

Chapter 1 Introduction.....2

 Overview 2

 Technical Report Approach 4

 NCLB Peer Review Guidelines and Established Standards..... 4

 Overview and Background of the Smarter Balanced Theory of Action 6

 Seven Principles of Smarter Balanced Underlying the Theory of Action..... 6

 Purposes for the Smarter Balanced Assessment System 7

 Figure 1. Overview of Smarter Balanced Theory of Action9

 Figure 2. Smarter Balanced Development Timeline.....10

Chapter Overview 10

Acknowledgments 12

 Smarter Balanced Work Groups 12

 Outside Groups and Organizations that Collaborated with the Smarter Balanced Assessment System . 12

 2014 Technical Advisory Committee.12

 Contributors to the Accessibility Accommodations Framework.....13

Other Acknowledgments. 13

References..... 14

Chapter 1 Introduction

Overview

The Smarter Balanced Assessment Consortium’s (Smarter Balanced) vision for a new generation assessment system—one that includes a set of balanced components that can be adapted to meet students’ needs across participating states. This is rooted in the need for valid, reliable, and fair assessments of the deep disciplinary understanding and higher-order thinking skills that are increasingly demanded by a knowledge-based global economy. This vision also is based on the belief that assessment must support ongoing improvements in instruction and promote meaningful learning experiences for students that lead to outcomes valued by all stakeholders. The overarching goal of Smarter Balanced is to ensure that all students leave high school prepared for postsecondary success in college or a career through a planned sequence of educational experiences and opportunities. To meet this goal, with support from institutions of higher education (IHEs) and workplace representatives, the Consortium built on the strong foundation in each participating state to create a high quality, balanced, multistate assessment system based on the Common Core State Standards (CCSS) in English language arts/literacy (ELA/Literacy) and mathematics. The role of the Consortium in this process was to guide the development and implementation of an assessment system that reshapes educational practice in participating states in strategic ways and leads to improved learning outcomes for students. Smarter Balanced provides options for customizable system components while also ensuring comparability of high-stakes summative test results across states. In addition, the Consortium is committed to creating a policy environment that fosters innovation while supporting the development of accountability systems that incentivize the right behaviors for students, teachers, and administrators and avoid ones that run counter to Smarter Balanced goals. The comprehensive assessment system proposed by the Consortium calls for strategic use of a variety of item types and performance events to measure the full range of the Common Core State Standards and to ensure accurate assessment of all students, including students with disabilities, English language learners, and low- and high-performing students. Smarter Balanced implemented a system that contains the following features

- assessing Common Core State Standards based computer adaptive summative and interim assessments that make use of technology-enhanced item types and human-scored performance events,
- interim/benchmark assessments that provide more flexible and in-depth and/or midcourse information about what students know and can do in relation to the Common Core State Standards,
- research-supported, resource-based, instructionally sensitive tools, processes, and practices developed by state educators that can be used formatively at the classroom level to improve teaching and increase learning,
- focused ongoing support to teachers through professional development opportunities and exemplary resource materials linked to the Common Core State Standards,

- online reporting system that enables educators' secure access to key information about student progress toward college- and career-readiness and about specific strengths and limitations in what students know and are able to do at each grade level, and
- cross-state communications to inform Stakeholders about Smarter Balanced activities and to ensure a common focus on the goal of college- and career-readiness for all students.

A key component of college- and career-readiness is the ability to integrate knowledge and skills across multiple content standards. Smarter Balanced addressed this capacity using new item types and performance tasks. Performance assessment emphasize the application of knowledge and skills and incorporate a range of non-selected-response tasks to evidence about students' abilities to solve substantive, meaningful problems. Performance assessments also give students opportunities to demonstrate their ability to find and organize information to solve problems, undertake research, frame and conduct investigations, analyze and synthesize data, and apply learning to novel situations. Smarter Balanced performance tasks involve interaction of students with stimulus materials and/or engagement in a problem solution, ultimately leading to an exhibition of the students' application of knowledge and skills, often in writing. Stimuli include a variety of information forms (e.g., reading & graphics) as well as an assignment or problem situation. As a result, performance tasks are an integral part of the Smarter Balanced test design. Further, performance tasks allow some types of peer-group work and collaboration with the teacher or other students prior to the actual scored assessment.

