Smarter Balanced
Assessment Consortium

End of Grant Report

JUNE 2015

CONTENTS

End of Grant Report

CONTENTS *continued*

End of Grant Report

EXECUTIVE SUMMARY

The Smarter Balanced Assessment Consortium has built a comprehensive learning and assessment system in English Language Arts/Literacy (ELA/L) and mathematics, the components of which work together toward the goal that all students leave high school prepared for postsecondary success in college or career through increased student learning and improved teaching. The comprehensive assessment system includes summative and interim assessments; a digital library of formative tools, processes, and practices; and a suite of open-source applications. The purpose of this report is to summarize the work done to build the assessment system for use beginning with the 2014-2015 school year.

The first chapter of this report discusses the principles underlying the Consortium's assessment system. The Consortium first articulated its vision for a world-class assessment system in the 2010 Race to the Top assessment grant application. The states that joined the Consortium committed to having common content standards and achievement standards. In addition, these states agreed that their students would have the same resources and accommodations available during testing. Finally, they agreed to support educators' professional development to better meet students' learning needs. Throughout the development of its comprehensive assessment system, the Consortium's work was guided by the following ideals:

1. Assessments are grounded in a thoughtful, standards-based curriculum and are managed as part of an integrated system of standards, curriculum, assessment, instruction, and teacher development;
2. Assessments produce evidence of student performance on challenging tasks that evaluate student achievement on the Common Core State Standards (CCSS);
3. Educators are integrally involved in the development and scoring of assessments;
4. The development and implementation of the assessment system is a State-led effort with a transparent and inclusive governance structure;
5. Assessments are structured to continuously improve teaching and learning;
6. Assessment, reporting, and accountability systems provide useful information on multiple measures that is educative for all stakeholders;
7. Design and implementation strategies adhere to established professional standards.

The comprehensive assessment system includes summative and interim assessments; a digital library of formative tools, processes, and practices; and a suite of open-source applications.

Chapter 2 articulates the purposes of the summative and interim assessments and then discusses the evidence that support those uses. For example, one purpose of the summative assessment is to provide valid, reliable, and fair information about students' ELA/L and mathematics achievement with respect to the CCSS. Throughout Chapter 2, we discuss the extensive work based on the principle of evidence-centered design done over the past four years to ensure that the assessment system produces results that support the intended uses. To ensure that the assessment results were appropriate for the intended purposes, the first step was to define the construct. In other words, what is meant by ELA/L and by mathematics? To begin this work, the Consortium enlisted content experts and educators to write content specifications for ELA/L and mathematics. The content specifications document was the first document where Consortium states articulated the claims they intended to measure, the evidence necessary to support those claims, and the assessment targets measured at each grade level.

Once the content specifications were developed, the Consortium brought educators together from member states to create achievement level descriptors (ALDs). The ALDs describe the knowledge, skills, and abilities that students should demonstrate on a standardized test in terms of levels or categories of performance. The creation of the content specifications and the ALDs laid the necessary groundwork for the item and task specifications. These specifications provided the guidance necessary for professional item writers and teachers to write the test questions. Test blueprints are the final foundational document of the Consortium's adaptive assessment system. The blueprints specify the type and range of content of the items that every student encounters on the assessments, thereby ensuring that the Consortium is measuring the same construct for all students.

The last portion of Chapter 2 discusses the digital library of formative tools. As with the interim and summative assessments, the Consortium articulated the planned purposes of the digital library prior to building it. The Consortium convened an expert committee to establish the quality criteria for all digital library material. Member states created State Leadership Teams, which received extensive training on the digital library and on the quality criteria. Within each member state, these teams created and trained State Networks of Educators. The materials in the digital library were created and vetted through the State Networks of Educators.

Chapter 3 discusses the work that was done to create test questions that measured the intended construct and that were accessible to all students. To do this, the Consortium recruited hundreds of teachers from member states to participate in item writing, content reviews, accessibility reviews, bias/sensitivity reviews, and item scoring. Educators, state partners, and assessment vendors developed the nearly 30,000 items in the Consortium's item pool.

This chapter also discusses the Consortium's open-source assessment delivery and reporting system. The Consortium's comprehensive tool for administering and reporting was built to allow maximum flexibility for our member states. States may use the system in its entirety, they may use components of the system, or they may use a completely different system so long as comparability from member to member can be demonstrated. Of particular importance was the Consortium's desire to maximize accessibility for all students. To this end, all member states agreed to the Consortium's Usability, Accessibility, and Accommodations Guidelines. The Consortium's open-source assessment delivery system was built according to these guidelines.

In creating the data warehouse and reporting system, the Consortium focused on the need for flexibility as each state has unique concerns regarding student data and reporting. The Consortium does not require that member states send personally identifiable information to the data warehouse. States may elect to use the Consortium's reporting system, integrate aspects of it into their own systems, or deploy their own reporting systems. The Consortium has created an Implementation Readiness Package to ensure that students receive comparable test scores regardless of the system used to deliver and score the assessments.

Chapter 4 discusses the population that took the spring 2014 field test and the expected population* who took the spring 2015 operational test. Over 4.2 million students participated in the spring 2014 field test, and nearly 7 million students participated in the spring 2015 operational test. The data from the spring 2014 field test was used to evaluate the quality of the item pools and to establish the final Smarter Balanced vertical scale.

Chapter 5 summarizes the Consortium's outreach and communication efforts. The Consortium serves a range of audiences, including state education agencies, students, parents, educators, district and school leaders, higher education faculty and leaders,

Assessments are grounded in a thoughtful, standards-based curriculum and are managed as part of an integrated system of standards, curriculum, assessment, instruction, and teacher development.

* The actual data will not be available until after the completion of this report.

business and community leaders, policy makers, and members of the general public. The Consortium uses various communication and outreach strategies in order to reach our stakeholder groups. Some strategies are common across groups, including communication through online/print media or through social media. Other strategies are tailored to specific groups (for example, newsletters targeted to teachers).

Chapter 6 discusses sustaining the work of the Consortium now that its first operational use during the 2014-2015 school year is past. Smarter Balanced remains a state-led organization with major budget, policy, and governance decisions resting with the governing states. Each Smarter Balanced member state has executed a memorandum of understanding with the University of California, Los Angeles (UCLA) to provide shared services to all member states. The member states (either individually or in groups) contracted with one or more organizations to administer the Smarter Balanced assessments to their students beginning in the 2014-2015 school year.

The Consortium is committed to allowing flexibility among members. To support this, Chapter 7 discusses the Consortium's technology standards for items and tasks as well as our Implementation Readiness Package (IRP). Through the IRP, the Consortium has defined the requirements for assessment delivery systems to deliver assessment items with authenticity; score items and tests properly; and deliver de-identified or identifiable assessment results to the Smarter Balanced Data Warehouse. The IRP allows vendors to self-test their assessment delivery systems. Member states must ensure that their vendors show evidence that their assessment delivery systems meet the requirements of the IRP.

The Consortium is committed to protecting the privacy, security, and confidentiality of student data. Chapter 8 shares the Consortium's policies on data privacy and data security. Smarter Balanced follows industry best practices for data privacy and adheres to all applicable federal laws and regulations that safeguard education records, including the Family Educational Rights and Privacy Act. In addition, the Consortium follows industry best practices for securing the storage and transmission of student records.

CHAPTER 1.
OVERVIEW OF THE SMARTER BALANCED ASSESSMENT CONSORTIUM

In 2010, the Smarter Balanced Assessment Consortium began with a simple, yet challenging vision: create a comprehensive learning and assessment system in English Language Arts/Literacy (ELA/L) and mathematics that satisfies the needs of students, parents, teachers, administrators, and policy makers. The components of this unified system would work together toward the goal that all students ***leave high school prepared for postsecondary success in college or career through increased student learning and improved teaching.*** In our original Race to the Top assessment application,[1] the Consortium identified four key deliverables necessary to achieve our goal:

- A comprehensively designed assessment system that balances summative, interim and formative components, provides access to all students (e.g., students with disabilities, English learners), and assesses the Common Core State Standards (CCSS);
- An online adaptive test administration with a secure item and performance task bank;
- A consolidated reporting system that enhances understanding of student progress toward college and career readiness for students, parents, educators, administrators, and policy makers;
- A system of professional development focused on assessment literacy.

As we document throughout this report, the Consortium has successfully produced each of these four key deliverables and created a comprehensive learning and assessment system designed to meet the needs of students, parents, teachers, administrators, and policy makers.

The components of this unified system work together toward the goal that all students *leave high school prepared for postsecondary success in college or career through increased student learning and improved teaching.*

CREATING SYSTEM CHANGE

Member states believe that by building a world-class comprehensive assessment system, we can influence and support positive change in schools, districts, and states. A comprehensive assessment system such as Smarter Balanced, however, is only one aspect of a larger commitment to education reform that incorporates change and growth throughout the education system. Therefore, in addition to participating in the consortium, member states have committed to shared policies in the following areas:

1. Common college and career ready content standards in ELA/L and mathematics;
2. Achievement levels that describe milestones of learning for all students;
3. Resources and accommodations that are available to students while testing;
4. Supporting professional development to help teachers and other educators better meet students' learning needs.

