, in credit-bearing entry-level courses in an institution of higher education (IHE), incorporating feedback from higher education and work force leads
    - 3. The assessment system assesses all students, including English learners and students with disabilities
    - 4. The assessment system produces data, including student achievement data and student growth data
    - 5. Other claims, as determined by the consortium
  - B. Assessment Development
    - 1. Assessment design
    - 2. Scoring and reporting
    - 3. Test security

Appendix A. Race to the Top Assessment Technical Review Process Report Notes

## **I. INTRODUCTION**

### **A. Background**

The American Recovery and Reinvestment Act of 2009 (ARRA) provided funding through the Race to the Top Assessment (RTTA) program to two consortia of states to develop comprehensive assessment systems by the 2014–15 school year that are valid, support and inform instruction, provide accurate information about what students know and can do, and measure student achievement against standards designed to ensure that all students gain the knowledge and skills needed to succeed in college and the workplace. These assessments are intended to provide administrators, educators, parents, and students with the information they need to continuously improve teaching and learning and help meet the President’s goal of restoring, by 2020, the nation’s position as the world leader in college graduates.

The consortia of states committed to developing and implementing an assessment system by the 2014–15 school year that—

1. Measures student knowledge and skills against a common set of college- and career-ready standards (as defined in the notice inviting applications (NIA)) in mathematics and English language arts in a way that—
  - a. Covers the full range of those standards, including standards against which student achievement has traditionally been difficult to measure;
  - b. As appropriate, elicits complex student demonstrations or applications of knowledge and skills;
  - c. Provides an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students; and
  - d. Can be used to provide an accurate measure of student growth over a full academic year or course;
2. Consists of assessment components in mathematics and in English language arts that include, for each subject, one or more summative assessment components that—
  - a. Are administered at least once during the academic year in grades 3 through 8 and at least once in high school; and
  - b. Produce student achievement data and student growth data (both as defined in the NIA) that can be used to determine whether individual students are college- and career-ready (as defined in the NIA) or on track to being college and career ready (as defined in the NIA);
3. Assesses all students, including English learners (as defined in the NIA) and students with disabilities (as defined in the NIA); and
4. Produces data, including student achievement data and student growth data, that can be used to inform—
  - a. Determinations of school effectiveness for purposes of accountability under Title I of the Elementary and Secondary Education Act of 1965, as amended (ESEA);
  - b. Determinations of individual principal and teacher effectiveness for purposes of evaluation;
  - c. Determinations of principal and teacher professional development and support needs; and
  - d. Teaching, learning, and program improvement.

### **B. Purpose**

As the two consortia, the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (Smarter Balanced), are developing their assessment systems, the U.S. Department of Education (Department) will support their work, analyze

their progress, and identify areas for technical assistance. To help accomplish this, the Department will utilize the services of nationally renowned assessment experts to help the Department conduct the RTTA Technical Review Process. These technical experts will meet at least annually during the life of the grant, beginning in early 2013, and will provide their individual evaluation of the grantees' progress developing a comprehensive assessment system that meets the RTTA absolute priority and identify areas where technical assistance may be needed to strengthen the consortia's development.

The Department acknowledges that much of the documentation used to confirm the technical quality of the assessment system, such as analysis of the differential item functioning (DIF) across groups of students or the equivalence of the forms for the operational test, will not be available until after the first operational administration of the assessment system. Similarly the alignment of the operational assessments to the academic content standards and whether the academic achievement standards measure college and career readiness cannot be determined during the development phase, but the consortia can and should be able to document its work during development that can help ensure the consortia's ultimately being successful in this activity. Documentation of the decisions and processes during the development of the assessment system provide important information about the quality of the consortium's assessment system. Both the consortia and the Department have a vested interest in ensuring that a high level of technical quality is being maintained throughout the development process, which will increase the likelihood that that the assessment system will yield valid, reliable, and fair results and meet the goals of the RTTA program.