The innovative and efficient use of technology serves as a central feature of this balanced assessment system. Some central notions concerning technology use are that

1. the Smarter Balanced system uses computer adaptive testing to increase the precision and efficiency of the tests.,
2. the expanded use of technology enables the Consortium's goals of developing innovative and realistic item types that ensure measurement of student achievement across a wide performance continuum and provides efficiencies and enhancements for teacher and administrator professional development and capacity building at the local level.
3. through the use of an interoperable electronic platform and leveraging of cross-state resources, Smarter Balanced delivers assessments and produces both standardized and customizable reports that are cost effective, timely, and useful for a range of audiences in tracking and analyzing the progress toward college- and career-readiness of individual students, student subgroups, classrooms, schools, districts, and states.

In summary, the Smarter Balanced learning and assessment system is grounded in a sound theory of action. This system promotes research-supported classroom practice and incorporates a balanced set of technology-enabled tools, innovative assessments, and state-of-the-art classroom support mechanisms that are intended to work coherently to facilitate teaching and learning. Over time, with a purposeful governing structure and Institutes of Higher Education in participating states, this assessment system will affect the improve teaching and learning consistent with Smarter Balanced's theory of action described below.

Technical Report Approach

The intent of this report is to provide comprehensive and detailed evidence in support of the validity of Smarter Balanced assessment program. Integral to this description is a discussion of validity and the Smarter Balanced test validation process. At the outset, it should be recognized that this process of demonstrating evidence in support of validity is an ongoing process. Validity information is provided here is primarily from the initial Pilot Test and the later Field Test phases. The Field Test reflects the final item statistics of record, reporting scale, and achievement levels for ELA/Literacy and mathematics.

To inform the Consortium, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999; 2014) hereafter referred to as the *Standards*) was used as the foundation for developing the necessary validity evidence. The 2014 version of the *Standards* was published just as this report was in the process of being finalized. Occasionally, this report referenced the 2014 *Standards* as well. Also referenced is the U.S. Department of Education (DOE)'s *Standards and Assessments Peer Review Guidance* (2009b), which stipulated the requirements for assessment programs to receive federal approval under the No Child Left Behind (NCLB) legislation. This information is consistent with the Joint Committee on Standards for Educational Evaluation's (JCSEE) *Program Evaluation Standards* (Yarbrough, Shulha, Hopson, & Caruthers, 2011) and the *Guiding Principles for Evaluators* (American Evaluation Association, 2004) which state "evaluators aspire to construct and provide the best possible information that might bear on the value of whatever is being evaluated" (p. 1). With respect to Smarter Balanced, this information is necessary for understanding the degree to which the Consortium is meeting its goals, and in some cases, what further tasks remain to improve the system as it evolves operationally.

NCLB Peer Review Guidelines and Established Standards

Among the principles underlying the Smarter Balanced theory of action is the adherence "to established professional standards" (Smarter Balanced, 2010, p. 33). In addition to adhering to the AERA et al. (1999; 2014) *Standards*, the Consortium will also meet the requirements of the U.S. Department of Education peer review process for NCLB assessments. Although these requirements were suspended as they undergo revision (Delisle, 2012), they remain important because they reflected the Department's most recent standards for ensuring quality and equity in statewide assessment programs. Validity evidence and the ongoing research agenda should incorporate the guidance provided in the *Standards and Assessments Peer Review Guidance* (U.S. Department of Education, 2009b). There is a great deal of overlap between the AERA et al. (1999; 2014) *Standards* and the U.S. Department of Education's *Peer Review Guidance*. However, the *Guidance* stipulates many important requirements. In particular, to meet these requirements the validity evidence and the ongoing research agenda should include the following

- evidence concerning the purpose of an assessment system and studies that support the validity of using results from the assessment system based on their stated purpose and use,
- strong correlations of test and item scores, with relevant measures of academic achievement and weak correlations with irrelevant characteristics, such as demographics (i.e., convergent and discriminant validity),
- investigations regarding whether the assessments produce the intended consequences;

- documentation of the definitions for cut scores and the rationale and procedures for establishing them,
- evidence concerning the precision of the cut scores and consistency of student classification,
- evidence of sufficient levels of reliability for the overall population and for each targeted subpopulation,
- evidence of content alignment over time through quality control reviews,
- evidence of comprehensive alignment and measurement of the full range of content standards, Depth of Knowledge, and cognitive complexity,
- evidence that the assessment plan and test specifications describe how all content standards are assessed and how the domain is sampled that lead to valid inferences about student performance on the standards, both individually and aggregated,
- scores that reflect the full range of achievement standards,
- documentation that describes how the assessments consist of a *coherent* system across grades and subjects including studies establishing vertical scales, and
- identification of how assessments provide information on the progress of students.