[1] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2011/12/Smarter-Balanced-RttT-Application.pdf.

GUIDING PRINCIPLES FOR THE SMARTER BALANCED ASSESSMENT SYSTEM

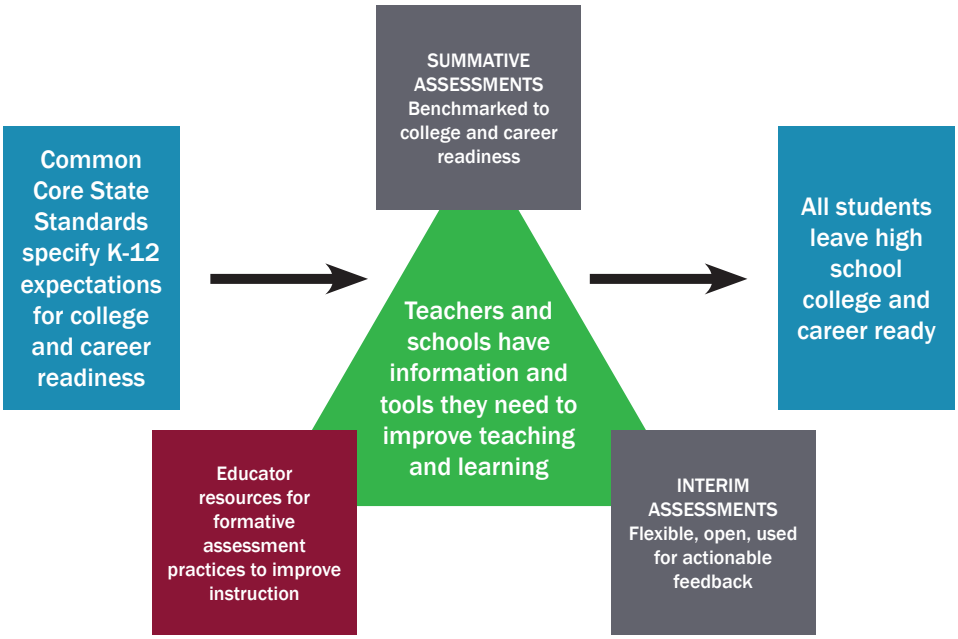
Our assessment system is shaped by principles shared by the assessment systems of high-achieving U.S. states and by high-achieving nations.[2] The Consortium's work was guided by the following ideals:

1. Assessments are grounded in a thoughtful, standards-based curriculum and are managed as part of an integrated system of standards, curriculum, assessment, instruction, and teacher development;
2. Assessments produce evidence of student performance on challenging tasks that evaluate student achievement on the Common Core State Standards (CCSS);
3. Educators are integrally involved in the development and scoring of assessments;.
4. The development and implementation of the assessment system is a State-led effort with a transparent and inclusive governance structure;
5. Assessments are structured to continuously improve teaching and learning;
6. Assessment, reporting, and accountability systems provide useful information on multiple measures that is educative for all stakeholders;
7. Design and implementation strategies adhere to established professional standards.

Designing and Developing A Comprehensive Assessment System

For the past four years, Consortium staff, member states, vendors, and, most important, educators have worked together to adhere to these principles, create the Consortium's products, and realize the promise of the Consortium's vision. The collective work of these groups has resulted in a comprehensive, technology-based assessment system that includes summative and interim assessments; a digital library of formative tools, processes, and practices; and a suite of open-source applications. Figure 1.1 shows the balanced assessment system.

FIGURE 1.1 BALANCED ASSESSMENT SYSTEM



[2] Smarter Balanced Race to the Top assessment application—see theory of action. http://www.smarterbalanced.org/wordpress/wp-content/uploads/2011/12/Smarter-Balanced-RttT-Application.pdf

End of Grant Report

The summative assessments are administered via a computer using adaptive test-delivery technology. Students taking a Smarter Balanced test answer test questions targeted to their ability level, and they engage in performance tasks structured around a classroom activity.

The comprehensive, technology-based assessment system is modular, and member states select the desired level of service. Members may purchase the base model, which provides access to only the summative assessment, or the full model, which provides access to the summative and interim assessments as well as the digital library. The summative and interim assessments cover grades 3 through 8 and high school, and the digital library provides materials for kindergarten through grade 12.

Members of the Consortium will maintain control of their data and will determine how results from both the summative and (optional) interim assessments will be provided to schools, districts, and their stakeholders. Members have the option of using the Consortium's reporting application, which allows teachers and administrators to easily see and understand the performance of their students, schools and states.

The Consortium's digital library of formative tools is available, if states opt to include access to interim assessments and the digital library in their membership fees, to educators in member states and provides a variety of resources designed to support and improve day-to-day teaching and learning. Through the digital library, educators may access exemplar instructional models, assessment literacy modules, and high-quality instructional resources and tools. In addition, the digital library serves as a network where specially trained educators in Consortium states may submit and share their own instructional resources. All submitted materials are reviewed for quality by a team of specially trained educators before they are posted. Rejected materials are returned to the submitting educator with actionable feedback for improving the resource until either it meets the quality criteria or is rejected again.

Role of Consortium Members

The Consortium was founded with the intent to share costs and improve quality. Maintaining membership is critical to meeting these goals. As of December 2014 the Consortium consists of 19 Governing members[3] and three Affiliate members. A key strategy to maintain membership is the Consortium's democratic governance system. Through work groups that included more than 80 individuals, experts from member states had an active role in policy and test design decisions. It was challenging but necessary to not only provide each state with a guiding voice in the process, but foster the shared sense of commitment and responsibility needed to successfully accomplish the goals of the project.

Work groups were focused around specific topics (e.g., Item Development, Reporting) and were composed of representatives from member states. Each Governing State committed to participate in two or more working groups while Advisory States were strongly encouraged to participate as fully as possible. This gave states a voice in shaping policy and test design decisions. States shared both common and state-specific concerns, and working groups hammered out solutions that were considerate of both state and consortium-wide concerns.

Member states had an active voice in shaping policy and test design decisions.

[3] Prior to November 1, 2014, states joined the Consortium as Governing or Advisory States. Territories and commonwealths of the United States and/or the Department of Defense Education Activity (DoDEA) joined as Affiliates. Governing States are fully committed to Smarter Balanced and have a vote in policy decisions, while Advisory States may participate in work groups and provide guidance for the development of the assessment system. After October 31, 2014, "Advisory State"was no longer a membership category. The membership category of "Affiliate" became available to states as well as territories and commonwealths of the United States, the DoDEA, and the Bureau of Indian Education (BIE).

An executive committee elected by Governing members manages the Consortium's day-to-day work. The Executive Committee approves Consortium expenditures based on the budget adopted by Governing members and advises Consortium staff regarding how to best meet the needs of Consortium members.

Consortium Members

Figure 1.2 shows the Consortium's members as of December 2014. This map has changed over the course of the Consortium's history. The Consortium began with 17 Governing members and 14 Advisory members. In spring 2015, 20 members participated in the summative assessment. Table 1.1 shows the changing membership since 2010.

FIGURE 1.2 MAP OF GOVERNING AND AFFLIATE MEMBERS DURING 2014-15 ADMINISTRATION

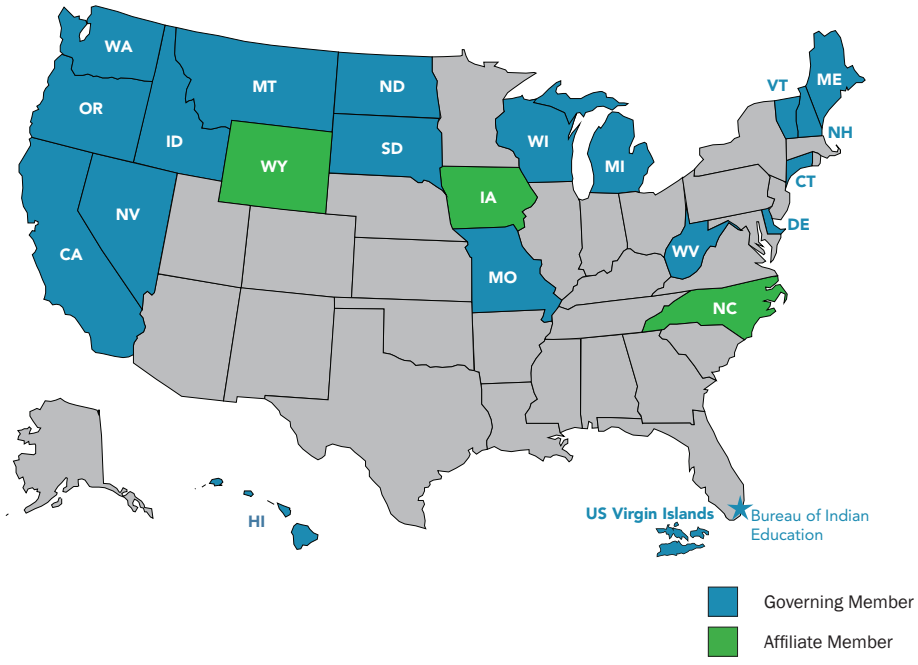


TABLE 1.1
CONSORTIUM MEMBERSHIP SINCE JUNE 2010

	APPLICATION DATA (JUNE 2010)	JULY 1, 2011	JULY 1, 2012	JULY 1, 2013	JULY 1, 2014	DECEMBER 2014
Governing States	17	19	22	21	21	19[4]
Advisory States	14	10	5	4	1	0
Affiliates				1 (territory)	1 (territory)	3

[4] Includes the Virgin Islands as a Governing member of the operational Consortium. As a territory, the Virgin Islands were precluded from receiving Federal Race to the Top Grant Funds. The Bureau of Indian Education joined the Consortium in 2015.