While there are many facets to ensuring that the assessment system is technically and psychometrically sound, the RTTA Technical Review Process will focus on two broad aspects: (a) the consortium's approach to ensuring that assessment results are valid for the intended purposes and (b) the consortium's approach to and process for developing assessment items and tasks that represent the full range of the college- and career-ready content standards. Section II of this document provides the issues or topics that the Department believes are of critical importance for the development of the assessment systems and which are pertinent to the consortium's current activities. These topics are not intended to be exhaustive of the questions or evidence that may be needed over the course of the development process to ensure the validity and technical quality of a final assessment system. For example, test security is an important consideration for ensuring the ultimate validity of the results; however, the Department understands that in the early stages of the development of the assessment system, the consortia have not established the test administration platform(s), which will inform the test security policies and procedures. As a result, this document may be updated prior to the RTTA Technical Reviews that may be conducted in future years.

The RTTA Technical Review Process is one component of the Department's overall RTTA program review process. The program review process is the overall method by which the Department provides oversight of and support for the consortia. It will be combined with other components, including ongoing, but at least monthly, conversations between the Department and the grantee; on-site program reviews by Department staff; "stocktake" meetings with the consortium and senior leaders in the Department; and the annual performance report. A description of the complete program review process can be found at: [www2.ed.gov/programs/racetothetop-assessment/performance.html](http://www2.ed.gov/programs/racetothetop-assessment/performance.html).

### **C. Role of Reviewers**

Seven nationally renowned technical experts will review each grantee's progress to date developing its assessment system. The technical experts should include psychometricians or experts in assessment design and development, experts in the field of educational testing validity, and experts in the field of how to include disadvantaged populations, such as English learners and students with disabilities, in assessment systems. The technical experts should also include individuals with an

understanding of the consortia's plans and the college- and career-ready content standards in English language arts and mathematics that the states in both consortia adopted as the basis for their assessment systems.

The role of the technical experts is to review evidence provided by the consortia and provide individualized feedback to the Department on whether the consortia are making progress meeting the RTTA absolute priority and designing next-generation assessment systems that conform to best practices for validity, reliability, fairness, alignment, and scoring and reporting, as outlined in the American Psychological Association's (APA) *Standards for Educational and Psychological Testing* (1999). When data are available or studies are complete, the technical experts should identify whether the studies are sufficient to address the question being considered and whether the consortium is responding appropriately to the results from the study. When data are not yet available or studies are not yet undertaken or complete, the technical experts should examine the consortium's approach and process to carrying out its work, including examining the study design and sampling procedures, to determine whether sufficient studies are planned. In this way, the technical experts' feedback to the Department will also serve as technical assistance to the consortium; the technical experts' comments should inform the consortium's on-going work.

The Department requires and expects the technical experts to provide individual feedback and is not seeking consensus advice or recommendations. The experts are each nationally renowned experts and the Department will ask each member to provide their individual feedback to the Department, which will be shared with the consortium.

#### Additional expertise

If the Department determines it is necessary, it may engage additional experts focused on specific components of the consortia's assessment system as part of the RTTA Technical Review Process. For example, the Department may determine that additional external expertise is required to analyze the consortia's progress with developing innovative, technology-enhanced items that measure the full range of the academic content standards and elicit complex student demonstrations or applications of knowledge and skills. The Department may engage additional content experts to examine a stratified sample of items from each consortium and additional supporting evidence.

In such cases, the Department will ask at least one of the technical experts to join the topic-specific experts to review the consortia's evidence. The Department will develop questions for the experts to consider and a template for them to record their analysis. Each will provide individual comments, which will be provided to the Department and shared with the RTTA technical experts as part of their analysis of the consortium's progress. These will also be shared with the consortium.

As is the case for the RTTA technical experts, these topic-specific experts will be asked for their individual analysis. The Department requires and expects the technical experts to provide individual feedback and is not seeking consensus.

#### **D. Review Process**

The technical experts will meet at least annually during the life of the grant, beginning in early 2013, to examine evidence provided by each consortium regarding its progress to date. The RTTA Technical Review Process will provide an opportunity for each consortium to engage in discussions with the technical experts concerning its approach and the evidence it submitted. Following the discussions, the technical experts will provide their individual analyses to the Department regarding the sufficiency of the current activities and additional activities or research each consortium may

consider undertaking. The Department will ask each consortium to submit to the Department a response to the technical experts' evaluations.

