

DATE: September 11, 2014
TO: K-12 Leads
FROM: Smarter Balanced
RE: Validity Overview

Since its inception, and consistent with advice from its Technical Advisory Committee [1], Smarter Balanced has taken a comprehensive approach to meeting the requirements for establishing valid measurement. Consistent with the literature on assessment validation (see Kane, 2013 [2]), Smarter Balanced executed a test development process that makes explicit the inferences (or “Claims”) that test users can make when interpreting student scores on the Smarter Balanced assessments [3]. This involved the development of a validity framework [4] that describes the work needed to evaluate these claims. The key steps in this development process are described in the Test Development section below.

Using Kane’s approach to test validation, Smarter Balanced has organized its work into two phases:

- Phase 1. Test Development
- Phase 2. Evaluation

Test Development Phase:

A valid process of test development requires adherence to best practices [5, 6], making possible the creation of tests that measure what is intended for a specified population. The various activities and analyses in the Test Development Phase can be viewed as establishing “Process Validity.” Evidence that Smarter Balanced has met Process Validity requirements is found in the documentation of key development activities (listed more or less in order of sequence):

- Establishment of a **validity framework** [4] to guide test development and ongoing research,
- Development of a comprehensive set of **content specifications** [3] that casts the Common Core content standards in terms of evidence statements about what students will be expected to know and do,
- Development of **test blueprints** [7] that specify number and types of questions to be presented to students,
- Development of **task models for items and stimuli** [8] that guide the writing and review of individual test items and passages,
- Conducting **content and bias/sensitivity reviews** [9, 10] to ensure that items and passages are aligned to Common Core content, are consistent with evidence statements of the content specifications, and are not biased in favor of or against students from different cultural and demographic backgrounds,
- Design and development of a **test delivery platform** [11], which is the software used to deliver the online assessments to students,
- Conducting **cognitive labs** [12] on item types and accommodations to uncover issues and opportunities with modes of presentation, tools, and other item features,
- Establishment of a comprehensive **accessibility framework and guidelines** [13] to ensure that the assessments are accessible to the widest possible array of students,

- Conducting **small-scale trials** [14] to investigate the use and feasibility of different item types,
- Design and development of **software for computer adaptive test delivery** [15], which is used to select items to be presented to students on the basis of both meeting the test blueprint and selecting items that maximize the accuracy of the student's score,
- Implementation of **large-scale pilot tests** [16] in order to collect data on the initial performance of items and the testing platform software,
- Implementation of **large-scale field tests** [17] to collect data on all items to evaluate their technical adequacy and their placement on a continuous growth scale from grades 3 through 11,
- Analysis of the **alignment** [18] among all components of the assessment design, and ultimately between the Common Core content standards and the tests students actually take, and
- Establishment of **internal validity** [16], or the degree to which the test functions as required, has sufficient reliability, and sufficient ability to measure the intended content and not unintended content. Internal validity was investigated using Pilot Test results to determine whether or not a given content area test (ELA/literacy or mathematics) measured the intended construct and not unintended constructs. Essentially, this is an investigation as to whether or not the test is measuring primarily one construct (i.e., if it is uni-dimensional). As indicated in the attached dimensionality paper, the evidence strongly suggests that the Smarter Balanced ELA/literacy and mathematics test are uni-dimensional. Test reliability will initially be modeled through simulations using the item pool after item review, which is due to be completed December 31, 2014. Operational test reliability will be reported in the technical manual following the first operational administration in spring 2015.

Evaluation Phase:

Once the Smarter Balanced assessments are administered operationally in spring 2015, it will be possible to determine “external validity,” which is the degree to which test results correspond to external indicators (consistent with expectations) [4]. For example, students who perform well on the summative test are expected to perform well in the classroom. These external research studies are listed in the attached Validation Worksheet document [19] (see the checkmark under column F for applicable activities). The information in this table shows the main validity activities established through the Smarter Balanced Validity Framework and the associated sources of evidence, past, present, and future. Because this type of evidence continues to be gathered through the operational administration of the assessments, this table mostly reflects future plans for external validity research.