ESTABLISHING A WORLD-CLASS COMPREHENSIVE ASSESSMENT SYSTEM

The Smarter Balanced assessment system was built on the foundation of the seven principles of assessment described above. In addition, the design and development of the system and each of its components were informed by the technical criteria for high-quality assessment systems established in the joint Standards for Educational and Psychological Testing,[5] the Operational Best Practices for Statewide Large-Scale Assessment Programs,[6] and the emerging criteria for high-quality next generation assessments, particularly those aligned to the Common Core State Standards.[7]

State-led Effort with a Transparent and Inclusive Governance Structure

Smarter Balanced is a consortium of 20 Governing members and three Affiliate members with a transparent, consensus-based governance structure. From each state, the Consortium has a K-12 State Lead, appointed by the State Superintendent (or similar position); and a Higher Education State Lead, appointed by the State Higher Education Commissioner (or similar position). On significant policy matters, governing states have one vote (the State Superintendent or delegate), and the vote must be unanimous to pass on the first vote. If a first vote is not unanimous, consortium leadership confers with dissenting states to determine if the dissent can be addressed. A second vote must receive a majority.

To assure that higher education input is incorporated into policy decisions on matters affecting higher education, the voting member from each governing state must indicate that the vote is agreed to by both the K-12 and Higher Education State Leads.

An Executive Committee composed of elected individuals from member states leads the Consortium. The Smarter Balanced Executive Committee oversees the development of the assessment system and management of the Consortium's funding. The Executive Committee is comprised of two co-chairs, four representatives from four separate Governing States, one representative from the Lead Procurement State (Washington), and two representatives from higher education. Executive Committee members and co-chairs are elected by the Governing States through votes of either the K-12 State Leads or the Higher Education State Leads (in the case of members representing higher education).

In order to maintain transparency with all member states, the Consortium has developed several methods of communication and outreach. All stakeholders have access to the Smarter Balanced website (www.smarterbalanced.org), which serves as a primary means for communicating up-to-the minute Consortium news as well as detailed content information. The Consortium also hosts weekly meetings for member states, and biannual collaboration conferences where member states and Consortium vendors can meet to discuss issues. The Consortium sends out weekly e-newsletters to member states and others with whom we collaborate and bimonthly e-newsletters to a broader audience including teachers.

Smarter Balanced is a consortium of 20 Governing members and three Affiliate members with a transparent, consensus-based governance structure.

Teachers are Key

From the beginning, the Consortium recognized the importance of teacher involvement in the design and development of the comprehensive assessment system. As teachers are the primary users of the assessment components and information from the

[5] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

[6] Council of Chief State School Officers [CCSSO] & Association of Test Publishers [ATP]. (2013). Operational best practices for statewide large-scale assessment programs. Washington, DC: CCSSO

[7] http://www.ccsso.org/Documents/2014/CCSSO%20Criteria%20for%20High%20Quality%20Assessments%2003242014.pdf

assessments, the Consortium believes teacher involvement is critical to building a system that will be useful in deepening their understanding of student learning and supporting their efforts to provide instruction aligned to the CCSS.

To this end, the Consortium has prioritized the involvement of classroom educators in the creation of our comprehensive assessment system. Hundreds of teachers, representing every member state, were involved at every stage of building the assessment. Teachers from across the Consortium wrote items for the Smarter Balanced interim and summative assessments. Educators from Consortium states scored student work following the field test, recommended cut scores for Consortium achievement levels (similar to cut scores already in use in many states to identify a student as "proficient"), and developed and reviewed materials for the digital library. In recognition of the vital role that educators play, Smarter Balanced created a digital library to increase access to high-quality materials for educators.

An Integrated System

All components of the Consortium's comprehensive assessment system are aligned to the Common Core State Standards (CCSS). The Consortium's summative and interim assessments were developed to measure the CCSS. In addition, the instructional supports and curricular materials offered in the Consortium's digital library are explicitly aligned to the CCSS. To further support teachers, the Consortium released sample items and performance tasks as examples of the knowledge and skills students are expected to demonstrate on the interim and summative tests. In addition, the Consortium has released training and practice tests to help students and teachers prepare for the summative assessments.

Since all components of the Consortium's comprehensive assessment system are aligned to the CCSS, teachers can use Smarter Balanced tools to support instruction and evaluate student progress throughout the school year. In planning instruction, teachers can align their instruction to the CCSS using materials from the digital library. Educators can use interim assessments to measure student performance on various aspects of the CCSS throughout the school year.[8] Since the interim assessment will provide timely feedback, teachers can adjust their instruction based on the results of the interim assessment. The summative assessment provides an end-of-year measure of performance on the CCSS. Schools can use this information as they adjust their curricular expectations, programming, policies, and school improvement plans from one year to the next.

The assessment system is also designed to support instruction and student learning across years. The CCSS are organized around learning progressions. In other words, the way students are expected to develop their content knowledge and skills within a content area across all grade levels. The design of Smarter Balanced resources as well as detailed achievement level descriptions at each grade level and the single scale used to report results from all Smarter Balanced assessments are all intended to help to tie the assessments to the learning progressions of the CCSS.

Produce Evidence of Student Performance on Challenging Tasks

The Consortium's overarching goal is that students, including students with disabilities and English language learners, leave high school well prepared for college or careers. Students who are well prepared for college or careers tend to demonstrate a variety of key cognitive strategies and content knowledge (Conley, 2012). These students will have deep knowledge of core concepts and ideas within and across content areas, and they will be able use various higher-level thinking skills such as analyzing, evaluating, and interpreting information. The CCSS embodies those principles of the importance of not only possessing but also applying deep content knowledge.

Teacher involvement is critical to building a system that is useful in deepening their understanding of student learning and supporting their efforts to provide instruction aligned to the CCSS.

[8] Only the fixed-form interim assessments was available during the first operational year (2014-15).

As a starting point in the design and development of the Smarter Balanced assessment system, the Consortium applied evidence-centered design principles to determine the types of evidence that would be required to demonstrate the knowledge and skills described by the CCSS and the types of assessments needed to provide that evidence.

Unlike traditional assessments that measure specific facts, concepts, and skills using largely multiple-choice items, the Consortium's assessments include significant numbers of additional types of test questions, such as short and long written responses, performance tasks and technology-enhanced test questions that challenge students to demonstrate deep knowledge of core concepts. Students engage in performance tasks that might, for example, provide them with the opportunity to demonstrate their skills in research or analyzing complex texts. As part of the integrated system, these types of measures are found on both the summative and interim assessments, and additional performance tasks are available for teacher use through the digital library.

Continuously Improve Teaching and Learning

The Consortium believes that effective assessments are designed to develop understanding of what learning standards are and what high quality work looks like. In addition, assessments should provide timely and accurate information on student achievement levels, on what growth is occurring, and on what is needed for student learning. The Consortium took several steps to ensure that its comprehensive assessment system was structured to continuously improve teaching and learning.

Consortium assessments were developed to adaptively measure the CCSS learning progressions in ELA/L and mathematics. Combining advanced measurement techniques and computer-based technologies, assessments are targeted to the skill level of an individual student. Teachers may assess student performance throughout the school year so that they can see what students know and can do in ELA/L or mathematics. This also allows teachers to adjust their teaching based on the needs of the students in their classrooms and to evaluate student progress and growth over time.

The Consortium's digital library provides examples of curriculum embedded assessment and assessment practices. All teachers in Smarter Balanced states paying membership fees for the interim/digital library package are provided access to the digital library, increasing equity of access to high-quality materials for all Smarter Balanced educators. This also provides Consortium educators with the opportunity to teach and evaluate students in ways that are directly aligned to the CCSS.

The digital library provides Consortium educators with the opportunity to teach and evaluate students in ways that are directly aligned to the CCSS.

As stated above, the Consortium values teacher involvement throughout the test's life cycle. Consortium teachers participated in scoring student work from the field test, and teachers can score student responses on the interim assessment. (States may elect to have their vendor score student responses on the interim assessment.) By being trained in and participating in scoring student responses to constructed-response items and/or performance tasks, teachers are able to more closely examine student work. This may increase teacher understanding of the knowledge and skills that are expected of students on these items and tasks. The Consortium shows its focus on teaching by providing teachers in Smarter Balanced states access to the digital library.