The overlap of this evidence with the AERA et al. (1999; 2014) *Standards* is large, and the anticipated revisions to this guidance will likely retain many of these key features. For example, in the temporary suspension of peer review, the U.S. Department of Education reiterated the following desired characteristics for a high-quality assessment system as

- valid, reliable, and fair for its intended purposes and measures student knowledge and skills against college- and career-ready standards,
- covering the full range of those standards, including standards against which student achievement has traditionally been difficult to measure,
- appropriate, eliciting complex student demonstrations or applications of knowledge and skills,
- providing an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students,
- assessing all students, including English Language Learners and students with disabilities,
- making provisions for alternate assessments based on grade-level academic achievement standards or alternate assessments based on alternate academic achievement standards for students with the most significant cognitive disabilities, consistent with 34 C.F.R. § 200.6(a)(2),
- providing an accurate measure of student growth over a full academic year or course,
- providing student achievement data and student growth data that can be used to determine whether individual students are college- and career-ready or on track to being college- and career-ready,
- producing data, including student achievement data and student growth data, that can be used to inform determinations of school effectiveness for purposes of accountability under Title I, individual principal and teacher effectiveness for purposes of evaluation; principal and teacher professional development and support needs; and program improvement.

These characteristics of high-quality assessment systems were given consideration in the development of the Smarter Balanced Assessment System to ensure that evidence was provided meets these high

standards. The Theory of Action and primary purposes and goals of Smarter Balanced are briefly described below.

Overview and Background of the Smarter Balanced Theory of Action

The Smarter Balanced Assessment Consortium supports the development and implementation of learning and assessment systems to reshape education in participating states in order to improve student outcomes. Through expanded use of technology and targeted professional development, the Consortium's Theory of Action calls for integration of learning and assessment systems, leading to more informed decision-making and higher-quality instruction and ultimately increasing the number of students who are well prepared for college and careers.

The ultimate goal of Smarter Balanced is to ensure that all students leave high school prepared for postsecondary success in college or a career through increased student learning and improved teaching. This approach suggests that enhanced learning will result from high-quality assessments that support ongoing improvements in instruction and learning and which are educative for students, parents, teachers, school administrators, members of the larger public, and policymakers. Meeting this goal will require reform and coordination of many elements across the education system. This goal includes but is not limited to a quality assessment system that strategically "balances" summative, interim, and formative components (Darling-Hammond & Pecheone, 2010). An assessment system is required that provides valid measurement across the full range of common rigorous academic standards, including assessment of deep disciplinary understanding and higher-order thinking skills that are increasingly demanded by a knowledge-based economy, and by the establishment of clear, internationally benchmarked performance expectations. Other elements that are outside the intended Smarter Balanced scope-of-work, but not outside its influence, are comprehensive professional development and valid accountability measures.

Seven Principles of Smarter Balanced Underlying the Theory of Action

The Smarter Balanced assessment is guided by a set of seven principles shared by systems in high-achieving nations and a number of high-achieving states in the U.S.

1. Assessments are grounded in a thoughtful, standards-based curriculum and managed as part of an integrated system of standards, curriculum, assessment, instruction, and teacher development. Curriculum and assessments are organized around a well-defined set of learning progressions along multiple dimensions within subject areas. Formative and interim/benchmark assessments and associated support tools are conceptualized in tandem with summative assessments; all of them are linked to the Common Core State Standards and supported by a unified technology platform.
2. Assessments produce evidence of student performance on challenging tasks that evaluate the Common Core State Standards. Instruction and assessments seek to teach and evaluate knowledge and skills that generalize and can transfer to higher education and multiple work domains. These assessments emphasize deep knowledge of core concepts and ideas within and across the disciplines—along with analysis, synthesis, problem solving, communication, and critical thinking—thereby requiring a focus on complex performances as well as on specific concepts, facts, and skills.