The RTTA Technical Review Process will follow this general timeline:

1. The Department, in consultation with PARCC and Smarter Balanced, will convene the technical experts for a two-day, in-person meeting.
2. At least two weeks prior to the review, using this document as its guide, each consortium submits documentation to the Department and the technical experts related to the progress and technical quality of its assessment development.
3. Using the template provided by the Department (see appendix A), the technical experts will independently examine the evidence prior to meeting in person and complete an initial draft of comments or questions to guide the in-person discussion. The technical experts consider the evidence in relation to the RTTA absolute priority, which lays out the intended purposes of the assessment system, and the *APA Standards for Educational and Psychological Testing* (1999) to examine the consortium's progress in creating, reviewing, revising, and validating the components of a high-quality assessment system.
4. The Department facilitates the meeting of the technical experts with the two consortia. The meeting will provide time for the technical experts to meet together to evaluate the consortium's evidence as well as to meet with representatives of each consortium. This meeting is a closed meeting in order to maintain the confidentiality of the consortia's sensitive materials and to provide an opportunity for a full discussion of the materials by the consortia members and Technical Reviewers.
  - **NOTE:** The technical experts are neither asked nor expected to develop a consensus regarding any recommendations or comments.
5. Within two weeks of the conclusion of the meeting, based on the evidence provided by the consortia and the discussion at the in-person meeting, the technical experts each submit to the Department their individual analysis for each consortium. The Department will share these with the consortium.
  - The Department may determine it is necessary to develop a Departmental summary of the technical experts' analyses. If that is necessary, this will be provided to the consortium along with the technical experts' individual analyses.
6. Within three weeks of receiving the report, the consortium submits a response to the Department addressing the technical experts' analysis or, if applicable, the Department's summary of the technical experts' analyses.
  - It is not expected that the consortium would be required to implement all suggestions raised by the technical experts (or the Department's summary, if applicable), but the consortium is expected to respond to the analysis and, where the consortium believes the suggested approach is not necessary or at odds with the consortium's current assessment development and approach, provide a rationale to support that decision.

## **E. How to Use This Document**

The consortium should use the guiding topics presented below in section II as the basis for identifying the evidence to provide to the technical experts. The Department has provided examples of evidence the consortium may provide to respond to specific topics or issues. These examples are not intended to be exhaustive. Where available, the consortium should provide the specific evidence, studies, or documents that demonstrate the consortium's progress. As noted above, the Department does not expect that the consortium has completed all of the work necessary for each component listed below; much of this work is iterative and will continue throughout the life of the grant and beyond. In such

cases, the consortium should include work undertaken to date and information about future work that is planned (including, as applicable) the timeline, approach, and key activities.

Section II of this document should also guide the technical review experts. The technical experts should examine the consortium's evidence in relation to the components of the RTTA absolute priority to determine whether the consortium has presented sufficient evidence or has established an approach that is likely to produce sufficient evidence to demonstrate the consortium is making progress toward meeting the RTTA absolute priority and creating a valid, fair, and reliable assessment system.

## II. GUIDING QUESTIONS FOR THE TECHNICAL REVIEW EXPERTS

### A. Validity Framework

As stated in the *Standards for Educational and Psychological Testing*, validity “refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests.” (p. 9) Consequently, the consortium’s approach to ensuring the validity of its summative assessment system, or its validity framework, is a critical component during and throughout the development of the assessment system. While the Department realizes that much of the validity evidence cannot be collected until after the field test or the first live administration of the summative assessments in the 2014–15 school year, the consortia should establish the validity framework early in the assessment development process to inform its research, development of individual items, and construction of assessments.

For the consortia, the RTTA absolute priority from the NIA (included on page 2 of this document) provides the key intended uses for the assessment system. The consortium should provide its validity framework for ensuring and confirming that its assessment system will meet the requirements laid out in the absolute priority.