Provide Useful Information on Multiple Measures

The Smarter Balanced assessment system is much more than an end-of-year summative assessment. The system is meant to provide educators with meaningful, timely feedback throughout the year through technology-based interim assessments and resources available in the digital library to support formative assessment. To provide appropriate and meaningful feedback, the Consortium also developed multiple levels of reports for states using the Smarter Balanced Reporting System. For teachers,

student reports provide specific information about individual and aggregate areas of strength and weakness. For administrators, school reports provide information to inform curricular and professional development decisions.

Unlike traditional assessment systems that provide a measure from a single point in time, information from the Consortium's interim and summative assessments provide teachers with ongoing information about learning and improvement. Teachers can collect information throughout the year in order to assess student knowledge and growth.

Adhere to Established Professional Standards

Throughout the design and development of the assessment system, the Consortium has adhered to established and emerging professional standards for large-scale comprehensive assessment systems. Equally important, decisions on the design and development of each of the components of the system were informed by a knowledge and understanding of the intended and likely uses of the assessments and the results of those assessments.

At the beginning of the project, the Consortium assembled a Technical Advisory Committee consisting of members who not only possessed the required knowledge and demonstrated experience on large-scale assessments, but who are also on the cutting edge of the design and development of next generation comprehensive assessment systems. The Technical Advisory Committee met regularly, providing technical advice and support on key decisions on all components of the assessment system.

The Consortium drew on experiences from other fields to develop a competitive, distributed process for procuring support from large-scale assessment contractors and other key contractors, ensuring that the highest levels of expertise were available for the design and development. Over the course of the project, a total of 21 contracts were awarded to contractors that specialize in areas such as item writing, test development, psychometrics, and communication. A complete list of contracts awarded is provided in Appendix A.

Throughout the design and development of the assessment system, the Consortium adhered to established and emerging professional standards for large-scale comprehensive assessment systems.

Ongoing Test Validation

Test validity is a primary concern of all testing programs. Early in the grant period, the Consortium commissioned Stephen Sireci, a nationally-recognized expert in test validity, to create an aspirational research agenda for both short- and long-term research activities that would collect and examine evidence of the validity of the Consortium's assessment system. The first step in this process was to articulate the purposes of our assessments (these are presented at the beginning of Chapter 2). The purposes of the assessments guide the type of evidence that should be collected.

Throughout Chapter 2, we discuss the extensive work done over the past four years to ensure that content in the assessment system was well aligned to the CCSS. Chapter 3 discusses the work that was done to create test questions that measured the intended construct and were accessible to all students. The work done thus far is foundational evidence as the Consortium continues test validation work in the future.

As the test becomes operational, the pressure to collect, document, and study evidence of validity will intensify. Member states will work with the Center for Research on Evaluation, Standards, and Student Testing (CRESST) to establish and execute a validity plan based on Sireci's original research agenda.

ORGANIZATION OF THIS REPORT

The chapters of this report will describe how the Consortium fulfilled the vision described above.

This chapter, Chapter 1, presented the Consortium's original goals and highlighted how these goals were achieved. This chapter also provided an overview of the principles that guided the Consortium's efforts. It also described how those principles resulted in the three main components of the Consortium's assessment system (the summative and interim assessments, and the digital library) that fulfilled their vision of a comprehensive assessment system that can be integrated with curriculum and instruction.

Chapter 2 provides an overview of the key decisions that formed the foundation of the assessment system. These include decisions related to:

- The intended purposes and uses of the assessment system;
- Claims about student learning that would have to be supported by the assessment system;
- The use of evidence-centered design (ECD) to guide the design of the assessments and the development of assessment tasks and items;
- The establishment of achievement standards for the Consortium.

Chapter 3 provides an in-depth discussion of how the Consortium implemented those foundational decisions to build the Smarter Balanced comprehensive assessment system. The chapter includes a description of key features of the assessment infrastructure such as the item pool, the online assessment and reporting system, and the steps taken to ensure quality data. Most important, Chapter 3 also describes the Consortium's commitment to providing all students with full access to the Smarter Balanced assessment system. The chapter provides an overview of the Consortium's policy on accessibility as well as a summary of the types of supports, accommodations, and universal tools that the Consortium provided to support the diverse group of students who participated in the field test.

Chapter 4 describes the population served by the Consortium. It examines the number and types of students who participated in one of the largest field tests ever administered in the United States. In particular, this chapter examines the diversity of the field test sample and how this supports the generalizability of the test results.

Chapter 5 describes the various levels of outreach and communication that the Consortium employed in order to reach our target audiences.

Chapter 6 summarizes the Consortium's plan for sustaining the work started in 2010 as well as plans for the continuing enhancement of the comprehensive assessment system.

Chapter 7 discusses the interoperability of the Consortium's technology-based assessments. This chapter examines how the Consortium maximized the interoperability of assessments across technology platforms and the ability for member states to switch assessments from one technology platform to another.

Chapter 8 summarizes the Consortium's approach to data privacy and security. In particular, this chapter discusses the Consortium's commitment to protecting students' personally identifiable information.

Chapter 9 presents the Consortium's final fiscal report.

CHAPTER 2.
THE FOUNDATIONS OF THE SMARTER BALANCED ASSESSMENT SYSTEM

DEFINING THE PURPOSES OF THE COMPREHENSIVE ASSESSMENT SYSTEM

The test scores from the summative assessments are being used for a variety of purposes, such as monitoring student progress toward college and career readiness, monitoring student proficiencies for federal accountability purposes, and measuring achievement on the CCSS. The interim assessment and digital library components of the assessment system were designed for the primary purpose of supporting instruction that will lead to improved student performance and, ultimately, to students meeting the high expectations of the summative assessment. In this chapter, we share the intended purposes of the summative assessment and explore how the assessment system was designed from the ground up to support these purposes.[9]

Purposes of the Summative Assessment

The Consortium's summative assessments are aligned to the CCSS in grades 3 through 8 and high school. Students in high school will take the summative test in grade 11. These assessments were built to support seven distinct, yet related, purposes:

1. The results from the summative assessments will provide valid, reliable, and fair information about students' ELA/L and mathematics achievement with respect to those CCSS measured by the ELA/L and mathematics summative assessments;
2. The results from the grades 3 through 8 summative assessments will provide valid, reliable, and fair information about whether students have demonstrated sufficient academic proficiency in ELA/L and mathematics to be on track for achieving college readiness;
3. The results from the high school summative assessments will provide valid, reliable, and fair information about whether students have demonstrated sufficient academic proficiency in ELA/L and mathematics to be ready to take credit-bearing college courses;
4. The results from the summative assessment will provide valid, reliable, and fair information about students' annual progress toward college and career readiness in ELA/L and mathematics from year to year;
5. The results from the summative assessments will provide valid, reliable, and fair information on how instruction can be improved at the classroom, school, district, and state levels;
6. The results from the summative assessments will provide valid, reliable, and fair information about students' ELA/L and mathematics proficiencies for federal accountability purposes and potentially for state and local accountability systems;
7. The results from the summative assessments will provide valid, reliable, and fair information about students' achievement in ELA/L and mathematics that is equitable for all students and all subgroups.

[9] Sireci, S. (2012). Smarter Balanced Assessment Consortium: Comprehensive Research Agenda, p.12. http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/08/Smarter-Balanced-Research-Agenda_Recommendations-2012-12-31.pdf

Purposes of the Interim Assessment

The Consortium's interim assessments are aligned to the CCSS in grades 3 through 8 and high school. These assessments were built to support four distinct, yet related, purposes:[10]

1. Provide valid, reliable, and fair information about students' progress toward mastery of the skills measured in ELA/L and mathematics by the summative assessments;
2. Provide valid, reliable, and fair information about students' performance at the content cluster level, so that teachers and administrators can track student progress throughout the year and adjust instruction accordingly;
3. Provide valid, reliable, and fair information about individual and group (e.g., school, district) performance at the claim level in ELA/L and mathematics, to determine whether teaching and learning are on target;
4. Provide valid, reliable, and fair information about student progress toward the mastery of skills measured in ELA/L and mathematics across all students and subgroups of students.

To ensure that the assessment results were appropriate for the intended purposes, the first step was to define the construct; in other words, what was meant by ELA/L and by mathematics. To do this, the Consortium used an evidence-centered design to build their assessment system.

