Research-Supported Guidance for the Development of a Comprehensive Assessment Framework Based on the Common Core State Standards

Report Submitted to:
SMARTER Balanced Assessment Consortium

September 30, 2010

Submitted by

Introduction

The SMARTER Balanced Assessment Consortium (SBAC, or the Consortium) was awarded funding for the development of a comprehensive assessment system through the U.S. Department of Education's Race to the Top competitive program. One of the requirements for funding is that all SBAC States must adopt the Common Core State Standards (CCSS) before December 2011. To date, 35 states and the District of Columbia have adopted the CCSS; many of these are SBAC members.[1] Each of these states now is in the process of developing a transition plan to support the gradual change from instruction and assessment based on state standards in English language arts (ELA)/literacy and mathematics to instruction and assessment based on a set of common standards in those content areas.

The Consortium seeks to support member States in transitioning to a CCSS-based common summative assessment by spring 2015. As a first step in this process, SBAC contracted with WestEd, the Consortium's Project Management Partner, to write a report about a type of transition or interim document—a framework—that can be used to guide development of SBAC assessments. In this report, WestEd seeks to provide SBAC decision-makers with research-supported recommendations for framework development and a series of guiding questions to support both immediate and more long-term work in "unpacking" the CCSS for assessment purposes. It is hoped that the recommendations in this report will support the SBAC in developing frameworks that have sufficient specificity to effectively guide test design and assessment practices across member States.

Report Organization

- **Section I** provides an introduction to assessment frameworks and reinforces the ways in which they can be used by highlighting exemplars from two key projects: (1) the National Assessment of Educational Progress (NAEP) and (2) the Principled Assessment Designs for Inquiry, or PADI (Mislevy & Haertel, 2006).
- Information about the content on which SBAC assessments will be based, the Common Core State Standards, is presented in **Section II**.
- WestEd refers to the specific document(s) that are intended to emerge as "comprehensive frameworks for assessment," or CFAs.[2] **Section III** describes the core components of a CFA.
- Because tests may serve different purposes, it is likely that distinct CFAs will need to be developed for each of the Consortium's proposed assessments (adaptive summative assessments at each grade, adaptive interim/benchmark assessments linked to learning progressions, and formative tools and processes). The primary

[1] See http://www.corestandards.org/in-the-states for an up-to-date map of state adoptees.

[2] This report provides guidance related specifically to the development of a framework to guide *assessment*. Frameworks also can be developed to guide the development of instructional and curricular materials.

developers of each of these assessment types also are likely to vary (e.g., contractors, panels of state and/or national experts). For these reasons, the Consortium may elect to use a different approach to CFA development for each of the assessments in their proposed system. To support informed decision-making, four defensible approach options are described in **Section IV**.

- **Section V** provides information intended to help SBAC determine an appropriate time line for CFA development.
- **Appendix A** provides a set of key resources for learning more about assessment frameworks.
- **Appendices B and C** contain verbatim text from the Consortium's recent Race to the Top Comprehensive Assessment Systems application that pertains to the purpose of an assessment framework and how assessment frameworks are intended to be used during the development of SBAC's core system components.

I. What is an Assessment Framework?

An assessment framework describes the basic design of a test. It can be a useful tool for defining and clarifying what the standards at each grade mean and in supporting a common understanding of the grade-specific priorities for assessment. When educators and other stakeholders provide input, the framework development process can enable alignment among standards, assessments, and instruction and ensure transparency about what is tested.

In general, an assessment framework describes the specific purpose(s) of the assessment and how results are intended to be used. It spells out the standards (knowledge, skills, and abilities) in a content domain that will be assessed at each grade and describes how student understanding of assessable standards will be measured. Traditionally, assessment frameworks serve as a blueprint for test development by describing the types of items that should be used on a particular test, the target characteristics of each item type (specifications), and the numbers of each item type that need to appear in an item pool or on a test form (blueprint). The framework then serves as a tool for communicating to test developers the intended content emphases and assessment objectives for each grade. Frameworks also may include information to guide test users by prescribing test administration practices and policies (e.g., allowable accommodations, calculator use).

The importance of using a framework to guide assessment development has been described in the Standards for Educational and Psychological Testing (AERA, APA & NCME, 1999). Benefits to framework use are highlighted below.

- **Serves as a "Bridge" Between Standards and Assessments.** According to SBAC's theory of action, the Consortium seeks to ensure "that all students leave high school prepared for postsecondary success in college or a career through increased student learning and improved teaching" (SBAC Race to the Top application, p. 5). This goal requires assessments to be part of a fully integrated system that also includes rigorous standards, aligned curriculum, quality instruction, and well-prepared teachers.

  The Consortium's Race to the Top Comprehensive Assessment Systems application explicitly calls for an interim document that connects the CCSS and effective measurement of those standards to increased postsecondary readiness through a coherent, logical, and transparent argument (see Appendix B). Through an evidence-based approach, development of a comprehensive framework for assessment (CFA) can provide an opportunity to make explicit the Consortium's intent to align all system components—including adaptive summative and interim/benchmark assessments, and formative tools, and professional development for teachers and administrators—and to begin collecting evidence from the outset to support this claim. Ultimately, SBAC States believe that one strategy for improving teaching and increasing student learning is through the implementation of defensible, high-quality assessment practices, and ensuring alignment across standards, assessment,

and instruction is a research-supported step in reaching that goal (National Research Council, 2001).

- **Provides Opportunities for Systematic "Unpacking" of the Standards for Assessment.** Adoption of the CCSS by SBAC States is only the first step in defining the content to be tested by SBAC assessments. The language, structure, and content of the CCSS also must be analyzed so their intent is fully understood by assessment and curriculum developers. Webb (2006) argues,

  > The extent to which valid inferences can be made from a test is closely linked to confidence that the appropriate content is included on the test and that the tasks that are included adequately sample the body of content knowledge from which inferences are made about students. (p. 157)

  Traditionally, framework development requires systematic interpretation of a set of standards; developers must "unpack" the standards to determine what students should know and be able to do at each grade. SBAC stakeholders at a number of levels can be involved in this important process. Depending on the CFA development approach ultimately selected (four options are presented in Section IV), framework development can provide a forum for discussing which content is eligible for assessment on different Consortium measures and how the critical knowledge and skills embedded in the CCSS at each grade can accurately and effectively be measured. For the summative assessments, framework developers must ensure that the full range of the CCSS will be assessed at each grade. Determining the eligible content for the interim/benchmark assessments and formative tools will require developers to make judgments about the clusters of related content as well as the vertical connections or learning progressions across grades embedded in the CCSS.

- **Documents Steps Taken in Reaching Assessment Goals.** Formal and systematic documentation of steps in the assessment development process is essential to sustaining the work of the Consortium and collaboration across SBAC states. Documentation enables continual improvement of the assessment system and eventual change in instructional practice. The CFA process and its outcomes facilitate these goals by ensuring transparency during development and documentation of decisions made.

A Framework Exemplar: The National Assessment of Educational Progress

The National Assessment of Educational Progress (NAEP) assesses samples of students in grades 4, 8, and 12 in all U.S. states and territories using common measures that are standardized so results can be compared over time and across states. The NAEP assessments are based on frameworks that incorporate research-based learning expectations at each grade; these assessments focus on the knowledge and skills that are

believed to be most important at grades 4, 8, and 12 in nine content areas, including reading and mathematics.

The NAEP Frameworks are documents written for general audiences that describe the content and design of subject area assessments (NAGB, 2008). They focus on the measurable indicators of student achievement. A companion document (*Specifications*) provides the detailed blueprint used for constructing an assessment. This technical document guides item development and decision-making about the appropriateness of different types of items and ancillary materials (e.g., passages). While neither of these documents prescribes teaching methods nor a specific curricular approach, together they effectively describe what will be assessed (the "cognitive targets," or mental processes or kinds of thinking underlying achievement in a content domain), how each target will be assessed, and how results will be reported (e.g., via a cross-grade scale or series of subscales).

The developers of the NAEP Technological Literacy Framework (WestEd, 2010) provided the following insight about what an assessment framework can and cannot do:

> The NAEP Framework is an assessment framework, not a curriculum framework. Although the two are clearly related, each has a different purpose and a different set of underlying assumptions. A curriculum framework is designed to inform instruction, to guide what is taught, and often, to guide how it is taught. It represents a very wide universe of learning outcomes from which educators pick and choose what and how they teach. An assessment framework is a subset of the achievement universe from which assessment developers must choose to develop sets of items that can be assessed within time and resource constraints. (p. 15)

While NAEP developers have elected to separate the "what" and the "how" into two separate documents, the CFA described in this report communicates both in one document. It will interpret the CCSS for test developers while also informing educators, test developers, policymakers, parents, and the public what will be measured at each grade by the SBAC assessments (summative, interim/benchmark, and formative measures). These efforts are intended to enable the development of tests that are maximally accessible to all students and that support stakeholder buy-in for Consortium assessment practices.

A Framework Exemplar: Principled Assessment Design for Inquiry, an Evidence-Centered Design Model

Mislevy and Riconscente (2005) describe an effective framework as one that "facilitates communication, coherence, and efficiency in assessment design and task creation" (Downing & Haladyna, 2006, p. 61). In its recent Race to the Top Comprehensive Assessment Systems application, SBAC described its intent to use an evidence-based model

during item development.[3] This model also can be incorporated during the development of the CFA, regardless of the specific approach ultimately adopted (see Section IV for options).

An example of a project in which evidence-centered design methodology was incorporated during framework development is the *Principled Assessment Design for Inquiry*, or PADI (Mislevy & Haertel, 2006; SRI, 2010). PADI is a theory-based approach that combines elements of cognitive psychology, research on science inquiry, contemporary measurement theory, and technology to design science assessments. It was developed to support assessment of concepts and skills that are difficult to assess with traditional measures. The PADI approach to standards-based assessment examines the claims about student capabilities implied by the standards and helps to identify the kinds of evidence needed to justify those claims. These steps require input from researchers, content area experts, and teachers. In this way, central science concepts and how students come to know them can be better understood by test developers as well as classroom instructors.

The evidence-centered design model is resource-intensive and is most effective when implementation is guided by experts with a significant prior knowledge base about how students learn and how different item types can be used to measure levels of expertise in the domains of ELA/literacy and mathematics (Mislevy, Almond, & Lukas, 2004). Yet its use during CFA development offers a number of advantages. Specifically, an evidence-based approach

- reinforces the Consortium's intent to incorporate both learning and measurement theory during test development;
- makes explicit the links between content and test items;
- supports comparability of test forms across students (for the adaptive assessments) and over time; and
- provides particular types of evidence/documentation for building a validity argument, i.e., to support claims that results from SBAC are valid for the purposes intended (Kennedy, 2007; Mislevy, Almond, & Lukas, 2004).

[3] While the scope of this report is limited to framework development, planning for item development is discussed in Section IV as part of each approach option. Also see pp. 77–78 of SBAC's application for Race to the Top funding for details about what the Consortium proposed in relation to item development.