In addition, the consortium may identify other uses for or interpretations of the scores. For example, both consortia intend to make specific claims or provide sub-scores about student knowledge or abilities (e.g., a student’s ability to write cogently or a student’s ability to model mathematics). Those uses should be identified by the consortium as part of its validity framework with a plan to support the valid use of the assessment system for that purpose.

The following questions should guide the review of the consortium’s progress toward demonstrating the validity of the proposed interpretation of test scores for each particular purpose:

- How does the consortium address this purpose in its research plan?
- What are the research questions guiding the consortium’s work? Are they sufficient to collect the necessary evidence to support the valid interpretation of test results for this purpose?
- What is the timeline for the data collection and the scope and sampling plan? Does this appear to be sufficient to address the research questions?
- If the research is complete, does the evidence adequately demonstrate the valid interpretation of test results for this purpose?
- What is the consortium’s plan (if the research is not yet complete) or action (if the research is complete) to use the results to improve on-going assessment development activities? What is the consortium’s plan to confirm the changes were sufficient to address any issues identified?
- Would additional research or study improve or inform the consortium’s plan for capturing necessary validity evidence?

The RTTA Technical Review will focus on the consortium’s ability to demonstrate the validity of using the assessment scores for the following purposes:

1. **The summative assessment measures student knowledge and skills against a common set of college- and career-ready standards (as defined in the NIA) in mathematics and English language arts.**
  - a. The assessment system covers the full range of those standards, including standards against which student achievement has traditionally been difficult to measure.

*Examples of evidence could include:*

- Test blueprint showing that the full range of the standards is covered on the assessments.
- Documentation that the specificity of the test blueprint ensures that the data collected provides sufficient information about student mastery of the concepts and skills assessed.
- Plan for demonstrating that, where content standards address more than one concept and skill, the assessment system measures the breadth of the standards such that the scores support inferences about a student's knowledge and abilities.
- Plan for confirming that the items measure the constructs they are intended to measure.
- If employing multiple forms of the assessment within and grade level, documentation that they are comparable in terms of content coverage and difficulty and yield comparable results.
- If employing a computer-adaptive design, documentation that all students have equal probability of exposure to all content standards, particularly content standards measuring cognitively complex knowledge or skills.

- b. The summative assessments provide an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students.

*Examples of evidence could include:*

- Test design and/or computer-adaptive algorithms showing that the full range of cognitive ability and difficulty, as defined by the objectives or targets, is measured for both high- and low-achieving students.
- Plan for evaluating (e.g., simulations, evidence from pilot or field testing) that the assessment system provides an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students.
- Changes made as a result of field trials and pilots to ensure an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students.

- c. The summative assessments provide an accurate measure of student growth over a full academic year or course.

*Examples of evidence could include:*

- Research plan to document that the summative assessments can provide an accurate measure of individual student growth across grades.
- Documentation regarding whether the assessment system supports measuring individual student growth between grade 8 and the high school assessments.

**2. The assessment system consists of assessment components in mathematics and in English language arts that include, for each subject, one or more summative assessment components that produce student achievement data and student growth data (both as defined in the NIA) that can be used to determine whether individual students are prepared for success, without remediation, in credit-bearing entry-level courses in an institution of higher education (IHE), incorporating feedback from higher education and work force leads.**

- a. Students who score proficient or above are college- and career- ready or on track to be college and career ready.

*Examples of evidence could include:*

- Methodology for setting academic achievement standards, research validating that methodology, and plan for validating the achievement standards and cut scores.
- Research plan for confirming the predictive validity of the assessment.

- For the high school assessment, documentation that individuals from IHEs, individuals knowledgeable about career-readiness requirements, and other relevant groups will be involved in the establishment of academic achievement standards.
- b. Test scores are related to outside variables as intended; intended and unintended consequences of the assessment are evaluated.
- Examples of evidence could include:*
- Documentation identifying intended consequences.
  - Data from item trials, cognitive labs, and the pilot or field tests verifying that items and sets of items are performing as intended for the given student population.
  - Data from item trials, cognitive labs, and the pilot or field tests verifying that test and item scores are related to outside variables as intended (e.g., scores are weakly correlated, if at all, with irrelevant characteristics, such as demographics).
  - Research plan for confirming the concurrent validity of the assessment results, such as by examining the relationship between the consortium’s assessments and other well-established assessments (e.g., NAEP or PISA).
  - Plans for using pilot and field test data to suggest a structure for evaluating intended and unintended consequences of operational testing.