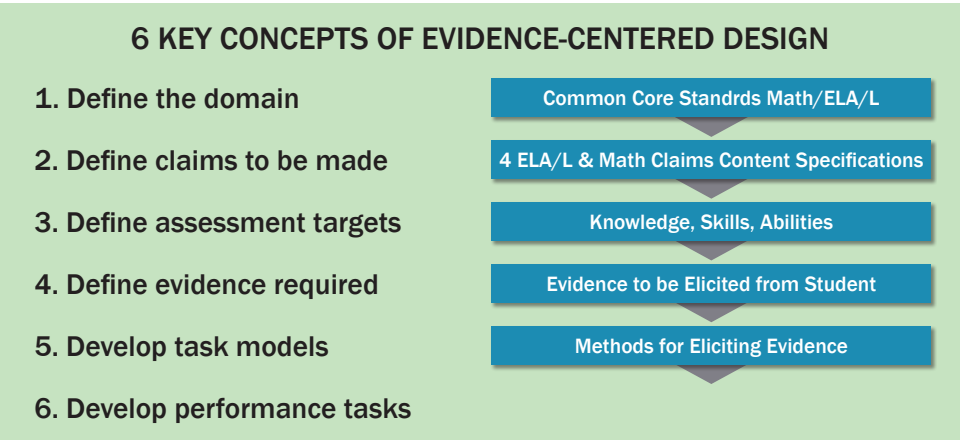
Adopting the Principles of Evidence-Centered Design

An assessment should be built on a foundation that articulates the content and cognitive processes being measured, how they are measured, and the relative importance of what is being measured. To establish that foundation, the Consortium discarded traditional approaches to assessment design and item development. Instead, the Consortium employed an evidence-centered design (ECD) approach – a modern approach to designing items and tasks. ECD has been used successfully in the design and development of educational assessments, but has not had widespread use in large-scale state assessment because it is labor-intensive and requires complex decisions. Relying on the breadth and depth of educators from across the Consortium states, the modern design of ECD was used.

As an initial step in the process, the Consortium identified six key concepts of ECD and applied them to the work of designing and developing the Smarter Balanced assessment system. These six, sequential steps in ECD and their relationship to Consortium tasks are shown in Figure 2.1.

The Smarter Balanced assessment is built on a foundation that articulates the content and cognitive processes being measured, how they are measured, and the relative importance of what is being measured.

FIGURE 2.1 KEY CONCEPTS OF EVIDENCE-CENTERED DESIGN



[10] Sireci, S. (2012). Smarter Balanced Assessment Consortium: Comprehensive Research Agenda, p.12. http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/08/Smarter-Balanced-Research-Agenda_Recommendations-2012-12-31.pdf

This chapter describes how the Consortium implemented the first five of these ECD steps (Step 6 is described in Chapter 3). In addition, this chapter discusses the Consortium's development of achievement-level descriptors and test blueprints.

DEFINING THE DOMAIN

The bedrock of the Consortium's assessment is the CCSS. All of the work started with the understanding that all member states would adopt the CCSS and that schools and districts within those states would align their curriculum and instruction to the CCSS. From the beginning, the Consortium intended to measure student achievement in terms of the CCSS.

The CCSS framed the knowledge and content students are expected to learn and demonstrate as they progress through school. The mathematics CCSS are based on research that detailed how "students' mathematical knowledge, skill, and understanding develop over time."[11] The ELA/L CCSS created anchor standards that are the skills that high school graduates should have in order to be college and career ready.[12] For each anchor standard, the ELA/L CCSS provides grade-specific standards that articulate what students are expected to learn in specific grades.

So, the CCSS delineates the knowledge and skills that students need to demonstrate in order to be college and career ready. In addition, the CCSS articulates the knowledge and skills students need to demonstrate in each grade to show that they are on track to achieving college and career readiness. The CCSS was the starting document for developing all of the Consortium assessments. It was a necessary starting place so that the results of the summative assessments would provide valid information with respect to the CCSS.

DEFINING CLAIMS, EVIDENCE, AND ASSESSMENT TARGETS

Although the CCSS define the domain of the assessment system at a high level, additional steps were necessary to apply the remaining steps in the ECD process. These steps include converting the high-level CCSS into specific content specifications and establishing claims, evidence, and assessment targets.

Content Specifications

Using the CCSS, the Consortium assembled a team of experts in the fields of mathematics, mathematics education, ELA/L, reading, elementary education, English language learners, students with disabilities, and assessment along with the primary authors of the CCSS to write content specifications for ELA/L and mathematics. This team worked together to create an initial draft of the content specifications in summer 2011.

In the content specifications document, the Consortium established the assessment claims discussed above along with the evidence that the Consortium would need to collect in order to support each claim. This evidence was collected through different types of test items and tasks, and this document delineated how the Consortium should provide evidence for each claim. All of the Consortium's content-related work that followed was built from the ideas first articulated in the content specifications.

The Consortium's Technical Advisory Committee and Consortium staff reviewed the initial draft of the content specifications. A revised version went through two rounds of public review that lasted nearly 30 days during which more than 200 individuals and organizations provided feedback on the content specifications. Using the public's feedback, the documents were revised and the claims were adopted by the Governing States in spring 2012.

More than 200 individuals and organizations provided feedback on the content specifications.

[11] http://www.corestandards.org/Math/

[12] http://www.teachingthecore.com/ccr-anchor-standards/

Claims

The interim and summative assessments serve as measures of how well students are learning the CCSS. These assessments are designed to measure particular "claims." A claim tells stakeholders what should be concluded from the test score. In both ELA/L and mathematics, the Consortium provides an overall claim for what the test score means as well as specific claims related to the content area. Each student's test score will show how well they have done on a particular claim. Table 2.1 below shows the claims that are being measured by the interim and summative assessments for both ELA/L and mathematics.

TABLE 2.1
SMARTER BALANCED ASSESSMENT CLAIMS

	ENGLISH LANGUAGE ARTS AND LITERACY	MATHEMATICS
Overall, Grades 3-8	Students can demonstrate progress toward college and career readiness in English language arts and literacy.	Students can demonstrate progress toward college and career readiness in mathematics.
Overall, Grade 11	Students can demonstrate college and career readiness in English language arts and literacy.	Students can demonstrate college and career readiness in mathematics.
Claim 1	**Reading:** Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts.	Concepts & Procedures: Students can explain and apply mathematics concepts and interpret and carry out mathematics procedures with precision and fluency.
Claim 2	**Writing:** Students can produce effective and well-grounded writing for a range of purposes and audiences.	Problem Solving:* Students can solve a range of complex well-posed problems in pure and applied mathematics, making productive use of knowledge and problem solving strategies.
Claim 3	**Speaking**[13] **and Listening:** Students can employ effective speaking and listening skills for a range of purposes and audiences.	Communicating Reasoning: Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others.
Claim 4	**Research/Inquiry:** Students can engage in research and inquiry to investigate topics, and to analyze, integrate, and present information.	Modeling and Data Analysis:* Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems.

**Claims 2 and 4 in mathematics will be reported jointly*

[13] Speaking is not being assessed currently as part of the summative assessment.

These claims were articulated prior to any other test development. Therefore, all work associated with developing the test, such as item writing and form construction, occurred after the Consortium specified its intended measures. Without this specification, it would have been difficult to build an aligned testing system.

Evidence

The content specifications described the sufficient relevant evidence necessary to draw inferences or conclusions about student attainment of each claim. The content specifications delineate the item types that could be used to provide evidence for each claim. For example, selected response, short constructed response, and even extended response items might be used to address Claim 1 in mathematics. By delineating the types of items and tasks needed for each claim *prior* to writing items, the content specifications laid the groundwork for ensuring that the Consortium's pool of items and tasks had the breadth and depth to measure and make interpretations about year-to-year student progress.

Accessibility. As part of the Consortium's approach to universal design, accessibility was built into the content specifications as a foundational document. The content specifications address accessibility issues that may be faced by students with disabilities or by English language learners (ELLs) when measuring each claim. For example, an important aspect of all mathematics claims is the ability for students to communicate *why* or *how* given procedures work. The authors of the content specifications suggested several ways to maximize access for students with disabilities, including allowing students with disabilities the use of scribes to express their views. The Consortium allows students with a documented need the ability to dictate their responses to a trained and qualified scribe for ELA/L non-writing items and for mathematics items.

As part of the Consortium's approach to universal design, accessibility was built into the content specifications as a foundational document. The content specifications address accessibility issues that may be faced by students with disabilities or by English language learners (ELLs) when measuring each claim.

Assessment Targets

In addition, the content specifications identify the grade specific aspects of each claim. Even though the broad content-related claims are constant across the grades, the types of fine-grained knowledge and skills being taught and measured will vary by grade level. This document specifies the knowledge and skills via assessment targets that are measured at each grade. The assessment targets define the range of content and the levels of cognitive skills assessed, and they describe the expectations of what will be assessed by the test questions and tasks within each claim. For example, when assessing ELA/L Claim 1 equal emphasis is placed on literary and informational texts in grades 3 through 5, slightly more emphasis is placed on informational text (55%) in grades 6 through 8, and greater emphasis is placed on a range of informational text (70%) in high school.

CREATING ACHIEVEMENT LEVEL DESCRIPTORS

Although the CCSS define the learning progressions and describe the discrete content and skills to be demonstrated at each grade level, the CCSS do not delineate how much students need to know in order to be considered ready for college and career. It is left to the Consortium to use available research and ultimately to define the knowledge, skills, and abilities students need to demonstrate to be considered ready for college and career. To do this, it was necessary to develop Achievement Level Descriptors (ALDs).[14, 15] The ALDs describe the knowledge, skills, and abilities that students should demonstrate on a standardized test in terms of levels or categories of performance.