3. Teachers are integrally involved in the development and scoring of assessments. While many assessment components are efficiently scored with computer assistance, teachers must also be involved in the formative and summative assessment systems so that they understand and can teach in a manner that is consistent with the full intent of the standards while becoming more skilled in their own classroom assessment practices.
4. The development and implementation of the assessment system is a state-led effort with a transparent and inclusive governance structure. Assessments are structured to improve teaching and learning. Assessments as, of, and for learning are designed to develop understanding of learning standards, what constitutes high-quality work, to what degree is growth occurring, and what is needed for further student learning.
5. Assessment, reporting, and accountability systems provide useful information on multiple measures that is educative for all Stakeholders. Reporting of assessment results is timely and meaningful—offering specific information about areas of performance so that teachers can follow up with targeted instruction, students can better target their own efforts, and administrators and policymakers can fully understand what students know and can do—in order to guide curriculum and professional development decisions.
6. Design and implementation strategies adhere to established professional standards. The development of an integrated, balanced assessment system is an enormous undertaking, requiring commitment to established quality standards in order for the system to be credible, fair, and technically sound. Smarter Balanced continues to be committed to developing an assessment system that meets critical elements required by US DOE Peer Review, relying heavily on the *Standards* as its core resource for quality design. Other key sources of professional standards that guide Smarter Balanced include a reasoning-from-evidence approach (National Research Council, 2001; Mislevy, Almond, & Lukas, 2004), *An Introduction to the Operational Best Practices for Statewide Large-Scale Assessment Programs* (Association of Test Publishers, Council of Chief State School Officers, 2010), and the American National Standards Institute (ANSI) endorsed *Student Evaluation Standards*, *Program Evaluation Standards*, and *Personnel Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994, 2002, 2008, respectively).

Purposes for the Smarter Balanced Assessment System

The Smarter Balanced purpose statements refer to three categories: (a) summative assessments, (b) interim assessments, and (c) formative assessment resources.

The purposes of the Smarter Balanced summative assessments are to provide valid, reliable, and fair information concerning

- students' ELA/literacy and mathematics achievement with respect to the Common Core State Standards as measured by the ELA/literacy and mathematics summative assessments in grades 3 to 8 and high school,
- whether grade 11 students have sufficient academic proficiency in ELA/literacy and mathematics to be ready to take credit-bearing, transferable college courses after completing their high school coursework,

- measurement of students' status prior to grade 11 to determine whether they have demonstrated sufficient academic proficiency in ELA/literacy and mathematics to be on track for achieving college and/or career readiness,
- measurement of students' annual growth toward college- and career-readiness in grade 11 ELA/literacy and mathematics,
- how instruction can be improved at the classroom, school, district, and state levels,
- students' ELA/literacy and mathematics proficiency for federal accountability purposes and potentially for state and local accountability systems, and
- equitable achievement for all students and subgroups of students in ELA/literacy and mathematics.

The purposes of the Smarter Balanced interim assessments are to provide valid, reliable, and fair information about

- student progress toward mastery of the skills in ELA/literacy and mathematics measured by the summative assessment,
- student performance at the Claim or cluster of Assessment Targets so teachers and administrators can track student progress throughout the year and adjust instruction accordingly,
- individual and group (e.g., school, district) performance at the Claim level in ELA/literacy and mathematics to determine whether teaching and learning are on target,
- teacher-moderated scoring of performance events as a professional development vehicle to enhance teacher capacity to evaluate student work aligned to the standards, and
- student progress toward the mastery of skills measured in ELA/literacy and mathematics across all students and subgroups.

The purposes of the Smarter Balanced formative assessment resources are to provide measurement tools and resources to

- improve teaching and learning,
- provide resources to teachers to help them monitor their students' progress throughout the school year,
- illustrate how teachers and other educators can use assessment data to engage students in monitoring their own learning, and
- help teachers and other educators align instruction, curricula, and assessments,
- assist teachers and other educators in using the summative and interim assessments to improve instruction at the individual and classroom levels,
- offer professional development and resources for how to use assessment information to improve teacher decision-making in the classroom.