**3. The assessment system is fair and produces valid measures for all students, including English learners and students with disabilities.**

- a. The assessment system is fair and accessible and produces valid scores for English learners.

*Examples of evidence could include:*

- Documentation that English learners (at different levels of English proficiency) participate in each stage in assessment development, including item tryouts, cognitive labs or interviews, usability tests, and pilot and field testing.
- Documentation of the conceptual defensibility and the practical feasibility for the accommodations offered for English learners, including the eligibility parameters.
- If applicable, documentation of the process to translate test materials.
- Documentation regarding how, when, and in what capacity people knowledgeable about English learners are included in the assessment development process. If applicable, this includes documenting the qualifications and experience of individuals in charge of translating test materials.
- Data from item trials, cognitive labs, and the pilot and field tests verifying that accessibility features (including technology features) and accommodations for English learners benefit students eligible for the feature and do not benefit other students.
- Data from item trials, cognitive labs, and the pilot and field tests verifying that the computer administration and other technology features do not impact English learners’ ability to demonstrate their knowledge and abilities.
- Data from item trials, cognitive labs, and the pilot and field tests identifying whether bias exists against English learners and the plan to investigate and eliminate that bias.

- b. The assessment system is fair and accessible and produces valid scores for students with disabilities.

*Examples of evidence could include:*

- Documentation that students with disabilities (with specific types of disabilities) participate in each stage in assessment development, including item tryouts, cognitive labs or interviews, usability tests, and pilot and field testing.
- Pilot and field test sampling plan for examining the impact of accessibility features, including the use of computer-based testing, on students with disabilities.

- Documentation regarding how, when, and in what capacity people knowledgeable about students with disabilities are included in the assessment development process.
  - Documentation of the conceptual defensibility and the practical feasibility of each of the accommodations offered for students with disabilities, including the eligibility parameters.
  - Data from item trials, cognitive labs, and the pilot and field tests verifying that accessibility features (including technology features) and accommodations for students with disabilities benefit students eligible for the feature and do not benefit other students.
  - Data from item trials, cognitive labs, and the pilot and field tests verifying that the computer administration and other technology features do not impact the impact of students with disabilities to demonstrate their knowledge and abilities.
  - Data from item trials, cognitive labs, and the pilot and field tests identifying whether bias exists against students with disabilities and the plan to investigate and eliminate that bias.
- 4. The assessment system produces data, including student achievement data and student growth data, that will validly inform—**
- a. Determinations of school effectiveness for purposes of accountability under Title I of the Elementary and Secondary Education Act of 1965, as amended (ESEA).  
*Examples of evidence could include:*
    - Research plan, with the timeline and method of collecting data, to examine whether the summative assessments can be used to determine school effectiveness.
    - Plan for documenting and reporting the standard error of measurement for aggregate scores used to determine school effectiveness.
  - b. Determinations of individual principal and teacher effectiveness for purposes of evaluation and identifying professional development and support needs.  
*Examples of evidence could include:*
    - Research plan, with the timeline and method of collecting data, to examine whether the summative assessments can be used to inform principal and teacher evaluation systems.
    - Research plan, with the timeline and method of collecting data, to confirm that the summative assessments are instructionally sensitive, meaning that students’ performance on the assessments accurately reflects the quality of instruction received.
    - Plan and rationale for how assessment data can and should be used to inform principal and teacher professional development and support needs.
- 5. Other claims, as determined by the consortium. The consortium should identify each additional claim (e.g., sub-scores that the consortium will report) about student achievement that will be made by the consortium.**
- a. For each claim, the consortium should provide the approach to determining the validity of the use of the scores for that purpose.  
*Examples of evidence could include:*
    - Study designs and plan and timeline for collecting data sufficient to evaluate the validity of the claim related to student performance.
    - Documentation indicating how the information will be included in score reports (e.g., sample reports or report templates).

