

Race to the Top Assessment Program Spring 2013 Technical Review for the Smarter Balanced Assessment Consortium

July 3, 2013

Introduction

In April 2013, the U.S. Department of Education (Department) created and convened the Race to the Top Assessment Technical Review as part of the Department's continuing work to support the two consortia of states developing next generation assessment systems, the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (Smarter Balanced). The review provided the consortia with an opportunity to share technical documentation and items being developed with external experts in assessment design and development, educational testing validity, accessibility, psychometrics, and English language arts and mathematics content.¹

The Department appreciates that the consortia are in the middle of their assessment development. As a result, much of the data and documentation to confirm the technical quality of the assessment system is not yet available, and the consortia will continue to revise and improve their development processes over the coming months. The review focused on two broad areas of assessment development: (1) the consortium's research confirming that assessment results will be valid for their intended uses and (2) the test development process, including reviewing a sample of items and tasks. The experts' individual analyses provided information that both the consortium and the Department can use during this critical period. We expect the Technical Review to help each consortium to identify whether it has sufficient processes in place, has documented its work and development decisions to date, and has sufficient research underway or planned to confirm the technical quality of its assessment system.

The Technical Review is one component of the Department's Race to the Top Assessment Program Review, which is the way we are working with both consortia to support and provide oversight of their work. The Department intends to convene experts periodically for additional review and comments. This review focused on the technical quality of the consortia's assessment development. It did not examine other critical components of developing a large, multi-state assessment system, such as procurement, project management, organizational effectiveness, technology development, and professional capacity, outreach, and engagement. These topics are addressed through other components of the program review; more information about the program review, including an analysis of the consortium's progress during the first two years of the grant, can be found in the annual consortium reports, available at www2.ed.gov/programs/racetothetop-assessment/performance.html.

About this Document

To assist the consortia in considering the reviewers' comments, the Department identified high-level topics that were discussed during the Technical Review and asked each consortium to respond, identifying actions that will strengthen its work. The Department prepared this document for the convenience of the reader based on the individual reviewers' comments and the responses from the consortium. For ease in understanding this document, the Department included a preliminary analysis for each major topic area (assessment development; accessibility and accommodations; and research and planning) based on the reviewers' individual analyses and the consortium's responses:

- Generally on track – the consortium's progress is on track and of high quality, though a few areas may need additional refinement or attention.
- Some aspects on track, other areas need additional focus – the consortium has made progress, though some areas need attention or refinement during development.

¹ The Department requested that the technical experts, each nationally renowned in their fields, provide individual feedback to the Department and the consortium. The Department did not seek consensus advice or recommendations.

- Needs attention – the consortium has made some progress but a number of key areas need attention.
- Needs urgent action – the consortium needs urgent action to address concerns in key areas of development.

The Department will use the reviewers’ analyses to guide its work with each consortium. The Department may identify areas where the consortia need additional assistance or focus as they incorporate Technical Review feedback into their ongoing assessment development. As part of the annual program review process, the Department will provide another opportunity for the Technical Review experts to analyze the continued progress of the consortia and to provide feedback to the Department and the consortia. The Department will also incorporate the Technical Reviewers’ feedback into ongoing program review routines, expanding attention to key areas of the reviewers’ feedback during monthly calls, annual site visits, and targeted support.

Assessment Development

Assessment Development

Highlights of Individual Reviewer Comments

- Many of the reviewers appreciated that Smarter Balanced developed a large number of items (5,000) for the spring 2013 pilot test and has thoughtfully attempted to incorporate lessons learned from this item development into future item development for the field test and operational assessment, creating archetype items to guide future development and bringing in content experts to inform the consortium’s assessment development work.²
- Reviewers appreciated that Smarter Balanced has created English language arts and mathematics content specifications that identify the claims about student achievement to be measured by the assessments and that translate the content standards into assessment targets to aid item development. They recommended that the consortium convene content experts to evaluate the assessment targets to ensure alignment with the college- and career-ready content standards the consortium selected, the Common Core State Standards (CCSS).
- Some reviewers recommended that the consortium study whether the 500 archetype items it is developing to guide and train item writers for the recently awarded contract for field test item development are sufficient to address the depth and breadth of the content standards. In addition, the reviewers encouraged Smarter Balanced to annotate the archetypes so that item writers and reviewers have clear guidance when using them.
- They recommended that, to enhance the quality and consistency of item development, training for item writers and reviewers should be improved and guidance documents should be clear, concise, and located in a central document for ease of access and reference.