The primary rationale of the Smarter Balanced assessments is that these aspects can interact to improve the intended student outcomes (i.e., college- and career-readiness). While there are many ways in which the Smarter Balanced assessment system can be deployed, one possible connection among these assessment components is presented in Figure 1.

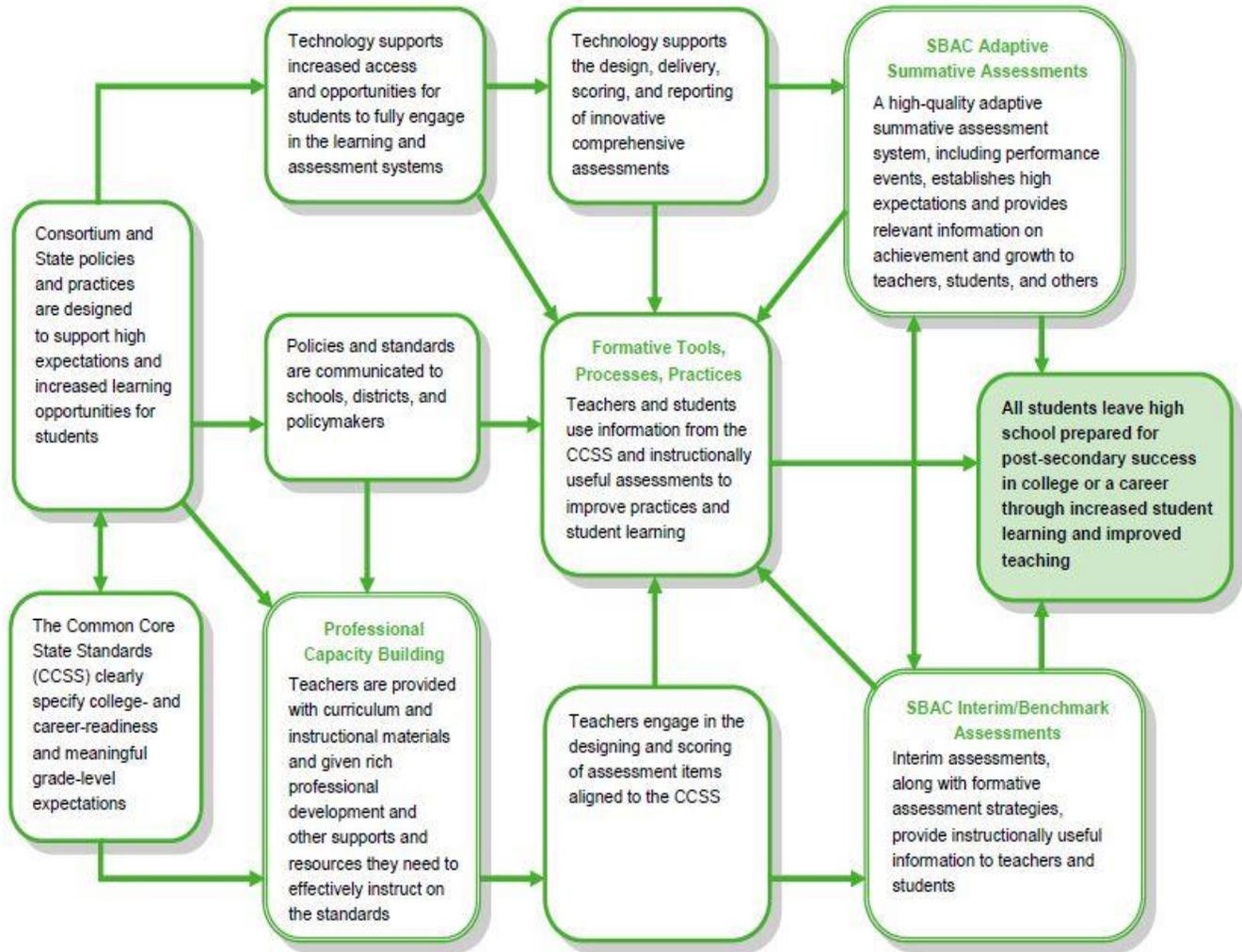


Figure 1. Overview of Smarter Balanced Theory of Action

The Smarter Balanced timeline that references critical aspects of development of the system, such as achievement level descriptors, item development and acceptance, Pilot and Field Test windows, and standard setting events, is given below.