## B. Assessment Design & Development

Assessment design and development informs the validity framework by establishing the consortium's process for developing test items that, when combined on a summative assessment, measure a student's knowledge and skills for a specific grade and subject. Ensuring that the consortium's process is technically and psychometrically sound is critical to meeting the RTTA absolute priority.

This section is focused on questions related to test design to determine whether the approach will yield an assessment system that yields valid and reliable results. For example, how will the consortium ensure that the items are measuring the construct they are intended to measure? Has the consortium developed a test framework or blueprint to measure the full range of the standards? How will the consortium develop items that are free of bias, to the extent possible, and maximize the accessibility of the items to the broadest population of students, particularly students with disabilities and English learners?

### 1. Assessment design

Describe the consortium's progress toward ensuring that the assessments:

- a. Are aligned to the academic content standards in terms of content (what students should know) and process (what students should be able to do).

*Examples of evidence could include:*

- Test blueprints that demonstrate the desired alignment of the assessments to the:
  - i. Description of the objectives or targets to be assessed.
  - ii. Degree of cognitive complexity expected by each objective.
  - iii. Intended coverage of objectives (e.g., the degree of emphasis placed on objectives).
  - iv. Intended item format by objective.
  - v. The number and proportion of items that have been developed, by item type, objective, cognitive complexity, and intended level of difficulty.
- Evidence that the items/forms follow the consortium's established item development approach (e.g., evidence-centered design).

*Examples of evidence could include:*

- i. Task or item models, with scoring rubrics.
  - ii. Item writing training materials.
  - iii. Samples of items for each type of item (selected response, constructed response, technology-enhanced items, and performance tasks) for grades 4, 8, and 11 (e.g., forms from the pilot test or item tryout that include the full range of item types). Information included with the item should be sufficient to evaluate whether it matches the task or item model, the intended content standards being assessed, and the intended degree of cognitive complexity and item difficulty.
  - iv. Process for evaluating the use of task or item models over time to ensure consistency.
  - v. Process for confirming that tasks and items match the approved models.
  - vi. Process for monitoring item writers' fidelity to task or item models.
  - vii. Process for evaluating consistency of reviewers' use of task or item models in reviewing tasks or items.
- Evidence that the assessment includes standards against which student achievement has traditionally been difficult to measure.
    - i. Test blueprint identifying difficult-to-measure standards.
    - ii. Documentation that the blueprint covers higher-order thinking skills consistent with the content standards.

- iii. Sample sets of items and tasks, including rubrics and sample student work where applicable, addressing difficult-to-measure standards.
- b. Elicit complex student demonstrations or applications of knowledge and skills.  
*Examples of evidence could include:*
  - Framework for designing items (e.g., content and item specifications).
  - Data from item trials, cognitive labs, and/or pilot testing validating that items of a higher cognitive complexity elicit the intended student responses.
- c. As much as feasible, are free of bias; all items and assessments are accessible to all students, including students with disabilities and English learners; and valid inferences about students' knowledge and skills can be made.  
*Examples of evidence could include:*
  - Process to maximize accessibility of items during the development process, including how Universal Design for Learning is included in the item development process.
  - Process to write, review, and evaluate items for bias and sensitivity, construct irrelevance, and grade appropriateness during item writing.
  - Guidelines or training materials for item reviewers.
  - Process to evaluate items for bias during pilot or field testing.

## 2. Scoring and reporting

Describe the consortium's progress toward ensuring that:

- a. Test users have clear explanations of and rationales for the meaning, intended interpretation, and use of data, as well as the limitations of specific assessment scores.  
*Examples of evidence could include:*
  - Sample reports or templates of reports showing the format for reporting achievement levels and interpretive, diagnostic, or descriptive information included in the report.
  - Research validating the meaning, intended interpretation, and use of data.
  - Plans and processes for communicating the explanations and rationales to stakeholders (e.g., interpretive guides, release of data sets, sharing of data for research).
- b. Assessment results are expressed in terms of the academic achievement standards, not just scale scores or percentiles.  
*Examples of evidence could include:*
  - Sample reports or templates of reports showing the format for reporting achievement levels.
  - Draft performance or achievement level descriptors.
  - Process for finalizing and formally adopting the performance level descriptors.
  - Plan to confirm that content addressed in the performance level descriptors from grade to grade is vertically moderated.
- c. The standard error of measurement has been determined.  
*Examples of evidence could include:*
  - Documentation regarding how the standard error of measurement will be calculated, reported, and used.
- d. Clear scoring rubrics are developed to evaluate whether the student's response is sufficient to demonstrate the knowledge and abilities tested by the item.