[14] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/11/Smarter-Balanced-ELA-Literacy-ALDs.pdf

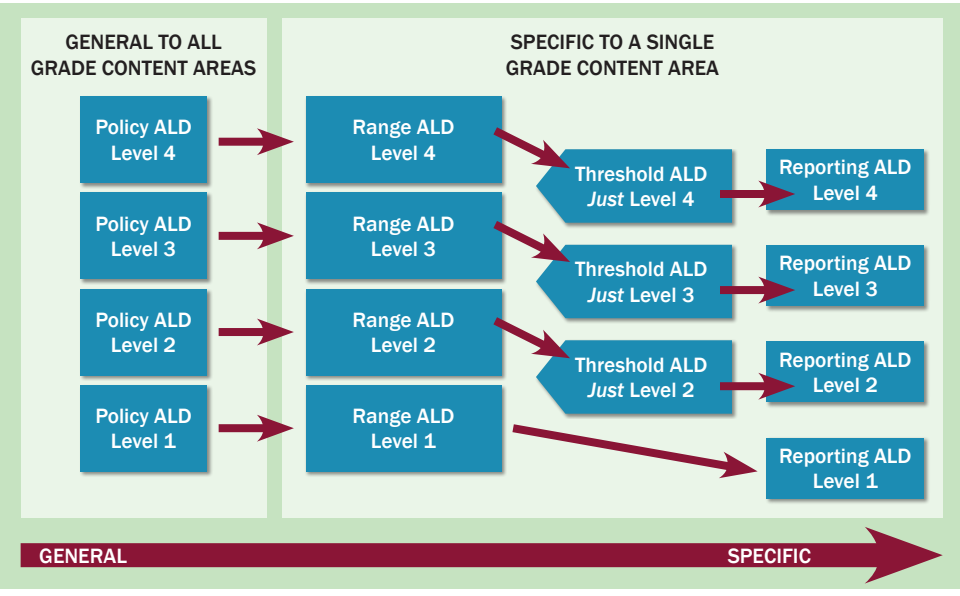
[15] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/11/Smarter-Balanced-Math-ALDs.pdf

The Consortium's ALDs defined four levels of performance on each of the claims: Level 1, Level 2, Level 3, and Level 4. The Consortium developed a system of interrelated ALDs that address distinct, but related purposes. These ALDs are as follows:

- Policy ALDs provide general descriptors that broadly articulate the goals for student performance and the general level of rigor expected in the final achievement standards. These descriptors set the tone for subsequent descriptors. Policy makers most often use these descriptors;
- Range ALDs are grade- and content-specific descriptions of student knowledge and skills that are used by test developers to guide item writing; these ALDs describe the cognitive and content rigor that is expected of all students in a particular achievement level, from the student who has just entered the achievement level to the student at the top of the achievement level;
- Threshold ALDs are derived from Range ALDs and describe the knowledge and skill of a student who can be considered to barely meet the bar for a given performance category (e.g., a barely proficient student). These ALDs were used at the achievement-level setting (see Chapter 3);
- Reporting ALDs are the ALDs most often encountered by stakeholders. These ALDs provide guidance on how to interpret student performance at a given level (e.g., Levels 1, 2, 3, and 4) on the test. Reporting ALDs were developed following the achievement-level setting held in October 2014.

Figure 2.2 presents an overview of the system of ALDs, indicating how the four types of ALDs are linked and how each ALD informs the development of the next. As depicted in the figure, the policy ALDs are overarching statements that encompass all grade content areas. The policy ALDs state educational goals regarding what students within the performance levels are expected to do. From these policy ALDs, the range ALDs are developed to incorporate grade- and content-specific information about the knowledge, skills, and processes that students are expected to demonstrate along the proficiency continuum. Range ALDs describe the types of evidence that items within an achievement level should elicit to support the policy claims, and for this reason, they support item writing. Range ALDs are built using Smarter Balanced content specifications and the CCSS. The threshold ALDs are the preliminary conceptualization of the minimum evidence a student should demonstrate from the range ALDs to meet an achievement level expectation. The reporting ALDs are the final indication of the threshold ALDs based upon the final approved cut scores.

FIGURE 2.2 OVERVIEW OF THE ALD SYSTEM



Developing the Consortium's Achievement Level Descriptors

Elementary, middle, and high school educators were chosen from all Consortium states to represent rural, suburban, and urban districts that had varying percentages of students receiving free and reduced lunch. The representatives of higher education were chosen to represent both two- and four-year colleges. These panelists represented all of the (then 21) Governing States, and all Governing States had at least one K-12 representative and one higher education representative attending the workshop.

During this workshop, the panelists created the first three types of ALDs for each claim (policy, range, and threshold). For the grade 11 claims, high school teachers and college faculty worked together to articulate the knowledge, skills, and processes that students would need to demonstrate in order to be considered ready for college and career. For each of the claims in grades 3 – 8, experienced educators in the specific and adjacent grades created the ALDs.

In addition to the ALDs, the grade 11 panelists also reviewed and revised the Consortium's operational definition of college content readiness and grade 11 policy framework (see the next section). Together, the operational definition and policy framework describe how colleges, universities, and schools are asked to interpret the results of the grade 11 assessment.

Following the workshop, the Consortium held a series of reviews of the ALDs, the operational definition of college content readiness, and the grade 11 policy framework. Members of the Consortium's Test Development and Validation work group, select members from the workshop, and Smarter Balanced staff participated in the first review and revision of the ALDs. The second review was open to the public, and 350 people representing K-12 and higher education contributed to the comments. The third review period was open only to the Governing States' K-12 and higher education leads as well as members of the Development and Validation work group. The members of the Executive Committee participated in the final review.

The ALDs were adopted by the Governing States at the March 2013 Collaboration Conference. The operational definition of college content readiness and the grade 11 policy framework were approved by the Governing States in April 2013.

Policy ALDs

The overall claim was delineated into the four levels as shown in Table 2.2 below. (The content claims within ELA/L and mathematics will be reported in relationship to the overall claim.)

TABLE 2.2
POLICY ALDS FOR GRADES 3 – 8 AND GRADE 11

HIGH SCHOOL	GRADES 6–8	GRADES 3–5
LEVEL 4 The student has exceeded the achievement standard and demonstrates the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in entry-level credit-bearing college coursework after high school.	**LEVEL 4** The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in entry-level credit-bearing college coursework after high school.	**LEVEL 4** The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in future coursework.
LEVEL 3 The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in entry-level credit-bearing college coursework after completing high school coursework.	**LEVEL 3** The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in entry-level credit-bearing college coursework after high school.	**LEVEL 3** The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in future coursework.
LEVEL 2 The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in entry-level credit-bearing college coursework after high school.	**LEVEL 2** The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in entry-level credit-bearing college coursework after high school.	**LEVEL 2** The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in future coursework.
LEVEL 1 The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in entry-level credit-bearing college coursework after high school.	**LEVEL 1** The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in entry-level credit-bearing college coursework after high school.	**LEVEL 1** The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy (mathematics) needed for likely success in future coursework.

Range and Threshold ALDs

Range ALDs were created for each assessment target, and Threshold ALDs were created for each content category associated with the specific claims. To create the Range and Threshold ALDs, the panelists worked from an abbreviated version of the content specifications in which the assessment targets were laid out side by side with the related CCSS. This method ensured a high level of fidelity to both the CCSS and to the content specifications.

Policy on College Content Readiness

The operational definition of college content readiness and the policy framework for college content readiness describe how colleges, universities, and schools should interpret student performance.

Operational Definition of College Content Readiness

The Consortium recognizes that college readiness encompasses a wide array of knowledge, skills, and dispositions, only some of which will be measured by the Smarter Balanced assessments. As a result, Smarter Balanced narrowed the focus of its "college readiness" definition to "content readiness" in the core areas of ELA/L and mathematics. Table 2.3 shows the college content readiness definition for ELA/L and mathematics.

TABLE 2.3
OPERATIONAL DEFINITION OF COLLEGE CONTENT READINESS

ENGLISH LANGUAGE ARTS/LITERACY	Students who perform at the College Content-Ready level in English language arts/literacy demonstrate reading, writing, listening, and research skills necessary for introductory courses in a variety of disciplines. They also demonstrate subject-area knowledge and skills associated with readiness for entry-level, transferable, credit-bearing English and composition courses.
MATHEMATICS	Students who perform at the College Content-Ready level in mathematics demonstrate foundational mathematical knowledge and quantitative reasoning skills necessary for introductory courses in a variety of disciplines. They also demonstrate subject-area knowledge and skills associated with readiness for entry-level transferable, credit-bearing mathematics and statistics courses.

Policy Framework
Table 2.4 shows the policy framework for college content readiness.