Highlights of Smarter Balanced Response

- The contract Smarter Balanced awarded in spring 2013 for field test item writing includes a review of the item writing and review process and the content specifications. Smarter Balanced will continue to make improvements to the item development process based on lessons learned from pilot testing and will utilize the expertise of the item quality review panel members as processes are revised.
- The consortium reported that the content specifications went through extensive review by content experts before the Smarter Balanced Governing States adopted the claims about student achievement in spring 2012. The assessment targets and evidence statements will be updated, as necessary, as lessons are learned during assessment development.

² As part of the recently awarded contract for item writing for the field test, Smarter Balanced created the Item Quality Review Panel consisting of national experts to collaborate with Smarter Balanced and its contractors to establish and maintain the criteria for high-quality items and to evaluate items and performance tasks during the development process.

- The item quality review panel that met in spring 2013 reviewed the content frameworks, including the claims, assessment targets, and alignment to the CCSS, and affirmed the consortium’s direction.
- Smarter Balanced will focus the archetype development on areas the consortium expects to be more challenging for item writers. This approach will emphasize the development of archetypes covering content and depth of knowledge that has historically been difficult to measure and that appropriately reflects the cognitive complexity of the CCSS.
- The consortium is currently developing annotated item development quality manuals to demonstrate to item writers the approaches that will elicit the best evidence of student learning.
- The Smarter Balanced item quality review panel has recommended to the item development contractor ways to improve training materials, including task models, item development criteria, and item specifications.

English Language Arts/Literacy Content

Highlights of Individual Reviewer Comments

- Reviewers commended Smarter Balanced for English language arts performance tasks that are appropriate, authentic, and drawing on textual evidence from multiple sources.
- Some reviewers recommended that Smarter Balanced decrease the number of commissioned texts and increase the number of high-quality, authentic, permissioned texts, and that the complexity of the permissioned texts be appropriate to each grade level.³

Highlights of Smarter Balanced Response

- Smarter Balanced is in the process of improving the design, specifications, and related administration materials and policies for English language arts performance tasks based on the analysis of the pilot test responses and test administration feedback. During the task development process, an emphasis will be placed on the authenticity of tasks that require the use of evidence from multiple sources.
- The spring 2014 field test will include a greater percentage of permissioned and public domain texts for the English language arts selected- and constructed-response items. The field test plan includes 50 percent permissioned, 20 percent public domain, and 30 percent commissioned texts. The item quality review panel recommended that texts be tagged with their text complexity to ensure grade-level appropriateness.
- Smarter Balanced will develop guidance for item development that specifies, for each passage selected for the field test, how to identify details in the text that generates questions worth answering.

Mathematics Content

Highlights of Individual Reviewer Comments

- Reviewers appreciated the range of item formats Smarter Balanced has developed and particularly how technology is used in the items.
- Reviewers recommended that the consortium consider assessing students, in some cases, at the “cluster”⁴ level of the mathematics CCSS so that the assessment system appropriately measures key concepts in the content standards at each grade.

³ Permissioned texts are published works for which a fee has been paid to obtain permission for their use. Commissioned texts are those contracted and written for a specified purpose, such as an assessment. Public domain texts are published works that are freely available to anyone.

⁴ In the Common Core State Standards, the standards are organized by “clusters,” which are groups of related standards (e.g., in grade 4, there are five clusters: operations and algebraic thinking; number and operations in base ten; number and operations, fractions; measurement and data; and geometry).

Highlights of Smarter Balanced Response

- Smarter Balanced will continue to revise and improve the mathematics item specifications for the use of technology by relying on its data review analysis and feedback from its member states.
- The consortium will conduct research during the field test in spring 2014 to ensure that the item selection process for the computer-adaptive test includes the full breadth of the CCSS, including at the cluster level. The item selection process will be revised as necessary based on information learned in the field test.

Preliminary Department Analysis: Some aspects on track, other areas need additional focus.

Next Steps: The Technical Review found that Smarter Balanced should focus greater attention on the clarity of its item development materials to ensure the alignment of items to the college- and career-ready standards. The consortium recently awarded the contract for field test item writing, which will first revise item development materials under the guidance of the item quality review panel. The Technical Reviewers highlighted the need for high-quality, authentic texts; the consortium's plans for item development for the field test increases the use of public domain or permissioned texts.