*Examples of evidence could include:*

- Guidelines for developing scoring rubrics.
- Sample item sets with scoring rubrics and sample student work along with data, where applicable.

- e. Procedures are in place for testing, evaluating, and validating the scoring of assessment results, including automated scoring.

*Examples of evidence could include:*

- Consortium's approach to using automated scoring (e.g., what types of items will be automatically scored; whether human raters will also be employed).
- Consortium's approach to evaluating the validity and accuracy of automated scoring, including identifying possible unintended consequences of automated scoring.
- Plan to ensure the inter-rater reliability between human scorers and between automated and human scoring, where applicable.

- f. Scoring and reporting structures are consistent with the sub-domain structures of the test design.

*Examples of evidence could include:*

- Documentation and a plan for the alignment between score reports and the sub-domain structures identified in the test design.

### **3. Test security**

Describe the consortium's progress toward ensuring that:

- a. Provisions are in place for preventing and mitigating the disclosure of test items or test forms, consistent with legal and professional ethics requirements, before and during test administration, particularly for performance tasks or other more memorable types of items, as well as after the test administration.

*Examples of evidence could include:*

- Plans and processes for identifying and mitigating the disclosure of test items or test forms.

- b. Item authoring and test delivery systems have proper safeguards and controls to minimize test security risks.

*Examples of evidence could include:*

- Description of security safeguards and explanation of how these safeguards and controls are being designed, tested, and validated prior to the operational administration.
- Plans and processes for ensuring the integrity of items and data from year to year.

- c. Clear test administration procedures and policies are established to reduce the risk of test security violations or item and test exposure.

*Examples of evidence could include:*

- Test administration policies and manual laying out security procedures before, during, and after test administration.
- Training guides and schedule for all relevant school and district personnel in proper test administration procedures.
- Description of post-administration analysis of test data.

## Race to the Top Assessment Technical Review Process Report Notes

**Consortium:**

**Date:**

**Reviewer:**

### **A. Validity Framework**

The following questions should guide the review of the consortium’s progress toward demonstrating the validity of the proposed interpretation of test scores for each particular purpose outlined in the table below:

- How does the consortium address this purpose in its research plan?
- What are the research questions guiding the consortium’s work? Are they sufficient to collect the necessary evidence to support the valid interpretation of test results for this purpose?
- What is the timeline for the data collection and the scope and sampling plan? Does this appear to be sufficient to address the research question(s)?
- If the research is complete, does the evidence adequately demonstrate the valid interpretation of test results for this purpose?
- What is the consortium’s plan (if the research is not yet complete) or action (if the research is complete) to use the results to improve on-going assessment development activities?
- Would additional research or study improve or inform the consortium’s plan for capturing necessary validity evidence?

**1. The summative assessment measures student knowledge and skills against a common set of college- and career-ready standards (as defined in the notice inviting applications (NIA)) in mathematics and English language arts.**

Measure	Consortium Material Submitted/Description of Planned Approach	Comments, Strengths, Recommendations, or Suggestions Regarding Consortium Materials
a. The assessment system covers the full range of those standards, including standards against which student achievement has traditionally been difficult to measure.		