Timeline

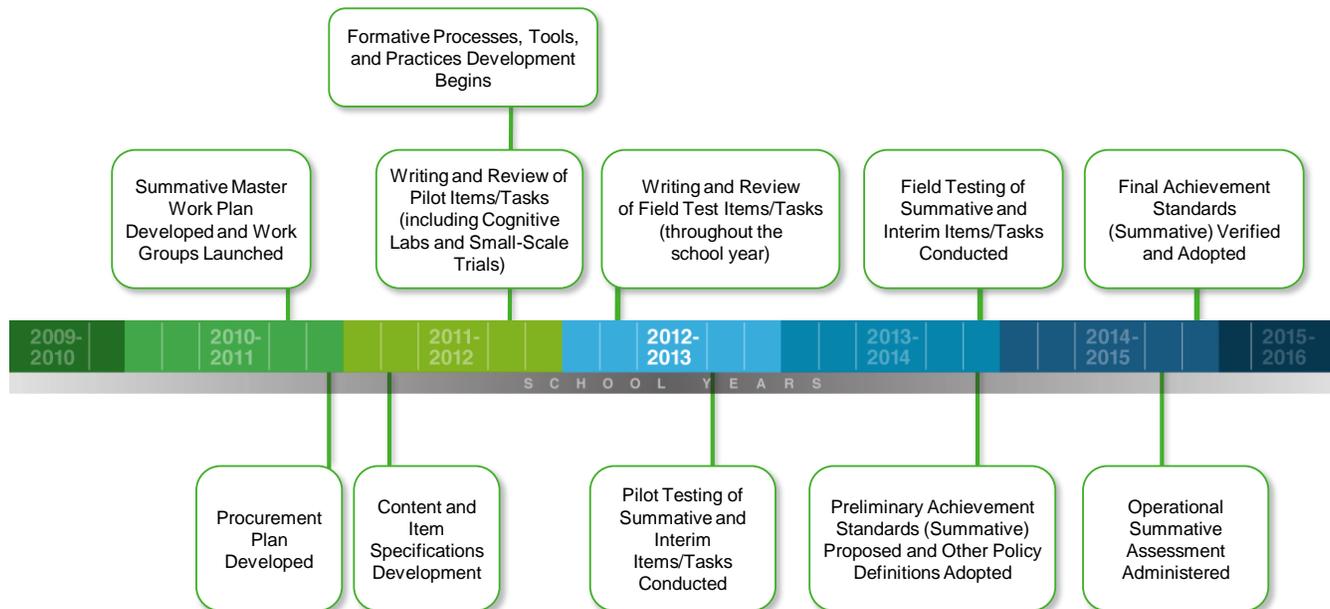


Figure 2. Smarter Balanced Development Timeline

Chapter Overview

Brief synopses of the other chapters contained in the Smarter Balanced Technical Report are given below in order to direct further review.

Chapter 2 Validity Various types of evidence are offered in support of validity in an on-going process of test validation. These sources are validity evidence from the *Standards* are based on (a) test content, (b) response processes, (c) internal structure, (d) relations to other variables, and (e) consequences of testing. As appropriate, some future sources of research and associated evidence are suggested. Since the evidence provided is consistent with a Field Test, essential elements necessary for validity are presented.

Chapter 3 Content Development The process of item and task development is described that presents evidence related to content related validity. A major component for item development was the content alignment studies exploring innovative item and task types writing items intended to measure the Common Core of State Standards. The rationale for and the development process for new types of machine scored and listening items, and classroom-based performance task are described.

Chapter 4 Test Design This chapter provides information pertaining to the content validity of the Smarter Balanced assessment system. Test design is primarily deals with test philosophy and purposes, the intended student population and other technical characteristics. There are two interrelated parts to the test design presented here. More traditionally, the first section pertains to the test blueprints and the entailing specifications such as the use of evidence centered design in Smarter Balanced test development. Secondly, the rationale for performance and CAT assessment components are presented. Summative and interim blueprints are presented here. Central organizing principles for the design and use of interim assessments are also given. Since CAT is an important part of the test design, recommended approaches for evaluating this test delivery approach and the accompanying item pool are presented.