TABLE 2.4
POLICY FRAMEWORK FOR COLLEGE CONTENT READINESS

LEVEL	POLICY ALD	DESCRIPTION	IMPLICATIONS FOR GRADE 12	IMPLICATIONS FOR HIGH SCHOOL GRADUATES WHO IMMEDIATELY ENTER HIGHER EDUCATION
4	Student demonstrates thorough understanding of and ability to apply the knowledge and skills associated with college content-readiness.	Student is exempt from developmental course work. (K-12 and higher education officials may jointly set grade 12 requirements to maintain the exemption.)	Within each state, students may be required to satisfactorily complete grade 12 English and/or mathematics courses to retain the exemption from developmental course work (higher education and K-12 officials may jointly determine appropriate courses and performance standards). Students are encouraged to take appropriate advanced credit courses leading to college credit while still in high school.	Colleges may evaluate additional data (courses completed, grades, placement test scores, writing samples, etc.) to determine appropriate course placement at or above the initial credit-bearing level.
3	Student demonstrates adequate understanding of and ability to apply the knowledge and skills associated with college content-readiness.	Student is conditionally exempt from developmental course work, contingent on evidence of sufficient continued learning in grade 12.	Within each state, higher education and K-12 officials may jointly determine appropriate evidence of sufficient continued learning (such as courses completed, test scores, grades or portfolios). Students are encouraged to take additional 4th year courses as well as appropriate advanced credit courses leading to college credit while in high school.	For students who demonstrate evidence of sufficient continued learning in grade 12, colleges may evaluate additional data (courses completed, grades, portfolios, placement test scores, etc.) to determine appropriate course placement at or above the initial credit-bearing level. For students who fail to demonstrate evidence of sufficient continued learning in grade 12, colleges also may evaluate the same types of additional data to determine placement in developmental or credit-bearing courses.
2	Student demonstrates partial understanding of and ability to apply the knowledge and skills associated with college content-readiness.	Student needs support to meet college content-readiness standard.	States/districts/colleges may implement grade 12 transition courses or other programs for these students. States also may choose to retest these students near the conclusion of Grade 12 (scoring will occur within two weeks, allowing opportunity for colleges to use scores the following fall).	Colleges may evaluate additional data (courses completed, grades, portfolios, placement test scores, etc.) to determine placement in developmental or credit-bearing courses.
1	Student demonstrates minimal understanding of and ability to apply the knowledge and skills associated with college content-readiness.	Student needs substantial support to meet college content-readiness standard.	States/districts/colleges may offer supplemental programs for these students. States also may choose to retest these students near the conclusion of grade 12.	Colleges may evaluate additional data (courses completed, grades, portfolios, placement test scores, etc.) to determine placement in developmental or credit-bearing courses.

DESIGNING THE ASSESSMENT, TASK MODELS, AND ITEMS

To build a summative assessment that measures the intended claims, supports the intended uses, and aligns to the seven principles of high-quality assessment systems (see Chapter 1 for a discussion of the seven principles), the Consortium's test development cycle was iterative, involving experts from various education-related fields, and was based on assessment-related research and best practices.

Item and Task Specifications[16]

The item specifications bridge the span from the content specifications and ALDs to the assessment itself. While content specifications established the Consortium's claims and the types of evidence that would need to be collected in order to support them, more specificity was needed in order to create items and tasks that measured the claims. Working with three vendors (ETS, Measured Progress, and CTB) with extensive experience in item writing, the Consortium's item development work group and performance task work group reviewed and approved preliminary item and task specifications for ELA/L and mathematics.

The original item and task specifications were developed in 2011. In early 2012, the Consortium held a series of showcases where they introduced the item and task specifications and collected feedback from member states. Using this feedback, the item and tasks specifications were revised during the first quarter of 2012.

Using the revised item and task specifications, a small set of items was developed and administered in fall 2012 during a small-scale trial. This provided the Consortium with their first opportunity to administer and score the new item types. During the small-scale trials, the Consortium conducted "cognitive laboratories" to better understand how students solve various types of items. A cognitive laboratory uses a think-aloud methodology in which students speak their thoughts while working on a test item. The item and task specifications were revised based on the findings of the cognitive laboratories. These revised specifications were used to develop items for the 2013 pilot test. Following rangefinding activities and analyses of the pilot test data, the item and task specifications were revised to support the creation of a robust item pool for the adaptive test.

The Consortium's item and task specifications are designed to ensure that the assessment items measure the assessment's claims. Indeed, the purpose of the item and task specifications is to define the characteristics of the items and tasks that will provide the evidence to support one or more claims. To do this, the item and task specifications delineate the types of evidence that should be elicited for each claim within a grade level. Then, they provide explicit guidance on how to write items in order to elicit the desired evidence.

The Consortium's item and task specifications are designed to ensure that the assessment items measure the assessment's claims. Indeed, the purpose of the item and task specifications is to define the characteristics of the items and tasks that will provide the evidence to support one or more claims.

In doing this, the item and task specifications provide guidance on how to realize the ideas first found in the content specifications. The item and task specifications provide guidelines on how to create the items that are specific to each assessment target and claim through the use of task models. A task model provides a description of an item/task's key features. These task models describe the knowledge, skills, and processes being measured by each of the item types aligned to particular targets. In addition, the task models sometimes provide examples of plausible distractors where applicable. Exemplar items are provided within every task model.

These task models were developed to delineate the expectations of knowledge and skill to be included on test questions in each grade. In addition, both the ELA/L and mathematics item specifications provide guidance on determining the grade

[16] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/ItemSpecifications/GeneralItemSpecifications.pdf

appropriateness of stimulus materials (the materials that a student must refer to in working on a test question). The task models also provide information on characteristics of stimuli or activities to avoid because they are not important to the knowledge, skill, or process being measured.

This is important because it underscores the Consortium's efforts to develop items that are accessible to the widest range of students possible; in other words, Consortium items are created according to the principle of universal design. As the name suggests, the concept of universal design aims to create items that accurately measure the assessment target for all students. At the same time, universal design recognizes that one solution rarely works for all students. Instead, this framework acknowledges "the need for alternatives to suit many different people."[17]

To facilitate the application of universal design principles, item writers are trained to consider the full range of students who may answer a test question. A simple example of this is the use of vocabulary that is expected to be known by all third-grade students versus only those third-grade students who play basketball. Almost all third-grade students are familiar with activities (e.g., recess) that happen during their school day, while only a subset of these students will be familiar with basketball terms like "double dribble," "layup," "zone defense," or "full-court press."

In addition to this, the item specifications discuss accessibility issues that are unique to the creation of items for a particular claim and/or assessment target. The accessibility concerns discuss the different supports that various groups of students may need to access the content of an item. By considering the possible supports that may be needed for each item, item writers are able to create items that will be accessible to almost all students.

The use of universal design principles allows the Consortium to collect evidence on the widest possible range of students. By writing items that adhere to the item and task specifications, the Consortium is assured that the assessments measure the claims and assessment targets established in the content specifications as well as the knowledge, skills, and processes found in the CCSS for all students for whom the assessment is appropriate.

Test blueprints are the final foundational document of the Consortium's summative assessment. The test blueprint, in part, guides the construction of each student's test.

CREATING THE TEST BLUEPRINT[18]

Test blueprints are the final foundational document of the Consortium's summative assessment. The test blueprint, in part, guides the construction of each student's test. The use of an adaptive test means that students will encounter different sets of items from their peers. Each student will take a unique test form that provides a reliable estimate of the student's ability. The test blueprints reflect the depth and breadth of the CCSS and include critical information about the assessment targets, the depth of knowledge associated with each assessment target, the number of machine scored and short text items, and the total number of items per content category. The blueprints specify the type and range of content and items that every student encounters on the assessments, thereby ensuring that the Consortium is measuring the same construct for all students.

Test blueprints guide the selection of test forms; thus, a test blueprint for an adaptive test necessarily differs from a test blueprint used for a traditional paper and pencil test. When a traditional paper and pencil test is used, a single form or a set of forms is selected. These forms can be reviewed and refined by experts in content and test design prior to test administration. With an adaptive test, thousands of forms will be created prior to the testing event. Like test blueprints for paper and pencil tests, blueprints for adaptive tests must be sufficiently detailed so that unique test forms

[17] Rose, D. & Meyer, A. (2000). Universal design for learning, associate editor column. Journal of Special Education Technology 15(1):66-67

[18] http://www.smarterbalanced.org/smarter-balanced-assessments/

measure the same construct; however, adaptive blueprints must also provide adequate flexibility so that thousands of forms may be constructed on the fly.

The Consortium's blueprints were built from the base of the CCSS and content specifications. The content specifications first communicated the content-related claims and the assessment targets being measured at each grade. The test blueprints operationalized these ideas by specifying a range of numbers of items that should be used to measure each claim, and within each claim, the range of items that should be used to measure each assessment target. In addition, the blueprint specifies a range of items at each depth of knowledge level that should be presented to each student in each claim and assessment target to assure that the skills (rather than just the content categories) presented to students are similar. As an example, Figure 2.2 shows a portion of the grade 3 mathematics test blueprint that specifies the number of items and tasks that should be used to measure each assessment target.

By specifying the approximate number of items and the expected depth of knowledge of the items, the test blueprints provide guidance to test developers and educators on the relative importance and rigor of the claims and assessment targets in each grade. In addition, the test blueprints clarify how the claims and assessment targets will be measured (through both the computer-adaptive test (CAT) and performance task portions of the test).

The Consortium's blueprints were created through a joint effort of the assessment, test design, item development, performance task, and psychometrics and validation work groups, along with the Consortium's lead psychometrician and content (i.e., ELA/L and mathematics) directors. Altogether, 20 states were represented in these work groups. The work groups contributed to the content and format of the blueprints. The Governing States unanimously voted to accept the draft version of the blueprints on November 20, 2012.