II. Understanding the Content to be Assessed: The Common Core State Standards

SBAC is committed to an assessment system with the CCSS at its foundation. In its Race to the Top application (2010), the Consortium states:

> *The assessment system is aligned to a common set of State standards that clearly specify college, career, and grade-level expectations.* A State policy that is fundamental to SBAC's Theory of Action is adoption of the Common Core State Standards (CCSS), which clearly specify college and career expectations as well as the knowledge and skills required at each grade level to meaningfully articulate progress toward these end-of-high-school expectations. These fewer, higher, and deeper standards—influenced by findings that high-achieving countries typically teach fewer topics more deeply—will serve as the basis for the comprehensive assessment system. And while it is critical that the assessment system validly reflects these standards, SBAC must interpret or translate these standards before they can be used effectively for assessment or instruction. Specific steps include the following: (1) Ensure that each member State adopts the CCSS by December 31, 2011; and (2) Translate the standards into content/curricular frameworks, test maps, and item/performance event specifications to provide assessment specificity and to clarify the connections between instructional processes and assessment outcomes. (p. 34)

To reach these goals, CFA developers need to closely examine the content on which SBAC assessments will be based. To support this initial step in CFA development, an introduction to the CCSS is provided below, followed by more specific information about the sets of standards in English language arts (ELA)/literacy and mathematics.

Goals of the CCSS

In collaboration, the Council of Chief State School Officers (CCSSO) and the National Governors Association (NGA) initiated and oversaw development of the CCSS. Following multiple rounds of development, expert review, public input, refinement, and validation, the standards in ELA/literacy and mathematics that emerged are intended to specify rigorous academic content knowledge and skills so that all students leave high school prepared to attend college or enter the workforce.[4] Specifically, the standards were developed to meet the following criteria (NGA & CCSSO, 2010a; CCSSI, 2010):

- Aligned with expectations for success in postsecondary endeavors in institutions of higher education (IHEs), the workplace, and the military, and for responsible citizenship.
- Clear, so educators and parents know what they need to do to help students learn and so test developers know what is most important in each content area to measure and track for reports on student progress toward college- and career-readiness.
- Consistent across all states, so that targets for learning and assessment are the same in all school communities;

[4] See Appendix A for links to lists of the experts (developer, reviewers, and validators) who contributed to the CCSS in ELA/literacy and mathematics.

- Inclusive of content knowledge as well as application, process, and higher order thinking skills.
- Internationally benchmarked; built upon strengths of standards of top-performing nations and intended to prepare all students for success in a global economy and society.
- Realistic, for effective use in guiding classroom instruction over the course of one school year and in measuring progress at key points in that school year.
- Evidence-based and research–supported.

The CCSS specify the academic expectations, knowledge, and skills that will help students succeed in postsecondary endeavors. It is important to note, however, that the CCSS do not specify *how* these expectations are to be met (NGA & CCSSO, 2010a). Therefore, member States will need strategic help from the Consortium in planning for and designing local systems that support successful transition to instruction and assessment based on the CCSS.

Information about the CCSS in Mathematics to Guide CFA Developers

The CCSS in mathematics for grades K–8 are organized by standards, clusters, and domains. The standards provide the most finely grained level of detail; they define what students should know and be able to do at each grade. Clusters are groups of related standards. Domains are larger groups of related standards that progress across grades. The CCSS for high school are organized by conceptual categories. Table 1 shows the K–8 domains and high school conceptual categories.

Table 1. Common Core State Standards in Mathematics, by Domain or Conceptual Category

	Domain (K–8) or Conceptual Category (HS)						
K	Counting and Cardinality	Operations and Algebraic Thinking	Numbers and Operations in Base Ten		Measurement and Data	Geometry	
1		Operations and Algebraic Thinking	Numbers and Operations in Base Ten		Measurement and Data	Geometry	
2		Operations and Algebraic Thinking	Numbers and Operations in Base Ten		Measurement and Data	Geometry	
3		Operations and Algebraic Thinking	Numbers and Operations in Base Ten	Numbers and Operations-Fractions	Measurement and Data	Geometry	
4		Operations and Algebraic	Numbers and Operations in Base Ten	Numbers and Operations-Fractions	Measurement and Data	Geometry	

	Domain (K–8) or Conceptual Category (HS)						
		Thinking					
5		Operations and Algebraic Thinking	Numbers and Operations in Base Ten	Numbers and Operations-Fractions	Measurement and Data	Geometry	
6		Expressions and Equations	Number System	Ratios and Proportional Relationships		Geometry	Statistics and Probability
7		Expressions and Equations	Number System	Ratios and Proportional Relationships		Geometry	Statistics and Probability
8		Expressions and Equations	Number System	Functions		Geometry	Statistics and Probability
HS		Algebra	Number and Quantity	Functions	Modeling	Geometry	Statistics and Probability

In the CCSS, each grade includes an overview of cross-cutting themes and critical areas of study. A set of Standards for Mathematical Practice is distinct from the K–12 content standards but connects with expectations in each grade (Achieve, 2010d).[5]

A recent presentation by noted mathematician Dr. Jere Confrey (2010) described a number of features of the CCSS in this content area:

- Focus attention on core concepts in numeration and their relationship to operations, especially as it relates to the structure of the number system.
- Develop place value in a coordinated and informed way across grades.
- Strongly link decimal reasoning to fractions.
- Introduce multiple measurement systems (e.g., metric and English) simultaneously.
- Directly link the number line to scales to improve students' visualization of number relationships.
- Support the articulation of key learning trajectories in numeration and geometry.

The NAEP Mathematics Framework served as a key resource during the development of the CCSS. It is not surprising, then, that a recent study by Achieve (2010b, p. 3) found much overlap between the CCSS in mathematics and the NAEP Mathematics Framework. Findings from this study suggest that the two documents describe similar learning expectations of comparable rigor for students at the end of grade 4, 8, and 12. However, since NAEP Frameworks are designed to measure cumulative learning at only three

[5] The *Standards for Mathematical Practice* focus on proficiency (problem solving, reasoning, modeling, mathematical decision making, and engagement), in addition to mathematical "habits of mind" such as making sense of problems, perseverance in solving problems, reasoning qualitatively, and making strategic use of tools.

particular points in time, grade-specific expectations and learning progressions are clearer in the CCSS.[6]

Comments about the nature and sequencing of content in the CCSS in mathematics, by grade span, are provided below (Achieve, 2010c; Confrey, 2010).

- *Grades K–5*: The CCSS for these grades provide a solid foundation in whole numbers, addition, subtraction, multiplication, division, fractions, and decimals.
- *Grades 6–8*: The CCSS for the middle grades describe important learning targets in algebra and geometry, per recommendations from high-performing nations. Variability, distributions, data use, statistical reasoning, and probability are introduced. The time line for teaching certain concepts (e.g., basic algebraic proficiency), however, is more aggressive in the late elementary and early middle grades and provides for more explicit study of geometric relationships.
- *High School*: The high school CCSS require students to practice applying mathematical ways of thinking to real world issues and in novel contexts and use mathematical modeling to analyze empirical situations. The CCSS are much less explicit about the preferred sequence and organization for high school mathematics content. It may prove challenging for districts to determine how best to organize their high school courses to align with the content embodied in the CCSS.

Confrey (2010) cited limitations of the CCSS in mathematics that include the following:

- Mathematics practices are isolated; they are presented in a separate document.
- Integration with development of technology skills and concepts is limited.
- Overreliance on additive structures limits early and foundational development of multiplicative structures related to ratio and rate and many algebra patterns of growth.
- Grain size of standards varies widely across grades, which necessitates careful unpacking.
- In places, language is unclear, so framework developers will need to confirm intent with CCSS developers and/or document rationale for agreed-upon interpretation of meaning.
- More focused on college than career.
- Framework developers will need to identify and/or verify the underlying learning progressions in mathematics as they determine priorities for testing at each grade. Appendix A of the CCSS in mathematics may be useful for this purpose.

Information about the CCSS in ELA/Literacy to Guide CFA Developers

Like the mathematics standards, the CCSS in ELA/literacy are grouped hierarchically into grade-specific standards, college- and career-readiness standards (CCRs, or anchor standards), and strands. At the most finely grained level of detail are the grade-level standards, which define end-of-year learning expectations in grades K–8 and high school.

[6] Similarly, a recent study found that the CCSS also are strongly aligned to the National Council of Teachers of Mathematics' *Focal Points* (Achieve, Inc., 2010a) and to recommendations from the National Mathematics Advisory Panel (Achieve, Inc., 2010c).

Related grade-level standards are grouped into CCRs. These overarching statements describe broad expectations that are consistent across grades. Related CCRs are organized into four strands: reading, writing, speaking and listening, and language.

Details about each content strand are provided below:

- **Reading**: Students are expected to show progressive development in comprehension by reading from a variety of texts of increasing complexity. Foundational skills as well as a range of skills associated with reading literature and reading informational text are addressed. Exemplar texts, selected using qualitative and quantitative measures and consideration of the interaction between reader and text, are provided at each grade to guide decision-making by assessment and curriculum developers (Achieve, Inc., 2010c.) According to the CCSS developers, the standards for this strand are aligned to the NAEP Reading Framework (NGA & CCSSO, 2010a).

- **Writing**: Students are expected to compose arguments, express opinions, and develop informative and/or explanatory compositions and narrative texts. In doing so, the focus is on developing a coherent, reasoned argument through presentation of evidence, conducting research, incorporating technology, and demonstrating progress toward college and career readiness in their work. Annotated student writing samples are included in an appendix. According to the CCSS developers, the standards for this strand are aligned to the NAEP Writing Framework (NGA & CCSSO, 2010a).

- **Speaking/Listening**: Students are expected to speak and listen in a variety of settings, both formal and informal, including whole-class discussions, small-group sessions, and one-on-one conversations. Effective communication practices are emphasized, with active engagement by both speaker (clear articulation of message and attending to feedback) and listener (interpreting and analyzing the message).

- **Language**: Students are expected to understand and apply conventions of standard English in writing and speaking effectively and to show progressive development in vocabulary acquisition through the mix of conversation, direct instruction, and reading called for in the standards. A focus of this strand is on using general academic and domain-specific language in written work and in discussions (Achieve, Inc., 2010c).

While **Media** and **Technology** are not designated as distinct strands in the CCSS, content associated with each is expected to be integrated with daily instruction through the four strands listed above.

Because the CCSS high school standards are organized into two grade spans (9–10 and 11–12), states and/or consortia will need to determine how best to organize the ELA/literacy standards into courses that build on one another in a sequence that leads to

college- and career-readiness (Achieve, 2010d). Care will need to be taken to coordinate the various strands (reading, writing, speaking and listening, and language) within each course, with appropriate weighting of each. It may prove challenging for districts to reorganize their high school courses to ensure full opportunity to learn the content embodied in the CCSS by the end of grades 10 and 12. These considerations have real implications for CFA developers.

The CCSS in this content area incorporate the traditional ELA content embedded in the four core strands above as well as the content from two sets of literacy standards: the **Standards for Literacy in History/Social Studies** and the **Standards for Literacy in Science and Technical Subjects**. Each set of literacy standards includes both reading and writing standards. The *reading* standards call for students to develop knowledge of domain-specific vocabulary; analyze, evaluate, and differentiate primary and secondary sources; and synthesize quantitative and technical information. The *writing* standards require students to use domain-specific vocabulary; write arguments on discipline-specific content and informative texts; and use data, evidence, and reasoning skills to support arguments and claims.