Race to the Top Assessment program  
 Technical Review Process

b. The summative assessments provide an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students.		
c. The summative assessments provide an accurate measure of student growth over a full academic year or course.		
<u>Summary</u>		

**2. The assessment system consists of assessment components in mathematics and in English language arts that include, for each subject, one or more summative assessment components that produce student achievement data and student growth data (both as defined in the NIA) that can be used to determine whether individual students are prepared for success, without remediation, in credit-bearing entry-level courses in an institution of higher education (IHE), incorporating feedback from higher education and work force leads.**

Measure	Consortium Material Submitted/Description of Planned Approach	Comments, Strengths, Recommendations, or Suggestions Regarding Consortium Materials
a. Students who score proficient or above are college- and career- ready or on track to be college and career ready.		
b. Test scores are related to outside variables as intended; intended and unintended consequences of the assessment are evaluated.		
<u>Summary</u>		

**3. The assessment system is fair and produces valid measures for all students, including English learners and students with disabilities.**

Measure	Consortium Material Submitted/Description of Planned Approach	Comments, Strengths, Recommendations, or Suggestions Regarding Consortium Materials
a. The assessment system is fair and accessible and produces valid scores for English learners.		
b. The assessment system is fair and accessible and produces valid scores for students with disabilities.		
<u>Summary</u>		

**4. The assessment system produces data, including student achievement data and student growth data, that will validly inform the following purposes.**

Measure	Consortium Material Submitted/Description of Planned Approach	Comments, Strengths, Recommendations, or Suggestions Regarding Consortium Materials
a. Determinations of school effectiveness for purposes of accountability under Title I of the Elementary and Secondary Education Act of 1965, as amended (ESEA).		
b. Determinations of individual principal and teacher effectiveness for purposes of evaluation and identifying professional development and support needs.		

Race to the Top Assessment program  
 Technical Review Process

Summary

5. **Other claims, as determined by the consortium. The consortium should list all additional claims about student achievement that will be made by the consortium and the approach to determining the validity of those claims. [Add rows as necessary.]**

Measure	Consortium Material Submitted/Description of Planned Approach	Comments, Strengths, Recommendations, or Suggestions Regarding Consortium Materials
a. [Consortium insert claim here]		
b. [Consortium insert claim here]		
c. [Consortium insert claim here]		
d. [Consortium insert claim here]		
e. [Consortium insert claim here]		
<u>Summary</u>		

**B. Assessment Design & Development**

**1. Assessment design.**

Measure	Consortium Material Submitted/Description of Planned Approach	Comments, Strengths, Recommendations, or Suggestions Regarding Consortium Materials
a. The assessments are aligned to the academic content standards in terms of content (what students should know) and process (what students should be able to do).		
b. The assessments elicit complex student demonstrations or applications of knowledge and skills.		
c. The assessments, as much as feasible, are free of bias; all items and assessments are accessible to all students, including students with disabilities and English learners; and valid inferences about students' knowledge and skills can be made.		
<u>Summary</u>		

**2. Scoring and reporting.**

<b>Measure</b>	<b>Consortium Material Submitted/Description of Planned Approach</b>	<b>Comments, Strengths, Recommendations, or Suggestions Regarding Consortium Materials</b>
a. Test users have clear explanations of and rationales for the meaning, intended interpretation, and use of data, as well as the limitations of specific assessment scores.		
b. Assessment results are expressed in terms of the academic achievement standards, not just scale scores or percentiles.		
c. The standard error of measurement has been determined.		
d. Clear scoring rubrics are developed to evaluate whether the student’s response is sufficient to demonstrate the knowledge and abilities tested by the item.		
e. Procedures are in place for testing, evaluating, and validating the scoring of assessment results, including automated scoring.		
f. Scoring and reporting structures are consistent with the sub-domain structures of the test design.		
<u>Summary</u>		

**3. Test security.**

Measure	Consortium Material Submitted/Description of Planned Approach	Comments, Strengths, Recommendations, or Suggestions Regarding Consortium Materials
a. Provisions are in place for preventing and mitigating the disclosure of test items or test forms, consistent with legal and professional ethics requirements, before and during test administration, particularly for performance tasks or other more memorable types of items, as well as after the test administration.		
b. Item authoring and test delivery systems have proper safeguards and controls to minimize test security risks.		
c. Clear test administration procedures and policies are established to reduce the risk of test security violations or item and test exposure.		
<u>Summary</u>		