List of Resource Documents

	Document Title	Link
1	Technical Advisory Committee	https://app.smartsheet.com/b/download/att/fAKbHm-fpdhkDOnJ_Vk_VENQu0cyHtpO2pqu5rfChl
2	Validating the Interpretations and Uses of Test Scores	Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. <i>Journal of Educational Measurement</i> , 50(1), 1-73.
3	Smarter Balanced ELA/Literacy and Mathematics Content Specifications	http://www.smarterbalanced.org/smarter-balanced-assessments/
4	Smarter Balanced Assessment Consortium: Comprehensive Research Agenda	http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/08/Smarter-Balanced-Research-Agenda_Recommendations-2012-12-31.pdf
5	Operational Best Practices for Statewide Large-Scale Assessment Programs - See more at: http://www.ccsso.org/News_and_Events/News_Archive/Operational_Best_Practices_Publication_Released_.html#sthash.q5SYLaa8.dpuf	Council of Chief State School Officers and Association of Test Publishers (2013). <i>Operational best practices for statewide large-scale assessment programs</i> . Washington, D.C.: Council of Chief State School Officers and the Association of Test Publishers.
6	The Standards for Educational and Psychological Testing	American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). <i>Standards for educational and psychological testing</i> . Washington, D.C.: American Educational Research Association.
7	Smarter Balanced Test Blueprints	http://www.smarterbalanced.org/smarter-balanced-assessments/
8	Smarter Balanced Item/Task Specifications	http://www.smarterbalanced.org/smarter-balanced-assessments/

	Document Title	Link
9	Smarter Balanced Item Review Process	https://app.smartsheet.com/b/download/att/aSLpUbp2q5AzUgjA7XgJfj09Me40mWvQFashzgsZwM
10	Smarter Balanced Bias and Sensitivity Review Training Module	https://app.smartsheet.com/b/download/att/67ef11VirjbLNnw3Stc6E58f_nJcivvQ8vGGUxIFy1M
11	Smarter Balanced Practice and Training Tests	http://www.smarterbalanced.org/practice-test/
12	Smarter Balanced Assessment Consortium: Cognitive Laboratories Technical Report	https://app.smartsheet.com/b/download/att/w61W8r8KYy62ct93_xwSf8XLtYWJtcwyONJiP_mbJlc
13	Smarter Balanced Assessment Consortium: Usability, Accessibility, and Accommodations Guidelines	http://www.smarterbalanced.org/wordpress/wp-content/uploads/2013/09/SmarterBalanced_Guidelines_091113.pdf
14	Smarter Balanced Assessment Consortium: Small-Scale Trials Technical Report	https://app.smartsheet.com/b/download/att/Nor5K7acF41b7Nm48RWfsEndVy_iHYWBQxbccgO3sYA
15	Smarter Balanced Adaptive Algorithm	http://www.smarterapp.org/documents/AdaptiveAlgorithm-Preview-v3.pdf
16	Pilot Test Data Analysis: Summary of Results	https://app.smartsheet.com/b/download/att/dXt1tJC9H1jOZ0tp_zDPiIFah-HeysfpD0CulqEqH_w
17	Smarter Balanced Field Test Report	https://app.smartsheet.com/b/download/att/B3G-t0a5Trb3uPcNcCJUgG43mWXLDeymLMfxSf-Owt8
18	Summative Assessment Alignment Study: Smarter Balanced TAC Update	https://app.smartsheet.com/b/download/att/I9Ps40mRLc3IIQTIsU12EG4PRhwfV-sJ0013uiVQhRk
19	Validation Worksheet	https://app.smartsheet.com/b/download/att/ASH12gnLXQbm5VUf6-iYNRk6Dm02min8M0Yf3w4Swws