To promote literacy across the curriculum, these standards are intended to be an element of instruction in social studies and science classrooms. Yet it is likely that member States will be challenged to determine how best to meaningfully teach and assess these literacy standards. The Consortium can support States during CFA development by clearly identifying the content from the literacy standards that will be measured by each of the SBAC assessments at each grade. CFA developers may find it helpful to monitor emerging recommendations from research and best practice communities about assessing literacy in history/social studies, science, and technology.

As with mathematics, developers of assessment frameworks in ELA/literacy also will need to identify and/or verify the underlying learning progressions as they determine priorities for testing at each grade. CFA developers may find the curriculum maps in ELA/literacy being developed by Common Core, Inc. (2010) to be helpful for this purpose (see Appendix A).[7]

Augmentation of the CCSS: Implications for CFA Developers

Each state who adopts the CCSS in ELA/literacy and mathematics may elect to supplement the CCSS with additional standards it believes are of critical importance and/or are strongly aligned with state priorities for assessment and instruction.[8] The state then will need to determine if and how to assess those augmented standards on state tests. Similarly, member States will need to reach agreement on this issue as a Consortium: will Consortium

[7] This is a Bill & Melinda Gate-funded initiative that is distinct from the Common Core State Standards Initiative.

[8] While a final ruling on the meaning of this clause is expected from USED, it has been interpreted by at least one state (CO) as meaning that the set of standards adopted must meet or exceed 85% agreement with the CCSS and by others (AZ, DE) as meaning that the state can add to or contribute to the CCSS up to a 15% maximum.

assessments be based solely on the content of the CCSS in ELA/literacy and mathematics as they currently exist or will these assessments measure CCSS content *plus* content from states, national organizations, or initiatives (e.g., STEM [Science, Technology, Engineering, and Mathematics])?

To support informed decision-making about CCSS augmentation, Consortium leaders likely will need a formal communiqué from USED to ensure that all expectations for CCSS adoption are clearly understood by State members and consistently interpreted during CFA development. The Consortium also may elect to reach out to the other federally funded consortium to discuss collaboratively the potential impact of augmentation and weigh tradeoffs prior to final decision-making. Cross-consortium discussions held at that time may support subsequent CFA development by reinforcing priorities for instruction and assessment, ensuring that potential implications of the decision to augment (or not augment) are fully evaluated, and bringing all advisory resources to the table for a common purpose. Undertaking these efforts prior to decision-making may help ensure that SBAC moves forward with plans for development of CCSS-based assessments that fully support the Consortium's goal of preparing all students to leave high school college- and career-ready.

III. Essential Components of a Comprehensive Framework for Assessment

Regardless of the approach for framework development ultimately adopted (see Section IV for descriptions of approach options), the Consortium will want its CFAs to include a number of key components to ensure their usefulness as tools for guiding the development of SBAC tests.[9] These include (1) detailed information about the content to be assessed; (2) the measurement theory of action underlying the assessment; (3) a description of the item types that will be used; (4) specifications for each item type; (5) a blueprint that describes the overall design for the item pool or test; (6) a plan and schedule for item development; and (7) documentation of decisions made, including options considered and rationales for choices. Each of these components is described in greater detail below.

(1) Making expectations for testing clear to stakeholders requires that CFA developers first specify the **content** in ways that precisely identify which standards will be assessed and how proficiency will be defined relative to that content. During CFA development, SBAC will need to define **learning expectations** for students at each grade based on the full range of the CCSS.[10]

Guiding questions to consider during this process include the following:

- According to the CCSS, what are the indicators of college- and career-readiness at each grade?
- Are these also the Consortium's priorities for assessment at each grade? What are the grade-specific learning outcomes it is seeking? Has the Consortium adequately defined the content and constructs that are the targets for assessment?
- How will the Consortium know if students have mastered the instructional priorities it identified? How will levels of achievement for each standard be differentiated?
- Which CCSS will be assessed on the SBAC summative assessments? On the interim/benchmark assessments? Via formative tools and processes?
- What steps will be taken to ensure that the Consortium has addressed the full range (depth and breadth) of the CCSS intended at each grade?

(2) CFA developers then will want to identify a **measurement theory of action** that describes how expertise or competence in a content domain (e.g., ELA/literacy) is measured.[11] Generally, test developers seek to measure students' level of mastery of

[9] Because the nature of these components will vary across content areas, separate CFAs for ELA/literacy and mathematics will be necessary, though they may share common elements (e.g., item types).

[10] A number of resources are included in Appendix A to support the Consortium with understanding college- and career-readiness expectations. For example, see the NGA Issue Brief on setting college and career-ready goals (NGA Center for Best Practice, 2010) and Conley (2010a, 2010b). A report on findings from a CCSS-focused validity study conducted by Conley and his colleagues at EPIC is expected in late 2010.

[11] An example is provided in Confrey (2010, slide 21).

a particular characteristic (e.g., reading comprehension). The measurement theory of action holds that students have different levels of this characteristic and, when measured appropriately, their "scores" on this characteristic will fall along a continuum of least to most. This theory of action is closely linked to the content to be assessed *and* to the specific item types (see Item Types below) used to measure the characteristics of interest, i.e., that can be used to elicit meaningful information (responses) from students about the precise location of their level of expertise (score) for that characteristic. According to the National Research Council (2001), demonstrating understanding of a theory of action underlying item and test development is an important piece of evidence to support the validity of results that emerge from these assessments.

Guiding questions to support informed decision-making include the following:

- How will each intended outcome defined above be measured? How will the Consortium confirm that its measures address all aspects of the intended outcome?
- What is the theoretical foundation or theory of action for SBAC's measurement plan? How is it linked to expertise in this domain?
- How will the Consortium test the accuracy and usefulness of its measurement theory of action for the purpose intended? What evidence should be collected?

(3) CFA developers also will need to determine which **item types** (e.g., open ended, selected response, short answer, extended answer, technology-supported constructed response, essay, presentation, demonstration, production, performance task or event) are developmentally appropriate for students in tested grades and most effective for measuring the content in the domain, including those standards that have been challenging to measure via traditional item types.[12] Using an evidence-based approach (e.g., the PADI example in Section I), each item type will be explicitly linked to a particular standard or group of standards and will fit the proposed measurement theory of action. Clear links among the target content, the measurement theory of action, and the item types must be evident.

The item types and responses to them may be expressed orally or in writing; responses can range from making a check mark to selecting a response choice, completing a project, or performing a task. Item templates (also called item shells or prototypes) will be created for each of the item types deemed appropriate for SBAC

[12] Unlike the types of items that teachers develop spontaneously and use formatively during the course of instruction, the types of items used for this purpose are planned carefully in advance so they meet specific characteristics for quality and consistency (see Specifications).

assessments.[13] Assumptions associated with each template will be tested informally through cognitive interviews with students or more formally through small-scale pilot administrations. Strategies for maximizing accessibility for special student populations (including strategies for differentiating responses) should be considered.

Guiding questions to support decision-making include the following:

- What item characteristics are needed to effectively measure the intended content of each SBAC assessment (summative, interim/benchmark, and formative)?[14]
- What item characteristics are needed to ensure that the access needs of all students, including students with disabilities and English learner students, are considered?
- Which item types most strongly demonstrate those characteristics identified as most important? What evidence supports claims that these item types are appropriate for a specific assessment purpose?
- What evidence links each item type with the measurement theory of action?
- What evidence links each item type with a CCSS standard or cluster of standards?
- Who will design the templates for each item type?

(4) **Specifications** for each item type help test developers by bringing consistency and quality assurance to the ways in which items will be presented, formatted, and used on each assessment. They describe the important criteria or dimensions (e.g., complexity, length, number of response options) for each item type that are needed to effectively measure different standards or groups of standards. Item-type specifications include guidelines for selection of associated stimuli (e.g., types and level of complexity of passages, graphics, or other support materials) and for the administration and scoring of each item type. Reliance on item-type specifications helps to ensure that all subsequent decisions reached during item development by teachers, expert panels, and/or consortium leaders in diverse locations are guided by the same set of pre-established guidelines. Specifications also may include allowable strategies for developing item types that are maximally accessible to special student populations.

Guiding questions for CFA developers to consider when writing specifications include the following:

- What are the presentation and/or formatting characteristics that should be prescribed for each item type?
- What assessment practices (e.g., calculator use) will be allowed?

[13] According to Darling-Hammond (2010, p. 34), "Use of carefully designed and tested task shells or templates can support the creation of tasks that are comparable across versions of the assessment."

[14] See Haladyna (2004), Chapter 3, for a full discussion of the strengths of different item types for different purposes.

- What types of stimuli (e.g., passages) will be used?
- How have the access needs of special student populations been considered?

(5) Framework developers then must create a **blueprint** that describes the overall design for each test or item pool. It prescribes the numbers of each item type needed, the balance of representation (e.g., percentages of items for each standard or groups of standards), and levels of cognitive demand (e.g., Webb's [2006] depth of knowledge levels of Recall, Basic Application of a Skill/Concept, Strategic Thinking, and Extended Thinking) to be assessed. Test or item pool breadth, depth, and length/size are defined. In keeping with the measurement theory of action, this combination of item types is expected to provide a comprehensive and coherent picture of what students know and can do in relation to the characteristic of interest.

Guiding questions to consider during this process include the following:

- How many of each item type/template will be needed for each assessment or item pool at each grade?
- How will the items be distributed across the CCSS at each grade?
- How will different item templates be combined on SBAC measures to address the full range of the CCSS at each grade? Have we addressed the full range (depth and breadth) of the CCSS intended at each grade?
- How can we ensure a balance between test length (for full coverage of the CCSS and to ensure sufficient reliability) and burden to students and schools?

(6) Finally, CFA developers will need to propose a **plan and schedule for item development** that will yield sufficient numbers of the right types of items with the right specifications to meet blueprint needs.[15] If the Consortium decides to collect items from member States, a consolidated item pool can serve as the starting point for development, and an inventory of the items in this pool can be used to set development targets for each grade in ELA/literacy and mathematics. These targets should provide for item overdevelopment, as nearly half of the new items may not be useable after review and refinement. SBAC then will need to identify the team(s) (e.g., a combination of teachers, content and measurement specialists, representatives from IHEs and the workplace, and contractors) that will be responsible for this work as well as a concrete plan and time line for getting the work done. The plan should describe how an evidence-based approach will be used to ensure that each item is strongly linked to the content intended to be assessed through the measurement theory of action. The schedule should allow sufficient time for training, multiple rounds of review by diverse audiences, cognitive

[15] The scope of this paper is limited to framework development. See pages 77–78 of the SBAC Application for Race to the Top funding for details about what was proposed with regard to item development.

interviews with students, small pilot tests, field testing, and post-field-test performance review.

Guiding questions to consider during this process include the following:

- How will the Consortium ensure development of sufficient numbers of different item templates to match blueprint needs? How many additional items of each type will need to be developed at each grade?
- Who will develop the items necessary to meet the target number at each grade? How will they be recruited?
- For each item template, to what general and specific quality standards should all developers be held? Will developers be expected to meet quotas?