Accessibility and Accommodations

Highlights of Individual Reviewer Comments

- Some reviewers appreciated that Smarter Balanced incorporated universal design and bias and sensitivity reviews into the item writing process from the beginning and built accessibility features into the assessment delivery system.
- Reviewers recommended that the consortium increase the involvement of experts in assessing English learners and students with disabilities in all aspects of the item and test development process.
- Reviewers recommended that Smarter Balanced add cognitive labs, item tryouts, and other research studies that include increased numbers of students with disabilities and English learners, particularly to evaluate their ability to access the content.

Highlights of Smarter Balanced Response

- Smarter Balanced has included several experts knowledgeable about assessing English learners and students with disabilities in its work, including editing and revising the Accessibility and Accommodations Framework, reviewing and approving the research plan, creating Spanish language glossaries for the pilot test, and as members of the item quality review panel, advisory panels for English learners and students with disabilities, and technical advisory committee (TAC).
- Smarter Balanced conducted cognitive labs and small scale trials before the pilot test in spring 2013. Subgroup analyses will be conducted as part of the pilot and field test data analysis. Additional cognitive labs will be conducted following the field test.

Preliminary Department Analysis: Needs attention.

Next Steps: The Technical Review found that Smarter Balanced needs to focus additional attention on ensuring that the assessment system is accessible for all students, particularly students with disabilities and English learners. Specifically, the consortium should increase and improve training for item developers to focus on accessibility and ensure it carries out sufficient research, including whether students with specific disabilities and English learners of varying levels of English proficiency can access the content of the assessments.

Research and Planning

Highlights of Individual Reviewer Comments

- Reviewers recommended that Smarter Balanced discuss the content coverage of the test blueprints⁵ with its TAC and with a panel of content experts, particularly to determine whether one literary and one informational text is adequate to cover the full range of content as intended by the CCSS.
- Regarding the computer-adaptive test (CAT) algorithm, some reviewers recommended that Smarter Balanced clarify its process for validating that the CAT engine will provide an accurate measure of performance for all students. Reviewers recommended that the algorithm should be reviewed by the Smarter Balanced TAC and by a panel of CAT experts and that the consortium should conduct simulations to evaluate student content mastery by achievement level (i.e., students at the 10th percentile, 20th percentile, etc.).
- Reviewers appreciated that Smarter Balanced has already awarded a contract for the development of the reporting system and that the work has begun.
- Some recommended that Smarter Balanced conduct additional research to validate its plan for reporting scale scores at the claim level and for combining the CAT and performance task portions of the assessments into a single scale score. Reviewers recommended that Smarter Balanced develop mock-ups of the score reports for all intended reporting levels and conduct focus groups to evaluate the reports.
- Reviewers commended Smarter Balanced on its innovative and promising framework for setting preliminary achievement standards, a framework that includes the increased involvement of educators via online components.
- They recommended that the consortium quickly establish the process and structure for standards-setting and initially pilot the process on a small scale. Reviewers recommended that the consortium review the process with its TAC.

Highlights of Smarter Balanced Response

- Smarter Balanced has updated its English language arts blueprints to add additional texts. As drafts of the blueprints are updated, they will be submitted to content and technical experts for their review and comments. The final blueprints will require approval by the governing states.
- The critical features of the item selection algorithm used by the CAT engine were reviewed by the TAC early in the process of development. Future reviews are planned as the algorithm is further defined. Simulations will be conducted of the test design, including the CAT algorithm.
- The consortium reported that evaluating the reliability of overall and claim-level scores will be part of the analyses conducted on the field test data. As the consortium evaluates its approach for combining scores, the TAC will be consulted on the approach.
- Smarter Balanced is actively engaged in designing score reports, including convening focus groups that include teachers, administrators, and parents, to gather input on initial designs.
- Smarter Balanced will continue to refine the design of the preliminary standards setting and conduct a field trial of the approach prior to the August 2014 standards setting event.

Preliminary Department Analysis: Some aspects on track, other areas need additional focus.

Next Steps: The Technical Review emphasized the importance of key components of the Smarter Balanced research plan, such as evaluating the CAT algorithm and the approach for setting achievement standards. Smarter Balanced revised its test blueprints and will continue to evaluate them during the development of the assessment system.

⁵ The Smarter Balanced test blueprints provide information on the English language arts and mathematics content that is to be included on the assessments, the emphasis and balance of content, and item types.