Chapter 5 Test Fairness Test fairness concerns whether score interpretations are valid for all relevant subgroups that minimizes construct irrelevant variance. The evidence for test fairness can consist of logical (e.g., bias review of items) or can be statistical in nature (e.g., differential item functioning, DIF). This chapter primarily presents Smarter Balanced Conceptual Framework for Usability, Accessibility, and Accommodations as well as DIF analysis.

Chapter 6 Pilot Test The Pilot Test design is described, sampling, and the subsequent data analysis. Two outcomes of the Pilot were a dimensionality study for each content area and a study on Item Response Theory (IRT) model choice. The outcomes from these two studies were recommendations for the use of unidimensional models and the two-parameter and generalized partial credit models for IRT scaling.

Chapter 7 Field Test Design Sampling and Administration The Field Test design is described that described the Linear-on-the-Fly test delivery to students that consisted of CAT items and the administration of performance tasks. Two major scaling phases are described that consisted of the IRT vertical scaling and the later horizontal calibration and linking of the much larger item pool. The sampling design is described and the associated decisions. Basic elements of the test administration and security are described.

Chapter 8 Data Step and Classical Test Analysis After the Field Test administration was completed, a data step was performed prior to classical analysis. The data step included decisions concerning the item and student inclusion/exclusion logic and the outcomes from the classical item and test analysis such as item p-values (difficulty).

Chapter 9 Field Test Scaling and Linking Analysis Construction of the vertical scale using the IRT models (i.e., 2PL/GPCM) is presented. Extensive description of the vertical scaling and supporting evidence is given. The outcomes from the second calibration of the item pool are given in terms of its IRT characteristics.

Chapter 10 Standard Setting The chapter describes the procedures used for the Smarter Balanced achievement level setting in the fall of 2014 and the outcomes in terms of the cut scores established.

Acknowledgments

Smarter Balanced Work Groups

The multifaceted nature of this project required that these contracts be organized into larger cross-disciplinary Smarter Balanced work groups that consisted of:

- Accessibility and Accommodations Work Group (A&A)
- Test Administration/Student Access Work Group (TASA)
- Item Development/Performance Tasks Work Group (ID/PT)
- Validation and Psychometrics/Test Design Work Group (V&P/TD)
- Formative Assessment Practices and Professional Learning/Transition to Common Core Work Group (FAPPL/TCC)
- Technology Approach/Reporting Work Group (Tech/RPT)

Outside Groups and Organizations that Collaborated with the Smarter Balanced Assessment System

Below is a partial list of individuals and groups that contributed time and expertise to the consortium.

2014 Technical Advisory Committee.

- Jamal Abedi, Ph.D. *UC Davis/CRESST*
- Randy Bennett, Ph.D. *ETS*
- Derek C. Briggs, Ph.D. *University of Colorado*
- Gregory J. Cizek, Ph.D. *University of North Carolina*
- David T. Conley, Ph.D. *University of Oregon*
- Linda Darling-Hammond, Ph.D. *Stanford University*
- Brian Gong, Ph.D. *The Center for Assessment*
- Edward Haertel, Ph.D. *Stanford University*
- Joan Herman, Ph.D. *UCLA/CRESST*
- G. Gage Kingsbury, Ph.D. *Psychometric Consultant*
- James W. Pellegrino, Ph.D. *University of Illinois, Chicago*
- W. James Popham, Ph.D. *UCLA, Emeritus*
- Joseph Ryan, Ph.D. *Arizona State University*
- Martha Thurlow, Ph.D. *University of Minnesota/NCEO*

Contributors to the Accessibility Accommodations Framework.