(7) Framework developers are responsible for **documentation** of all decisions made during all steps in this process. As appropriate, rationales may be provided to support the defensibility of decisions related to the high-stakes assessments and to reinforce the Consortium's commitment to transparency in testing.

Guiding questions to consider during this process include the following:

- How will decisions about assessable standards, measurement theory of action, and item types be documented?
- Who will review and approve all specifications and blueprints? How will changes be recorded?
- How will the item development process be documented?
- How will decisions based on reviews, interviews, pilot testing, and field testing be documented and their impact monitored?

IV. Building the Comprehensive Framework for Assessment: Four Approaches

A number of different approaches may be used to develop a CFA.[16] Each option is associated with specific strengths and limitations, especially in relation to stakeholder investment, time required for completion, and cost. The Consortium may decide on three different approaches for developing the CFAs for SBAC's three core components (i.e., summative assessments, interim/benchmark assessments, and formative tools). The Consortium also may have compelling reasons for adopting different approaches for developing the CFAs in the two content areas (i.e., ELA/literacy and mathematics). In general, however, it may be more efficient and less confusing to stakeholders if the same approach is adopted for the development of frameworks in both ELA/literacy and mathematics for a specific assessment (e.g., the summative assessment).

Four approaches to CFA development are discussed in this section: (A) an expert panel approach, with oversight from a CFA steering committee; (B) a contractor approach; (C) a committee approach involving State representatives; and (D) a cross-consortia team approach with state representatives from SBAC and the Partnership for the Assessment of Readiness for College and Career (PARCC). Though all four share common elements, each is unique in particular ways.

Overarching Guiding Questions

The following questions are intended to guide Consortium leaders as they consider the various options for CFA development presented in this section:

- Of the options described below, which approach best fits with the Consortium's theory of action?
    - Who is most qualified to interpret and "unpack" the CCSS?
    - Who will weigh tradeoffs and make final decisions (e.g., finding the right balance between test length, optimal coverage of standards, and burden to students and schools)? Who will "own" the final document?
    - How will input from IHEs, professional organizations, teachers, researchers, curriculum designers, assessment experts, and students be considered and/or incorporated?
    - How can contractors be most effectively used to alleviate burden to states?
    - Does one option better support consensus building across states on the intent of the standards?

- Of the options described below, which best meets the Consortium's timeline for development?
    - Is one option more cost-effective than another?

[16] The Consortium also may find that a strategic combination of elements from different approaches best meets its needs.

- Do any require time lines for completion that make them unrealistic for meeting Consortium needs?
- Is sufficient time allowed for multiple rounds of review?

- How will the Consortium ensure that the framework development process is transparent?
  - How will decisions be documented and member states kept informed about the status of the framework(s)?
  - What process will be used to update the framework(s) when needed?

Approach A:
Expert Panels Develop the CFA, with Oversight from a Steering Committee

In this approach, a Consortium-appointed steering committee delegates assignments to small expert panels that have been targeted for having unique expertise to complete specific tasks at different stages of framework development. The effectiveness of this approach hinges on identifying the right sets of experts to work together to complete the work of framework development, with oversight from a CFA steering committee.

***The CFA Steering Committee*.** At the foundation of this proposed approach is an 8–12 person CFA steering committee that delegates development work to smaller teams of experts ("mini-panels"). The CFA steering committee recruits experts for the mini-panels, ensures that each mini-panel has the resources and support needed to complete a specific task or phase of work, synthesizes all contributions into a coherent framework, and ensures that all steps in the development process are documented.[17] Roles are negotiable; e.g., the group may decide that two members work together to oversee one expert mini-panel. Members of the CFA steering committee are State representatives who are elementary and secondary teachers in ELA/literacy or mathematics; state- or district-level content experts in ELA/ literacy, mathematics, or curriculum and instruction specialists; measurement or assessment specialists; representatives from IHEs and the workplace; and/or experts in the assessment needs of special student populations (e.g., students with disabilities or English learner students). Each is identified through nominations from States and selected by the SBAC Executive Committee. To the extent possible, the CFA steering committee is representative of all governing States.

The Expert Mini-Panels. The CFA steering committee recruits experts for the mini-panels. A reasonable plan would be to have one expert mini-panel focused on each of the key components of the framework described in Section III: (1) content to be assessed (CCSS); (2) measurement theory of action; (3) item types; (4) item type specifications; (5) item pool/test blueprint; and (6) item development plan.[18] Each task-focused mini-panel, comprised of 4–6

[17] The resources in this report may be useful for this purpose, including the list of CCSS developers, reviewers, and validators (NGA & CCSSO, 2009b).

[18] Each mini-panel is responsible for documenting steps taken and decisions made during completion of its work.

nationally recognized experts in a particular area, conducts its own review of research literature and examines best practices in the field in relation to the group's focus. The mini-panel then completes all work on that component. Each mini-panel is responsible for building on the work completed and/or under development by other mini-panels, documenting all issues raised, steps taken, and decisions made, and supporting review processes by the CFA steering committee and/or broader SBAC audiences. To the extent possible, the mini-panels should be representative of governing States.

***Steps in Drafting the CFA*.** The proposed steps in this process are outlined below.

(1) Each mini-panel submits its work to the CFA steering committee for review, revision, or refinement (as needed) and approval. Via an iterative process, the first draft of the CFA then is a synthesis of all steering committee- approved work from the mini-panels.
(2) The CFA steering committee submits this first draft to the SBAC Executive Committee for review. Appropriate time (2 weeks) is allocated to allow for a thorough review.
(3) A period of review by the governing States (3–4 weeks) then follows.
(4) All feedback is directed to the CFA steering committee to incorporate, as appropriate.
(5) A period for public review can be incorporated, if desired, to allow for input from a broader variety of fields and content areas. This is best accomplished through an online public comment forum.

The Final CFA. Once all feedback from reviews is incorporated, the CFA steering committee submits the final document(s) to the Executive Committee for review and approval. Once approved, the Executive Committee posts the final document(s) on the SBAC website and call for a State vote on formal adoption.

Strengths, Limitations, and Potential Tradeoffs. This approach has the clear advantage of tapping the particular expertise of those who have been targeted as having unique training and experience to complete specific tasks at different stages of CFA development (e.g., unpacking the CCSS or identifying item types). Additionally, to the extent possible, the mini-panels would work concurrently so as to avoid a bottle-neck effect in which the work of one group holds up all of the others. This approach also allows States to be significantly involved on the CFA steering committee or "somewhat" involved by naming representatives to the mini-panels or monitoring feedback and comments during the review period(s). One disadvantage of this model is that it will require a significant time commitment from the CFA steering committee, but this time may be well spent if the Consortium reaps the benefits of having the right experts doing the right work. With strong recruitment strategies in place, SBAC may find that expert mini-panelists agree to donate their time to this important work, as each will be involved in completing tasks associated with only one CFA component (e.g., developing a blueprint). This could result in an overall cost savings over other approaches.

Option: Use of One Expert Panel Instead of Multiple Mini-Panels. A variation of this model is to follow a more traditional plan in which one expert panel is recruited by Consortium leaders to develop all components of the CFA. While this model requires less work by the CFA steering committee in synthesizing emerging work from the mini-panels, the Consortium runs the risk of not having the right experts in place for specific tasks (e.g., identifying the measurement theory of action). A bottle neck effect also could emerge if one panel is responsible for all components and falls behind in its schedule, thereby costing the Consortium time and money.

Approach B:
A Contractor Develops the CFA

In this model, a vendor selected though a competitive process completes all work associated with CFA development. The effectiveness of this approach hinges on (1) identifying the best contractor (s) to complete the work of CFA development and (2) monitoring deliverable development.

Identifying the Contractor. A subcommittee of SBAC leaders is named to manage all tasks associated with CFA development. This CFA subcommittee, in collaboration with WestEd (as the Project Management Partner) and OSPI (as fiscal agent), develops an RFP to solicit responses from vendors who are qualified to complete the work of CFA development. The RFP can be crafted to ensure that the contractor uses a specific approach (e.g., the expert panel approach), if desired. Each response received through this competitive bid process is screened by WestEd for completeness. All complete proposals then are forwarded to the sub-committee for review. Subcommittee members, with support from WestEd, develop a rubric to support consistent critique of proposals across reviewers and ask each governing State to nominate two representatives to review proposals. State representatives use the rubric to rate each proposal and provide the names of the top two candidates to the subcommittee. The subcommittee, with support from OSPI, reviews the recommendations and selects one candidate for presentation to the Executive Committee. If the first candidate is approved, OSPI names the winning contractor. If this candidate is not approved, the subcommittee provides the Executive Committee with the name of the second candidate. This process continues until OSPI formally names the winning contractor. The winning contractor is announced to member States, and SBAC and OSPI enter into negotiations with the winner.

Steps in the Development Process. Proposed steps are as follows:

(1) Contractor development of the CFA begins, with oversight from WestEd (project management), OSPI, and the CFA subcommittee (substantive issues). WestEd provides guidance to ensure that CFA development proceeds in the manner expected.

(2) The contractor submits a first draft to the subcommittee according to an agreed-upon schedule. The subcommittee reviews the draft to ensure it is ready for review by States.
(3) The draft is posted electronically for State review. With support from WestEd, States are asked to solicit feedback and comments on the draft CFA from target populations such as teachers; district-level curriculum/instruction specialists and measurement/assessment specialists; representatives from state IHEs and the workplace; and experts in the assessment needs of special student populations (e.g., students with disabilities or English learner students). A reasonable time period (e.g., four weeks) should be allowed for this review process.
(4) WestEd collects feedback and comments from States and forwards it in a consolidated format to the subcommittee.
(5) The subcommittee reviews all feedback and directs the contractor to make approved edits.
(6) The contractor submits a revised draft to the subcommittee.
(7) The subcommittee sends the revised draft to identified experts for vetting, reviews all feedback, and advises the contractor on edits, as appropriate.
(8) A period for public review can be incorporated at this point to allow for input from a broader variety of fields and content areas, if desired. This is best accomplished through an online public comment forum.

***The Final CFA*.** The contractor submits the final document(s) to the Executive Committee for review and approval. The contractor is expected to make any changes to the CFA requested by SBAC leaders. Once the deliverable is approved, the Executive Committee posts the final document(s) on the SBAC website and calls for a State vote on formal adoption.

Strengths, Limitations, and Potential Tradeoffs. An advantage to this approach is that the bulk of the work is completed by a contractor, with oversight from an SBAC subcommittee and WestEd; this approach would help alleviate concerns about burden to States. In addition, many contractors have experience developing test specifications and blueprints for states, and they can bring this expertise to the CFA development process. If the contractor also is responsive to SBAC needs, this can be a truly cost-effective option. A number of disadvantages are clear: (1) few contractors will bring expertise in development of key CFS components such as knowledge about the CCSS, measurement theories of action, or innovative item types; (2) time for expert vetting must be built into the process to ensure trustworthiness of the deliverables; (3) acquiring the services of the most qualified contractor may be costly; and (4) the contractor may have more "ownership" for the final CFAs than SBAC States as this approach requires the least input from state representatives during decision-making. Finally, to make this approach work, the SBAC subcommittee and WestEd will need to commit sufficient time and resources to guide and monitor the contractor's work at all stages.

Approach C:
A Committee of State Representatives Develops the CFA

The committee approach borrows a number of elements from the expert panel approach but taps state-level representatives to serve on a committee that completes all work associated with CFA development. The effectiveness of this approach hinges on identifying a team of State representatives who can work together effectively and have the expertise and time to develop all components of the CFA described in Section III: (1) content to be assessed (CCSS); (2) measurement theory of action; (3) item types; (4) item type specifications; (5) item pool/test blueprint; (6) item development plan; and (7) documentation of steps taken and decisions made during the development process. Proposed steps in this process are described below.

***The Nomination Process*.** The Executive Committee asks each governing State to identify candidates who are most qualified to support CFA development and to submit the names of three qualified candidates for consideration for appointment to the CFA development committee. State nominees may be elementary or secondary teachers in ELA/literacy or mathematics; state- or district-level content experts in ELA/literacy or mathematics; curriculum and instruction specialists; measurement or assessment specialists; representatives from IHEs and the workplace; or experts in the assessment needs of special student populations (e.g., students with disabilities or English learner students). The Executive Committee reviews all nominees and selects a team of 12–15 representatives to serve on the CFA development committee. To the extent possible, the committee should be representative of governing States. Once CFA development committee members have been announced, they begin the work of developing the CFA.

***CFA Draft Development and Review*.** The proposed steps in this process are described below.

(1) CFA development committee members attend a kick-off meeting. The kick-off meeting sets the tone for this approach, which is designed to allow for cross-state sharing of ideas, discussion of priorities, and consensus building.
(2) CFA development committee members conduct a review of research and best practices associated with development of each component in the CFA.[19]
(3) The CFA committee works on one component at a time, beginning with unpacking the content to be assessed (CCSS) and moving through articulation of the measurement theory of action, identification of item types, development of item specifications and blueprints, and formulation of an item development plan. As a group, they discuss all tasks and work collaboratively to make and document decisions about all components.
(4) Component by component, a first draft is developed.

[19] The resources in this report may be useful for this purpose, including the names of CCSS developers whom the Consortium may want to interview for clarification when interpreting the standards (NGA & CCSSO, 2009b).

(5) CFA committee members review the completed first draft as a group and edit, as needed.
(6) The CFA committee submits the first draft to the SBAC Executive Committee for review. Appropriate time (2 weeks) is allocated to allow for a thorough review.
(7) Feedback from the Executive Committee is discussed by the CFA development committee and incorporated, as appropriate.
(8) A period of review (3–4 weeks) by the governing States then follows.
(9) All feedback is directed to the CFA committee to incorporate, as appropriate.
(10) The CFA committee sends the revised draft to nationally known content area and assessment experts for vetting. The CFA committee reviews all feedback and incorporates edits, as appropriate.
(11) A period for public review can be incorporated, if desired, to allow for input from a broader variety of fields and content areas. This is best accomplished through an online public comment forum.

The Final CFA. Once all feedback from reviews is incorporated, the CFA committee submits the final document(s) to the Executive Committee for review and approval. Once approved, the Executive Committee posts the final document(s) on the SBAC website and call for a State vote on formal adoption.

Strengths, Limitations, and Potential Tradeoffs. This approach offers two key advantages that are linked to the SBAC theory of action: (1) States are actively involved in decision-making about what will be assessed and how students will be asked to show what they know and can do during testing and (2) States are working together collaboratively to complete all phases of CFA development. This enables grassroots support from States for emerging Consortium assessments and reinforces SBAC's mission of seeking to build the professional capacity of its member States. This strength also is its greatest limitation, as a significant commitment of time and effort will be needed from all CFA committee members. If the Consortium elects this approach, they may want to consider providing CFA committee members with a stipend to show appreciation for their important work. Though this approach will not be as costly as the contractor approach, it is likely to be the most time-consuming, as one of the known challenges of consensus building is completing work in a timely fashion. This will be compounded by the need for expert vetting of the committee-developed CFA to ensure usefulness and trustworthiness for the purposes intended.

Approach D:
A Cross-Consortium Team Uses a Common Approach to Developing the CFAs for Different Summative Assessments

In general, this approach is similar to Approach C (the committee approach), but the committee in this case includes members of both federally funded consortia (SBAC and the

Partnership for the Assessment of Readiness for College and Career, or PARCC). The key challenge for the team in this approach is that members must support one another in developing CFAs for different assessments, though both are based on the same content (CCSS). The ideal outcome from this process would be CFAs that differ only in terms of the item pool/test blueprint. The effectiveness of this approach hinges on team members' expertise and their ability to support one another in developing CFAs for different assessments using a common approach. This team of state-level, cross-consortium representatives must have the collective experience and training to work together effectively to develop all components of a CFA as described in Section III: (1) content to be assessed (CCSS); (2) measurement theory of action; (3) item types; (4) item type specifications; (5) item pool/test blueprint; (6) item development plan; and (7) documentation of steps taken and decisions made during CFA development. Proposed steps in this process are described below.

***The Nomination Process*.** The governing bodies from each consortium ask governing states to identify candidates who are most qualified to support CFA development and to submit the names of two qualified candidates for consideration for appointment to the CFA development team. Nominees may be elementary or secondary teachers in ELA/literacy or mathematics; state- or district-level content experts in ELA/literacy or mathematics; curriculum and instruction specialists; measurement or assessment specialists; representatives from IHEs and the workplace; or experts in the assessment needs of special student populations (e.g., students with disabilities or English learner students). The governing bodies for each consortium independently review all nominees from their respective governing states. From these nominees, each governing body selects the most qualified ten candidates with varied areas of expertise or experience to serve as CFA developers. At the end of this process, 20 highly qualified representatives (ten per consortium) are named to serve on the CFA development team. To the extent possible, the team should be representative of governing states across both consortia. Once team members have been announced, they begin the work of developing the CFA.

***CFA Draft Development and Review*.** The proposed steps in this process are described below.

(1) Team members attend a kick-off meeting. The kick-off meeting provides an opportunity for cross-consortium sharing of ideas, discussion of priorities, and consensus building in relation to the intent of the CCSS.[20] Roles and responsibilities are clarified so that all group members are prepared to work together productively on all activities, with the exception of the item pool/test blueprints. Decisions about this component only are made in a closed-door session among members of a single consortium.

[20] A significant period of time should be allowed for this discussion. It is the primary rationale for this approach.

(2) Team members review existing research and best practices associated with each component of a CFA (e.g., CCSS, measurement theories of action, item types, etc.).[21]
(3) Team members work together strategically on one component at a time, beginning with unpacking the content to be assessed (CCSS) and moving through articulation of the measurement theory of action, identification of item types, development of item specifications and blueprints, and formulation of an item development plan. As a group, they discuss the task at hand and work collaboratively to make and document decisions as they arise. As necessary, given that the CFAs will be guiding the construction of different assessments, decision-making may include only members of one consortium.
(4) Component by component, a first draft is developed for each assessment.
(5) Team members meet to review the completed first drafts for each assessment as a group and make edits, as needed.
(6) The team submits the first draft for each assessment to each consortium's respective governing bodies for review. Appropriate time (3–4 weeks) is allocated to allow for a thorough review.
(7) Feedback from the respective governing bodies is discussed by the team and incorporated, as appropriate.
(8) A period of review (3–4 weeks) by governing states in both consortia then follows.
(9) All feedback is directed to the team to incorporate in the two draft CFAs, as appropriate.
(10) The team sends the revised draft CFAs to selected nationally known content and assessment experts for vetting. As a group, the team reviews all feedback and incorporates edits into the individual CFAs, as appropriate.
(11) A period for public review can be incorporated, if desired, to allow for input from a broader variety of fields and content areas. This is best accomplished through an online public comment forum.

The Final CFAs. Once all feedback from reviews is incorporated, the cross-consortium team submits the final document(s) to each consortium's respective governing bodies for review and approval. Once approved, the governing bodies post the final document(s) on consortium websites and call for a State vote on formal adoption.

Strengths and Limitations. This approach offers the key advantage of actively involving governing states from both consortia in decision-making about how the CCSS will be assessed at each grade. This enables grassroots support from states for emerging consortium assessments and fosters cross-consortia consistency in interpretation of the intent of the CCSS and in framework components (e.g., item types). Team members will face a significant commitment of time and effort, but it should be viewed as a privilege to serve on this team and representatives should be compensated. Though this approach will incur

[21] The resources in this report may be useful for this purpose, including the names of CCSS developers whom the Consortium may want to interview for clarification when interpreting the standards (NGA & CCSSO, 2009b).

costs associated with travel to meetings, it likely will not be as costly as the contractor approach. However, it could become time-consuming if discussions are not productive and work is not completed in a timely fashion. This will be compounded by the need for expert vetting of the two team-developed CFAs to ensure usefulness and trustworthiness for the purposes intended.

V. Building the Comprehensive Framework for Assessment: Considerations in Setting a Time Line for Development

Like most states, the Consortium needs a transition plan to help it prepare for CCSS-based common assessments in 2014–15. SBAC's transition plan will include a few critical first steps to address immediately, many short-term tasks, and a prioritized list of long-term commitments. With the support of WestEd (as Project Management Partner), a timetable for completing this work already is being developed. CFA development is among the Consortium's early priorities.

It will take time to plan for the development of the CFAs that will guide Consortium testing. SBAC leaders will need to review proposal commitments, monitor States' adoption of the CCSS, and implement immediate strategies for communicating with States, teachers, parents, and other stakeholders. They will need to weigh the relative strengths and limitations of the different approaches and select the approach that best fits Consortium values and needs.[22] Once the approach is selected, SBAC leaders will need to sort out immediate priorities from more long-term needs, using specific questions to guide planning:

- Should the Consortium develop both content areas simultaneously or one at a time?
- Should the Consortium develop by grade span (e.g., K–2, 3–5, 6–8, and HS) within each content area or work on all grades at once?
- Should the Consortium initially develop only the CFAs for the summative assessments? What are the potential consequences of delaying development of frameworks for the interim/benchmark assessments and formative tools, given that they must be closely tied to yet-to-emerge curriculum and instructional strategies? How can lessons learned during CFA development for the summative assessments be applied to development of frameworks for other assessments?

Then, depending on the approach selected, the Consortium will need to recruit experts, develop RFPs and select vendors and/or state representatives to serve as committee members. For these reasons, the Consortium will want to plan for 8–10 weeks for preliminary decision-making and preparation for CFA development. The actual work of developing the CFAs for the summative assessments in ELA/literacy and mathematics alone then will take another 12–24 months, depending on the approach adopted.

[22] As previously stated, the Consortium may elect to adopt a different approach for each content area after weighing the tradeoffs associated with that decision. It also may decide that a strategic combination of elements from different approaches best meets its needs.

References

Achieve, Inc. (2010a). *Comparing the Common Core State Standards in mathematics and NCTM's Curriculum Focal Points.* Washington, DC: Author. Retrieved from http://www.achieve.org/CCSSandFocalPoints

Achieve, Inc. (2010b). *Comparing the Common Core State Standards in mathematics and the NAEP Framework.* Washington, DC: Author. Retrieved from http://www.achieve.org/files/CCSSMathandNAEP.pdf

Achieve, Inc. (2010c). *On the road to implementation: Achieving the promise of the Common Core State Standards*. Washington, DC: Author. Retrieved from http://www.achieve.org/files/FINAL-CCSSImplementationGuide.pdf

Achieve, Inc. (2010d). *Understanding the Common Core State Standards.* Washington, DC: Author. Retrieved from http://www.achieve.org/files/CCSSJune22010FINAL.ppt

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: APA.