In February 2012, the Smarter Balanced Assessment Consortium Accessibility and Accommodations Work Group began work on developing the Accessibility and Accommodations Framework. The primary goal of this effort was to develop uniform accessibility and accommodation policies and guidelines that will be adopted and used by all Smarter Balanced states. Recognizing the diversity in policies and practices that currently exist across states, the legal issues that must be addressed by the policies, the mixed research findings regarding many accommodation practices, and the differences in opinion regarding accommodation policies, the work group undertook an iterative process designed to gather input from a large and diverse audience. This effort began by contracting with Measured Progress and its partners, who included:

- Members of the Measured Progress Innovation Lab who conducted work in accessibility in digital environments, developed the Accessible Test Design model, and were leaders in developing the Accessible Portable Item Protocol (APIP) Standard,
- Experts at Educational Testing Service who have conducted a variety of studies on test accommodations and accessibility for students with disabilities and for students who are English language learners, and who have developed industry-recognized guidelines for accessibility in the context of assessment,
- Experts at the George Washington University Center for Equity and Excellence in Education, who are nationally recognized experts in accessible assessment for students who are English language learners and who have worked with several states to develop policies on test accommodations for students who are English language learners, and
- Experts affiliated with the National Center on Educational Outcomes who have conducted extensive reviews of state-test accommodation policies, worked with the Assessing Special Education Students (ASES) work group of the Council of Chief State School Officers (CCSSO) to develop test accommodation policies, and closely monitored research on test accommodations.

In addition to these partners, an expert panel was formed composed of the following members:

- Jamal Abedi -assessment of English language learners,
- Martha Thurlow -assessment of students with disabilities,
- Sheryl Lazarus -test accommodations for students with disabilities,
- Stephanie Cawthon -accommodations for students who communicate in American Sign Language
- Richard Jackson -accommodations for students with visual impairments,
- Rebecca Kopriva -assessment of students who are English language learners, and
- Stephen Sireci -validity of test accommodations.

Other Acknowledgments.

This technical report leveraged the *Smarter Balanced Comprehensive Research Agenda* by Stephen G. Sireci (2012) as the primary validity framework and sources of evidence. Input was provided on critical aspects of the program and this report by the Smarter Balanced Technical Advisory Committee.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.
- Darling-Hammond, L., & Pecheone, R. (2010). *Developing an Internationally Comparable Balanced Assessment System that Supports High-Quality Learning*. Retrieved from <http://www.k12center.org/publications.html>.
- Delisle, D.S. (2012, December). Letter to Chief State School Officers. Washington, DC: U.S. Department of Education.
- American Evaluation Association (2004). *Guiding Principles for Evaluators*. Washington, DC: Author.
- Mislevy, R.J., Almond, R.G., & Lukas, J. (2004). *A Brief Introduction to Evidence-Centered Design*. (CSE Technical Report 632). Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST).
- National Research Council. (2001). *Educating Children With Autism*. Committee on Educational Interventions for Children with Autism. Division of Behavioral and Social Sciences and Autism. Washington, DC: National Academy Press.
- No Child Left Behind Act of 2001*, 20 U.S.C. 70 6301
- Smarter Balanced Assessment Consortium (2010, June 23). Race to the Top Assessment Program Application for New grants: Comprehensive Assessment Systems, CFDA Number: 84.395B. OMB Control Number 1810-0699.
- U.S. Department of Education (2009a, November). *Race to the Top Program Executive Summary*. Washington, DC: Author.
- U.S. Department of Education (2009b, January). Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001. [Revised December 21, 2007 to include Modified academic achievement standards. Revised with technical edits, January 12, 2009] Washington, DC: Author.
- U.S. Department of Education (2010, September). U.S. secretary of education Duncan announces winners of competition to improve student assessments. Downloaded October 3, 2012 from <http://www.ed.gov/news/press-releases/us-secretary-education-duncan-announces-winners-competition-improve-student-asse>.
- Joint Committee on Standards for Educational Evaluation, American Psychological Association Student Evaluation Standards, Program Evaluation Standards, and Personnel Evaluation Standards (2008).

Sireci, S. G. (2012). Smarter Balanced Assessment Consortium: Comprehensive Research Agenda. Report Prepared for the Smarter Balanced Assessment Consortium.

Standards and Assessments Peer Review Guidance (U.S. Department of Education, 2009b).

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., and Caruthers, F. A. (2011). *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users* (3rd ed.). Thousand Oaks, CA: Sage. 34
C.F.R. § 200.6(a)(2)

An Introduction to the Operational Best Practices for Statewide Large-Scale Assessment Programs (2013). ATP& CCSSO.