Common Core, Inc. (2010). The *Common Core curriculum mapping project*. Retrieved from http://www.commoncore.org/maps/

Common Core State Standards Initiative. (2010). *Messaging toolkit: Common Core State Standards Initiative.* Retrieved from http://www.fldoe.org/board/meetings/2010_06_15/toolkit.pdf

Confrey, J. (2010). *What now? Implications, obligations and opportunities for curriculum after the release and adoption of the Common Core State Standards.* Presentation at the Center for the Study of Mathematics Curriculum's "Curriculum Design, Development and Implementation in an Era of the Common Core State Standards" Conference. Arlington, VA. Retrieved from http://www.mathcurriculumcenter.org/conferences/ccss/JConfrey.htm

Conley, D. (2010a). *Creating college and career readiness.* Presentation at Beaverton School District on March 4, 2010. Eugene, OR: EPIC.

Conley, D. (2010b). *Beyond business as usual—Key state actions to boost college and career readiness.* Eugene, OR: EPIC.

Darling-Hammond, L. & Pecheone, R. (2010). *Developing an internationally comparable balanced assessment system that supports high-quality learning.* Presentation at the National

Conference on Next Generation K–12 Assessment Systems, Center for K–12 Assessment & Performance Management with the Education Commission of the States (ECS) and the Council of Great City Schools (CGCS), Washington, DC. Retrieved from http://www.k12center.org/rsc/pdf/Darling-HammondPechoneSystemModel.pdf

Haladyna, T. (2004). *Developing and validating multiple-choice test items.* Third Edition. Mahwah, NJ: Erlbaum.

Kennedy, C. (2007). *The BEAR assessment system: A brief summary for the classroom context.* BEAR Report Series, 2005-03-01. Berkeley, CA: University of California.

Mislevy, R., Almond, R. & Lukas, J. (2004). *A brief introduction to evidence-centered design.* (*CSE Report 632).* Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, (CRESST), Graduate School of Education & Information Studies, UCLA. Retrieved from http://www.cse.ucla.edu/products/reports/r632.pdf

Mislevy, R. & Haertel, G. (2006). *Implications of evidence-centered design for educational testing (Draft PADI Technical Report 17).* Menlo Park, CA: SRI International. Retrieved from http://padi.sri.com/downloads/TR17_EMIP.pdf

Mislevy, R. & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology (PADI Technical Report 9).* Menlo Park, CA: SRI International. Retrieved from http://padi.sri.com/downloads/TR9_ECD.pdf

National Assessment Governing Board. (2010). *NAEP frameworks*. Retrieved from http://www.nagb.org/publications/frameworks.htm

National Governors Association & Council of Chief State School Officers. (2009). *Common Core State Standards Initiative validation committee announced*. Washington, DC: CCSSO. Retrieved from http://www.nga.org/portal/site/nga/menuitem.6c9a8a9ebc6ae07eee28aca9501010a0/?vgnextoid=f541ea15a18e3210VgnVCM1000005e00100aRCRD&vgnextchannel=759b8f2005361010VgnVCM1000001a01010aRCRD

National Governors Association & Council of Chief State School Officers. (2010a). *The Common Core State Standards Initiative.* Presentation retrieved from http://www.corestandards.org

National Governors Association & Council of Chief State School Officers. (2010b). *Designing common state assessment systems.* Washington, DC: CCSSO. Retrieved from http://www.nga.org/Files/pdf/1004NGACCSSOASSESSMENTS.PDF

National Governors Association Center for Best Practices. (2010). *Setting statewide college- and career-ready goals.* Washington, DC. Retrieved from http://www.nga.org/Files/pdf/1008COLLEGECAREERREADYGOALS.PDF

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

SMARTER Balanced Assessment Consortium. (2010). Application notice inviting applications, Race to the Top program, Comprehensive Assessment Systems. CFDA Number: 84.395B. Submitted by Washington State, as fiscal agent, on June 23, 2010. Washington, DC: USED.

SRI, International. (2010). *Principled Assessment Designs for Inquiry* (PADI). Retrieved from http:// http://padi.sri.com/

Webb, N.L. (2006). Identifying content for student achievement tests. In S.M. Downing & T.M. Haladyna (Eds.), *Handbook of test development* (pp. 155–180). New Jersey: Lawrence Erlbaum Associates.

WestEd. (2010). *Technology and engineering literacy framework for the 2014 National Assessment of Educational Progress.* (Pre-Publication Edition). Redwood City, CA: Author.

White, April D. (2010). *Implementing the Common Core Standards.* Blueprint. Number 4. The James B. Hunt, Jr. Institute for Educational Leadership and Policy. Retrieved from http://www.huntinstitute.org/elements/media/files/Blueprint-Number-4-June-2010.pdf

Wilson, M., Herman, J., Schneider, S., Shavelson, R. & Timms, M. (2003). *A framework for assessment in science.* U.S. Panelist Papers**.** DFG-NSF Conference: An International Conference on Research and Development in Mathematics and Science Education. Leibniz-Institute for Science Education (IPN), Kiel, Germany. Retrieved from http://cltnet.org/archive/dfgnsf/docs/KielUSPlenaryPapers.pdf

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Appendix A:
Annotated Resources Related to Development of a Comprehensive Assessment Framework

Achieve, Inc. (2010). *Comparing the Common Core State Standards in mathematics and NCTM's Curriculum Focal Points.* Washington, D.C.: Author. Online: http://www.achieve.org/CCSSandFocalPoints

> This brief compares the CCSS in mathematics to NCTM's *Curriculum Focal Points* in terms of rigor, content, coherence, focus, and specificity. Achieve's analysis suggests that the two are similar in rigor, content, coherence, and focus; however, their levels of specificity differ due to their unique purposes.

Achieve, Inc. (2010). *Comparing the Common Core State Standards in mathematics and the NAEP Framework.* Washington, D.C.: Author. Online: http://www.achieve.org/files/CCSSMathandNAEP.pdf

> This brief compares the CCSS in mathematics to the NAEP Framework in mathematics in terms of rigor, coherence, and focus. Achieve's analysis suggests that students who master the CCSS will be well prepared for the NAEP exam and that the CCSS and NAEP Framework are similarly focused.

Achieve, Inc. (2010). *On the road to implementation: Achieving the promise of the Common Core State Standards*. Washington, D.C.: Author. Online: http://www.achieve.org/files/FINAL-CCSSImplementationGuide.pdf

> This guide to implementing the CCSS into states' college-and career-ready agendas focuses on aligning instructional materials and assessments with the CCSS from both planning and policy perspectives.

Achieve, Inc. (2010). *Understanding the Common Core State Standards.* Washington, D.C. Online: http://www.achieve.org/files/CCSSJune22010FINAL.ppt

> This presentation provides a brief overview of the CCSS and describes the development process in each content area.

Common Core, Inc. (2010). The *Common Core curriculum mapping project*. Online: http://www.commoncore.org/maps/favicon.ico

> The Bill and Melinda Gates Foundation funds this teacher-led initiative to develop curriculum maps in ELA. They translate the CCSS for each grade into unit maps that teachers can use to plan their instruction, craft their own more detailed curriculum, and create lesson plans.

Common Core State Standards Initiative. (2010). *Mathematics standards, Appendix A: Model course pathways in mathematics.* Online: http://www.achieve.org/files/CCSSIMathematicsAppendixA.pdf

This appendix provides critical information about learning progressions in mathematics to guide course development and instruction at the high school level.

Common Core State Standards Initiative. (2010). *Messaging toolkit: Common Core State Standards Initiative.* Online: http://www.fldoe.org/board/meetings/2010_06_15/toolkit.pdf

This document is an informational guide to the CCSS. The intent of the document is to help CCSS advocates best argue their support of the initiative. The toolkit includes answers to a comprehensive list of FAQs, a section on CCSS myths and facts, key talking points, suggested activities, and samples of letters to the editor, etc. The information in this toolkit provides a strong foundation for understanding the CCSS.

Confrey, J. (2010). *What now? Implications, obligations and opportunities for curriculum after the release and adoption of the Common Core State Standards.* Presentation at the Center for the Study of Mathematics Curriculum's Conference on Curriculum Design, Development and Implementation in an Era of the Common Core State Standards. Arlington, VA. Online: http://www.mathcurriculumcenter.org/conferences/ccss/JConfrey.htm

This 85-slide PowerPoint analyzes the CCSS for mathematics by exploring five strategies for systemic change and curriculum: phasing the implementation; articulating learning trajectories from domains; linking curriculum and classroom assessment; defining and deploying the 15% maximum supplement to the CCSS in pursuit of a broader STEM agenda; and using longitudinal data systems to study curricular effectiveness.

Conley, D.T. (2010). *Creating college and career readiness.* Presentation at Beaverton School District on March 4, 2010. Eugene, OR: EPIC.

This presentation describes EPIC's methods and approaches for defining and assessing college- and career-readiness. It outlines key dimensions and principles of college- and career-readiness and describes EPIC's College Career Ready System, including its College-readiness Performance Assessment System (C-PAS) and its College Career Ready School Diagnostic.

Conley, D.T. (2010). *Beyond business as usual—Key state actions to boost college and career readiness.* Eugene, OR: EPIC.

This presentation is based on two of Conley's books: *College Knowledge* and *College and Career Ready*. Conley discusses concerns that more students are attempting

college today, yet many are struggling to be successful at that level. He defines college-readiness and its dimensions. Two exemplars of state policy frameworks for college- and career-readiness (TX and OH) are provided.

Darling-Hammond, L. & Pecheone, R. (2010). *Developing an internationally comparable balanced assessment system that supports high-quality learning.* Presentation at the National Conference on Next Generation K–12 Assessment Systems, Center for K–12 Assessment & Performance Management with the Education Commission of the States (ECS) and the Council of Great City Schools (CGCS), Washington, DC. Online: http://www.k12center.org/rsc/pdf/Darling-HammondPechoneSystemModel.pdf

This report describes the authors' priorities for developing a balanced assessment system that promotes high-quality learning. The authors' "theory of action" is an integrated system of curriculum and assessment to include summative and formative assessments, professional development, and reporting systems. The paper also suggests roles for states working within consortia, suggests how the system would operate, and provides answers to guiding questions, an overview of goals, a timeline, and related costs.

ETS, Pearson & College Board. (2010). Thoughts *on an assessment of Common Core Standards.* Princeton, NJ. Online: http://www.ets.org/s/commonassessments/pdf/ThoughtsonAssessment.pdf

This paper provides an opening discussion of issues and ideas related to the design of a comprehensive assessment system that may be of interest to various stakeholders. The document describes a high-level framework for a CCSS-based assessment, addressing such topics as the role of summative and formative assessments, test design, grades and subjects tested, test delivery, scoring, reporting, item types, accessibility and accommodations, etc. It provides a good introduction to assessment design issues.

Haladyna, T. (2004). *Developing and validating multiple-choice test items.* Third Edition. Mahwah, NJ: Erlbaum.

This text situates test development activities, including item development, squarely within the validity argument. As such, developers are responsible for ensuring that specific steps are taken before, during, and following the work of developing items. Specifically, developers are accountable for defining what they are seeking to measure (e.g., in the case of achievement testing, the knowledge, skills, and abilities in a content domain). Items that appear on achievement tests must be subjected to an iterative development process that includes multiple rounds of review and refinement. Outcomes from each round are documented as evidence to support the

validity of that item for the intended purpose. The evidence then is accumulated to support use of that item on a high-stakes achievement test. The four types of problems that can undermine the validity of items are discussed, as are the strengths of different item types for different purposes.

Kennedy, C. A. (2005). *The BEAR assessment system: A brief summary for the classroom context.* BEAR Report Series, 2005-03-01. Berkeley, CA: University of California. Online: http://bearcenter.berkeley.edu/publications/BAS_Summary.pdf

Kennedy summarizes the BEAR Assessment System's principles and building blocks for developing assessments both aligned with classroom curriculum and activities and designed to produce data useful for teachers to measure student learning.

Kennedy, C. A. (2007). *The BEAR assessment system: A brief summary for the classroom context.* BEAR Report Series, 2005-03-01. Berkeley, CA: University of California.

In this updated version of the 2005 report, the author describes ways in which the development principles and building blocks of the BEAR Assessment System applies to classroom-based assessment of progression in student learning.

Marion, S. & Shepard, L. (2010). *Let's not forget about opportunity to learn: Curricular supports for innovative assessments.* Draft document. Dover, NH: Center for Assessment.

This paper suggests approaches to aligning assessment with curriculum. It is intended to inform consortium design considerations, specifically in regard to the incorporation of performance tasks, interim assessments, and formative probes.

Mislevy, R., Almond, R. & Lukas, J. (2004). *A brief introduction to evidence-centered design. (CSE Report 632).* Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, (CRESST), Graduate School of Education & Information Studies, UCLA. Online: http://www.cse.ucla.edu/products/reports/r632.pdf

This report provides a general overview of evidence-centered design (ECD). The authors describe assessment as a special case of evidentiary reasoning in which validity is used to bridge the gap between what is observed via assessments and what is inferred from the results. They describe primary models for the design and delivery of coherent assessments: the conceptual assessment framework (CAF) and the four-process architecture delivery. These models are the core components of an ECD framework, which the authors believe to be integral to carrying out the actual steps of designing and implementing an assessment.

Mislevy, R. & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology (PADI Technical Report 9).* Menlo Park, CA: SRI International. Online: http://padi.sri.com/downloads/TR9_ECD.pdf

The authors use examples from the PADI project to explain how the evidence-centered assessment design (ECD) framework can be used to design different types of computer-based assessments.

Mislevy, R. & Haertel, G. (2006). *Implications of evidence-centered design for educational testing (Draft PADI Technical Report 17).* Menlo Park, CA: SRI International. Online: http://padi.sri.com/downloads/TR17_EMIP.pdf

The authors, researchers at the University of Maryland and SRI International, describe the layers in evidence-centered assessment design (ECD) and associated Principled Assessment Design for Inquiry (PADI) project tools, with a focus on the significance for large-scale assessments using computer-based tasks.

National Assessment Governing Board. (2004). *NAEP 2009 science framework development: Issues and recommendations.* Online: http://www.nagb.org/publications/reports-papers.htm

This report, commissioned by NAGB, describes processes used in the design and development of the 2009 NAEP Framework in science and the NAEP science assessment.

National Governors Association & Council of Chief State School Officers. (2009a). *CCSSI K-12 standards development teams.* Online: http://www.nga.org/Files/pdf/2010COMMONCOREK12TEAM.PDF

Names and contact information for all members of the CCSS development committees in ELA/literacy and mathematics are provided.

National Governors Association & Council of Chief State School Officers. (2009b). *Common Core State Standards initiative Validation committee announced*. Washington, DC: CCSSO. Online: http://www.nga.org/portal/site/nga/menuitem.6c9a8a9ebc6ae07eee28aca9501010a0/?vgnextoid=f541ea15a18e3210VgnVCM1000005e00100aRCRD&vgnextchannel=759b8f2005361010VgnVCM1000001a01010aRCRD

This press release announces the names of the members of the Validation Committee for the Common Core State Standards Initiative. This panel reviewed the standards development process and the college- and career-readiness standards. It includes biographical information on each of the Validation Committee members.

National Governors Association & Council of Chief State School Officers. (2010). *Designing common state assessment systems.* Washington, DC: CCSSO. Online: http://www.nga.org/Files/pdf/1004NGACCSSOASSESSMENTS.PDF

This paper briefly describes the vision and shared priorities of CCSSO and NGA regarding the design and development of next generation assessment systems. It details the similarities and differences between the two major assessment consortia that will be developing next generation assessments for the CCSS (SBAC and PARCC).

NGA & CCSSO Center. (2010). *The Common Core State Standards Initiative.* PPT Presentation. Online: http://www.corestandards.org

This presentation briefly discusses the new CCSS. The slides provide a basic overview of the standards development process, as well as the design, organization, and content of the standards.

NGA Center for Best Practices. (2010). *Setting statewide college- and career-ready goals.* Washington, DC. Online: http://www.nga.org/Files/pdf/1008COLLEGECAREERREADYGOALS.PDF

This policy brief suggests how to establish college- and career-ready performance goals at the state level. Five specific steps are presented, as well as five performance measures on which to report.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

This seminal work describes the assessment triangle, with the cornerstones of cognition (how students represent knowledge and develop competence), observation (tasks that allow the collection of evidence about what a student knows and can do in a domain through observation of a student's response to questions), and interpretation (methods for drawing inferences about the meaning of the evidence observed). These three elements must be explicitly connected in an assessment framework. The authors also describe the theoretical foundation for a measurement model in which theory about how people develop competence in a domain helps RFA developers understand how different types of measures are more useful than others in eliciting information about what students know and can do in a domain. These practices are part of building a body of evidence to support claims that valid inferences can be drawn from results.

WestEd. (2010). *Technology and engineering literacy framework for the 2014 National Assessment of Educational Progress.* (Pre-Publication Edition). Redwood City, CA: Author.

This 158-page document details processes and outcomes from recent work developing the NAEP Framework in Technology and Engineering Literacy that will guide development of a test to be administered in 2014. The content identified as appropriate for assessment is identified, as are the types of items and measures most appropriate for assessing this content. WestEd used a steering committee; an expert panel approach that included technology specialists, engineers, teachers, IHE and workplace representatives, and national organization leaders; input from 18 outreach meetings held across the nation, and stakeholder surveys. The developers recommend using a model that is "inclusive, deliberate, and designed to achieve as much broad-based input as possible" (1–11).

White, April D., Ed. (2010). *Implementing the Common Core Standards.* Blueprint. Number 4. The James B. Hunt, Jr. Institute for Educational Leadership and Policy. Online: http://www.hunt-institute.org/elements/media/files/Blueprint-Number-4-June-2010.pdf

This article proposes methods of CCSS alignment analysis for states to conduct in order to facilitate their implementation of the CCSS. The authors highlight the importance of strong alignment between standards, curriculum, and assessment, and describe the six basic elements of standards alignment analysis. Specifically, options for contributing to an assessment consortium are discussed. The article will be useful to State leaders and Consortia members seeking to use alignment studies to support assessment design.

Wilson, M., Herman, J., Schneider, S., Shavelson, R. & Timms, M. (2003). *A framework for assessment in science.* U.S. Panelist Papers**.** DFG-NSF Conference: An International Conference on Research and Development in Mathematics and Science Education. Leibniz-Institute for Science Education (IPN), Kiel, Germany. Online: http://cltnet.org/archive/dfgnsf/docs/KielUSPlenaryPapers.pdf

The Center for Assessment and Evaluation of Student Learning (CAESL) proposes a framework for assessment that designers can use when developing specifications for district and state science assessments aligned to a set of standards.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Dr. Wilson spells out the steps he recommends in building coherent assessments from which trustworthy inferences can be drawn about what students know and can do. He first recommends identifying the construct(s) to be assessed and developing maps

that fully lay out all elements of the content domain. Next, items are designed that are linked conceptually to the content domain and that fit the model (theory of action) for measurement in that domain. Considerations related to the types of information and level of specificity needed for each item type are provided to ensure that outcomes lead to the development of items that are consistent in quality and usefulness and item pools that are able to accommodate blueprints for a variety of assessment purposes. Pp. 59–61 delineate recommended steps for preparing a panel for item development.

Appendix B:
What Did the Consortium Say It Would Do in Relation to an Assessment Framework?

Comments and Page References from the SBAC's Application for Race to the Top Funding

From p. 73:

System development starts with clear specification of the constructs to be measured, intended users and uses, and target student population, so that items and performance events comprising each assessment can be purposely designed to address the intended learning construct(s) in ways that are consistent with intended uses and maximally accessible for all students. Our approach brings together the last decade's work on the logic of assessment as a process of reasoning from evidence with principles of evidence-based design and model-based assessment (NRC, 2001; Mislevy & Riconscente, 2006; Baker, 2007). Its hallmarks include clear specification of progressions of learning expectations, test blueprints with clear rules for sampling the full domain of those learning expectations, learning-based item and performance event design templates, and the potential for innovation and efficiency through technology-enhanced items and automated scoring.

From pp. 74–76:

Building the assessment frameworks, test blueprints, item templates, and specifications. Linking standards and assessments, as well as making expectations clear for teaching and learning, requires that we first specify standards in ways that precisely identify what is to be assessed and how proficiency is defined relative to that content (i.e., at what levels of cognitive demand are students expected to know that content and how should they be able to apply it during testing?). For example, the CCSS for mathematics lay out the major domains of mathematics and the clusters and standards for each grade constitute the content of what students are supposed to know, but we must also specify what students are expected to be able to do with the content. Building from the CCSS, we propose to convene key stakeholders and content specialists to develop assessment frameworks that precisely lay out the content and cognitive demands that define college- and career-readiness for each grade level. We then will systematically develop items and performance events that collectively provide evidence for valid conclusions about what a student knows and can do in relation to college- and career-readiness.

The vertical pathways to college- and career-readiness embedded in the CCSS across grades also enable us to hypothesize implicit learning progressions (i.e., empirically validated descriptions of how learning typically unfolds within a curricular domain [Darling-Hammond & Pecheone, 2010]) and what prior understandings are most essential for future performance. Similarly, within grades, with the help of teachers and learning theorists, we will identify within year learning progressions on which to base the

summative and I/B assessments and formative tools. We plan to validate and strengthen these hypothesized progressions as one element of our planned research agenda.

Once the comprehensive assessment framework that defines learning expectations is developed, assessment blueprints will be constructed that specify how items and performance events can work in various combinations to measure the full range of the CCSS. Performance event design templates then can be created to provide more concrete detail about the nature of test items and/or performance events that best measure different types of knowledge and skills. In this way, the content and cognitive demand are used in combination to describe and measure the development of competency within a domain (NRC, 2001).

Based on prior research and learning theory (Baker, 2007), we plan to develop reusable event design templates for generating a variety of selected- and constructed-response items and performance events. These templates will be designed to measure specific types of understanding (e.g., conceptual, procedural, or strategic knowledge) or skills (e.g., problem solving, reasoning, or argumentation). This type of reusable event design template or shell invokes a consistent set of principles, criteria, and expectations for defining appropriate content limits and generating events and scoring criteria to address particular cognitive demands that can be adapted for use across subject matter domains and grade levels. For example, using a common definition of problem solving, an appropriate event design might feature a non-routine, authentic problem (at various degrees of near or far transfer) that requires students to integrate core concepts and principles they have learned and apply reasoning to develop a novel solution strategy. In doing so, they follow the general path of identifying the problem within the novel context, identifying critical variables and their relationships, constructing a suitable representation of the problem, solving the problem, justifying their strategy, and/or communicating results (see, e.g., Mayer, 1992; NRC, 1999; Hiebert et al., 1996).

In the past, such authentic problem-solving tasks have been difficult for State assessment programs to implement in ways that are appropriate for large-scale assessment. Complex unstructured problems may have enabled high-competency students to show what they know and can do, but may have left low-ability students unable to respond. The remedy of using highly structured tasks may have provided access for low-ability students, but it does not reflect true problem solving. Over the past two decades, this picture has improved. The technology for building such items and tasks has evolved significantly, to the point where it may increase access and generalizability across tasks (Baker, 2007; Darling-Hammond & Pecheone, 2010). That progress allows our performance event designs, coupled with the use of technology, to offer a possible solution: technology-enhanced adaptive items and performance events can provide additional challenge to those students who are ready to move forward, while at the same time providing progressive scaffolding and hint systems to customize difficulty and diagnosis for other students.

Building in support for flexibility and accessibility, the templates will provide item developers with efficient, standard routines for item development and scoring, which ultimately may provide capacity for automated development. Further, by supporting constancy and comparability in the nature of the learning assessed from year to year, this approach also will allow for better measurement of growth over time along key domains and/or learning progressions. Initially, we will develop detailed design templates for the summative performance events to support comparability from year to year. However, these templates also provide an easy link among summative, I/B, and formative functions and classroom instruction. That is, teachers can use the design templates to generate rich classroom assignments and curriculum-embedded assessment tasks.

Building from the CCSS, the Consortium will develop cognitive models for the domains of ELA/literacy and mathematics that specify the content elements and relationships reflecting the sequence of learning that students would need to achieve college- and career-readiness. Implementation of these models will require close collaboration with IHEs, workplace representatives, cognitive scientists, classroom teachers, experts in assessing English learners and students with disabilities, and content experts in ELA or mathematics who have studied the CCSS and have deep understanding of the knowledge, skills, and cognitive processes that characterize the pathways in grades 3 through 8 and high school that lead to the college- and career-ready outcomes specified in the standards. Once the content of the CCSS has been studied, an assessment framework will be developed that specifies the item types that can provide the necessary evidence of student achievement.

The next critical step in the process is the development of item and performance event shells. These will provide constancy in the set of constructs measured by a set of items and events. They also will allow for better measurement of achievement and growth, development of equivalent tasks, and a framework to guide the development of scoring rubrics. The item/event shells will be complemented by the development of item specifications that will delineate the content limits to be reflected in the items. Once the assessment framework, item/event shells, and item specifications have been established, the test blueprint will be developed that will specify the number and types of items to be established in the summative item banks. An overview of the development process for the summative assessments can be found in Appendix A4-2 of the SBAC application to USED for Race to the Top Comprehensive Assessment System funding.

From p. 111:

Also, in the name of consistency, SBAC will work with participating States to develop common frameworks for assessment, guidelines for effective formative assessment practices, and research-supported exemplars of curricular/instructional materials to support teachers' professional growth.

From p. 112:

SBAC's professional capacity-building efforts will begin immediately upon project funding. As the CCSS are the fundamental building blocks of the assessment system, our first priority will be to monitor ongoing State- and district-level efforts to unpack these standards at each grade level. As a first step, the Professional Capacity and Outreach Working Group will collect, review, and share exemplary resources from participating States. Web-based discussions will focus on the concepts, skills, and knowledge embedded in the CCSS; the learning pathways students must follow to reach valued college- and career-ready outcomes; and strategies for developing CCSS-aligned curriculum and instruction. SBAC will work with State teachers, technical advisors, and representatives from IHEs and the workplace to develop curriculum frameworks that address the full range of the CCSS and that clearly articulate what students must learn within and across grade levels. For the interim/benchmark assessments, the Professional Capacity and Outreach Working Group also will support the curriculum alignment to the learning progressions and the CCSS domains of ELA/literacy and mathematics.

Appendix C:
What Did the Consortium Say It Would Do For Each System Component in Relation to Item Pools or Item Types?

Comments and Page References from the SBAC's Application for Race to the Top Funding

Adaptive Summative Assessments

From pp. 52–53:

The adaptive summative assessment will include selected-response items, technology enhanced items, and extended constructed-response items and will be combined with the performance events to provide for the assessment of declarative and procedural knowledge along with the extended problem solving, inquiry, and synthesis expected for college- and career readiness. The proposed distribution of item types is based on the number of CCSS, the preliminary analysis of their content, and the need to produce a scale that will include enough score points to reliably measure student abilities. Sample items for each item type included in the assessment system are presented in Appendix A3-3.

Whereas selected-response items have routinely been a part of assessments of student learning, their emphasis has too often been on low-level skills reflecting knowledge of discrete pieces of information that are not critical for subsequent student learning. The emphasis here will be on the development of items that reflect important knowledge and skills consistent with the expectations of the CCSS. Items can be developed to address knowledge and skills from more than one standard. The appropriate and judicious use of selected-response items provides for a cost effective means to address content in terms of test development, administration, and scoring.

The Consortium is committed to making effective use of technology by including items that capitalize on the capabilities of the assessment platform. The effective use of technology can expand not only the nature of the content that can be presented but also the knowledge, skills, and processes that can be assessed (Quellmalz & Moody, 2004). Technology-enhanced items can take advantage of drag-and-drop, hot spot, and simulation technologies along with the use of online tools to measure content that was previously not assessed or was assessed through constructed response item formats requiring more elaborate scoring procedures. While most of the work with the technology-enhanced items has been done in the area of science, we envision the development of technology-enhanced items in both English language arts and mathematics. For example, a video presentation of a speech could be combined with a reading passage to provide students the opportunity to integrate reading and listening skills per the CCSS and evaluate content across diverse media.

Our assessment design calls for the use of extended constructed-response items. These items will be used to assess knowledge and skills not easily assessed with selected-

response or technology-enhanced items. Constructed-response items allow students to demonstrate their use of complex thinking skills such as formulating comparisons or contrasts; proposing cause and effects; identifying patterns or conflicting points of view; categorizing, summarizing, or interpreting information; and developing generalizations, explanations, justifications, or evidence-based conclusions (Darling-Hammond & Pecheone, 2010). These complex thinking skills are consistent with the expectations for college- and career-readiness and will be included in both the English language arts and mathematics assessments.

The summative assessment system also includes the strategic use of performance events in both English language arts and mathematics. We anticipate two performance events per content area for grades 3–8 and up to six performance events in each content area by the end of grade 11. These events will evaluate the CCSS in ways that require more student-initiated planning, management of information and ideas, interaction with other materials and/or people, and production of more extended responses (e.g., oral presentations, exhibitions, product development, in addition to more extended written responses) that reveal additional abilities of students (Darling-Hammond & Pecheone, 2010) not captured by the other item types included in the summative assessment.

Interim/Benchmark Assessments

From p. 42:

The second key component of SBAC's assessment system is adaptive I/B assessments, built around learning progressions. Learning progressions describe how learning typically unfolds (Darling-Hammond & Pecheone, 2010, p. 9). Results from these assessments can be used by teachers to develop targeted instructional strategies aimed at addressing specific gaps in understanding to see where each student currently is situated on the learning continuum toward college- and career-readiness. These I/B assessments play a key role in our system because they provide more finely grained information about student progress toward college- and career readiness than can be provided by the summative assessment. They can be administered at various points within the instructional year to support effective instructional decisions at the student level. This pool of items will contain the same types of items and performance events as the summative pool and will include all released summative items and events.

From p. 77:

The I/B assessments are designed to provide a measure of progress toward the summative goals and to help identify learning gaps and guide instruction (Rabinowitz, 2009). Consequently, I/B assessments will be developed to zero in on a student's current level of understanding in each learning progression in English language arts and mathematics.

Learning progressions are empirically validated descriptions of how learning typically unfolds within a curricular domain or area of knowledge and skill (Darling-Hammond & Pecheone, 2010). By design, a set of items developed to measure learning progressions for the purposes of the I/B assessments will be deep in terms of content coverage for each content cluster in the CCSS—a larger number of items will be used to measure small, incremental differences in what students know and can do. Item clusters will be developed that can hone in on students' precise level of understanding of those linked pieces of knowledge and/or demonstration of skills that constitute a progression.

The I/B assessments will require the specification of the learning progressions as a first step in the development process. Once identified, these learning progressions will be mapped to the CCSS and the evidence-based model again will be applied to determine the knowledge and/or skills a student must demonstrate to show mastery of the steps in the learning progression. As with the summative assessment, an assessment framework, test blueprint, and item specifications will be developed to guide the development of the I/B assessment item bank.

From pg. 74:

The vertical pathways to college- and career-readiness embedded in the CCSS across grades will enable the identification of learning progressions (i.e., empirically validated descriptions of how learning typically unfolds within a curricular domain [Darling-Hammond & Pecheone, 2010]) and what prior understandings are most essential for future performance. Then, within grades, with the help of teachers and learning theorists, they can identify within-year learning progressions that also will be included in the framework documents. These steps will require a review and synthesis of literature on the development of expertise in each content domain.

Formative Tools and Processes

From pp. 42–43:

The third component of the Consortium's assessment system is a set of formative tools, processes, and practices that support the needs of teachers on a daily basis as they help students learn and progress on their path toward college- and career-readiness. The goal of our formative assessment component is to enable full implementation of the learning and assessment system by helping teachers and administrators effectively use data from the summative and I/B assessments and to build their capacity to collect evidence during instruction that can be useful in diagnosing students' learning needs. To build these capacities, the Consortium will develop and disseminate research-supported tools, resources, and materials. It also will provide opportunities for training and collaboration through professional learning work groups to support the development of assessment literacy; an understanding of the content and performance expectations of the CCSS;

development of model curriculum units; alignment of curriculum, instruction, and assessment; development of teacher capacity in the reasoning-from-evidence approach to item development supported by the Consortium; an understanding of the expectations of student performance through the effective use of scoring activities and released events with annotated student work; and an in-depth understanding of the learning process to sustain improved instructional practice. By combining these formative practices and tools with the summative and I/B assessments, we are developing a system for learning and assessment that will lead to more informed decision making and will result in higher-quality instruction, and thus, higher levels of student achievement. Further, SBAC is committed to developing an assessment system that reflects the principle of responsible flexibility. We want to provide States within the Consortium flexibility in implementing the components of our assessment system in a way that best meets their needs.