The original draft blueprints were updated following the spring 2013 pilot test, and they were again updated following the spring 2014 field test. The revisions to the blueprints were minor, and the major ideas first articulated in the 2012 blueprints remained intact. The blueprints used in summative tests in spring 2015 are the February 2015 revisions which are also largely consistent with the original ideas.

FIGURE 2.3 PORTION OF A TEST BLUEPRINT

TARGET SAMPLING MATHEMATICS GRADE 3

Claim	Content Category	Assessment Targets	DOK	Items		Total Items
				CAT	PT	
1. Concepts and Procedures	Priority Cluster	B. Understand properties of multiplication and the relationship between multiplication and division.	1	5-6	0	17-20
		C. Multiply and divide within 100.	1			
		I. Geometric measurement: understand concepts of area and relate area to multiplication and to addition.	1, 2			
		G. Solve problems involving measurement and estimation of intervals of time, liquid volumes, and masses of objects.	1, 2			
		D. Solve problems involving the four operations, and identify and explain patterns in arithmetic.	2	5-6		
		F. Develop understanding of fractions as numbers.	1, 2			
		A. Represent and solve problems involving multiplication and division.	1, 2	2-3		
	Supporting Cluster	E. Use place value understanding and properties of operations to perform multi-digit arithmetic.	1	3-4		
		J. Geometric measurement: recognize perimeter as an attribute of plane figures and distinguish between linear and area measures.	1			
		K. Reason with shapes and their attributes.	1, 2			
		H. Represent and interpret data.	2, 3	1		

CREATING THE DIGITAL LIBRARY

The Digital Library is a critical component of the Consortium's system of assessments. It is an online collection of instructional and professional learning resources contributed by educators for educators. These resources are aligned with the intent of the CCSS and help educators implement the formative assessment process to improve teaching and learning.

Purposes of the Digital Library

The purpose of the Consortium's Digital Library is to provide tools and resources that:

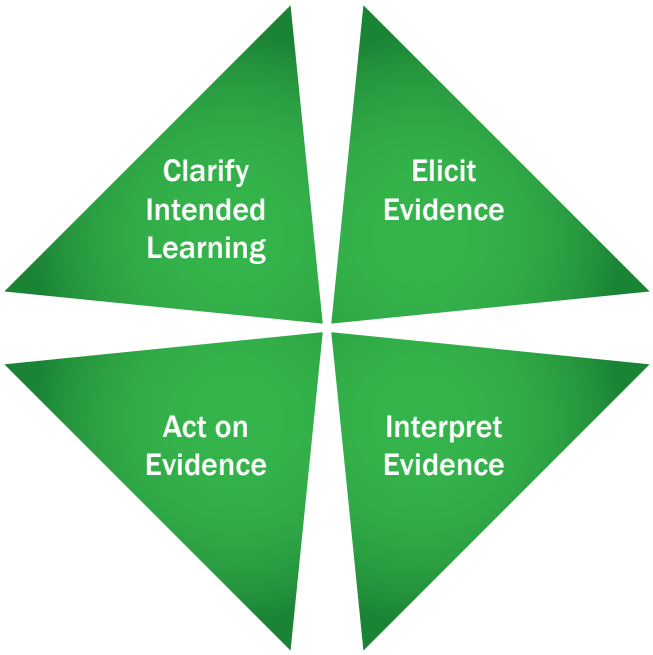
1. Improve teaching and learning;
2. Monitor student progress throughout the school year;
3. Help teachers and other educators align instruction, curricula, and assessment;
4. Help teachers and other educators use the summative and interim assessments to improve instruction at the individual student and classroom levels;
5. Illustrate how teachers and other educators can use assessment data to engage students in monitoring their own learning.

Formative Assessment

Formative assessment is a deliberate process used by teachers and students during instruction that provides actionable feedback used to adjust ongoing teaching and learning strategies to improve students' attainment of curricular learning targets/goals. The four attributes of the formative assessment process, represented in Figure 2.4, are:

- Clarify intended learning;
- Elicit evidence;
- Interpret evidence;
- Act on evidence.

FIGURE 2.4 ATTRIBUTES OF THE FORMATIVE ASSESSMENT PROCESS



There are two important components of clarifying intended learning: learning goals and success criteria. Learning goals state what students will know by the end of the lesson, and success criteria define the evidence that teachers and students will use to determine how students are progressing toward the learning goals. Eliciting evidence provides information that teachers and students need to determine where students are in their progress toward the intended learning goals. Teachers and students interpret the evidence that they collect to determine where students are in relation to the learning goals and success criteria, identifying what students do or do not understand. Acting on evidence occurs when teachers and students use actionable feedback to determine the next steps to move learning forward. Each resource in the Digital Library addresses one or more of the four attributes of the formative assessment.

Creating the Materials for the Digital Library

The Consortium worked with an expert panel, member states, K-12 educators, and higher education faculty to create the materials for the digital library. The expert panel helped the Consortium create a Quality Criteria Guide[19] for the professional learning and instructional resources that populate the Digital Library. This guide was designed to assist educators and professionals when submitting and reviewing formative resources for the Digital Library.

Each member state formed State Leadership Teams (SLTs) of 8-12 members consisting of K-12 educators and higher education faculty. The Consortium provided the SLTs with extensive training on the digital library and on the quality criteria. The SLTs were responsible for recruiting and training educators for the State Networks of Educators (SNEs) that were formed within each member state.

The SNEs involved nearly 1,200 K-12 educators and higher education faculty who were trained on the digital library and on the quality criteria. Only educators who are part of SNEs may submit materials to the Digital Library. Once an educator submits materials to the Digital Library, they will go through three levels of review before those materials can be used. Materials that are not accepted are returned to the submitter with feedback and suggestions for improvement.

[19] https://www.k12.wa.us/SMARTER/pubdocs/quality-criteria-guide.pdf

CHAPTER 3.
BUILDING THE ASSESSMENT SYSTEM

The Consortium spent almost four years carefully constructing the content and content delivery elements of the system for the interim and summative assessments. The Consortium's approach to the delivery of content has three interlocking components: the item pool, the online assessment delivery system, and the reporting system.

In creating the assessment system, we built an item pool designed to assess the breadth of the CCSS for all students for whom the assessment is appropriate, including students with disabilities, English language learners, and English language learners with disabilities. Educators from member states created many of these items. The items will be delivered through an online, open-source computer-adaptive testing system that accurately and efficiently measures student achievement and growth. Educators will receive standardized and customizable reports that are cost-effective, timely, and useful. These reports will allow teachers, students, and parents to track progress toward academic readiness for postsecondary education, whether in the form of college or career training.

THE ITEM POOL

An item pool refers to a collection of test questions (known as items) measuring the same content area (e.g., mathematics) within the same grade. (As will be explained below, the use of off-grade-level items is allowed in some instances.) The quality of the items is a primary concern when building an item pool. The Consortium took multiple steps to ensure the quality of the items in our item pool. Building on the ongoing process of developing item/task specifications and test blueprints described in the previous chapter, we used an iterative process for creating and revising each item as well as the collection of items. The Consortium tested items and refined its approach to item development through three steps: small-scale tryouts, a large pilot test, and the largest ever field test of a K-12 assessment. The pilot and field test included over 700,000 and 4.2 million students, respectively. During each phase, the Consortium used cognitive laboratories to understand the strategies that students used to respond to the items. By incorporating this tiered and iterative approach, the item and task specifications that guided the development of the final operational pool were improved based on the lessons learned during these important tryouts.

The pilot and field test included over 700,000 and 4.2 million students, respectively.

Teachers were integrally involved in the creation of the item pool from beginning to end. Some participated in the processes described in the previous chapter and others developed many of our items through a rigorous item writing process. Still other educators reviewed the items for accuracy and appropriateness of the content knowledge and skill level required to respond to the items. Another group reviewed items for potential issues of bias in favor of or against any demographic group of students, and accessibility for students with disabilities, English language learners, and English language learners with disabilities. Teams of educators reviewed items for content, bias, and accessibility prior to administration to any students. Following the pilot and field test administrations, items were again reviewed if pilot or field test data indicated a potential problem. Finally, teachers participated in scoring constructed-response items to ensure that the items could be properly scored given their scoring rubrics.

In this section, we will examine the primary role that educators played in creating the field-test item pool by writing, reviewing, and scoring items. This section will end by examining the current composition of the item pool.

Item Writing

The job of writing all of the items and performance tasks was no small undertaking, and the Consortium worked with educators throughout the test development cycle to write items. Prior to the spring 2013 pilot test, the Consortium engaged 136 educators in K-12 and higher education from 19 member states to write items. Prior to the spring 2014 field test, 184 educators in K-12 and higher education from 16 member states participated in item writing. All K-12 participants:

- Were certified/licensed to teach ELA/L and/or mathematics in a K-12 public school;
- Were currently teaching in a public school within a Smarter Balanced Governing State;
- Had taught ELA/L and/or mathematics in grades 3 through 8 and/or high school within the past three years (second-grade teachers were also recruited to participate in the development of grade 3 items and/or tasks);
- Had previously reviewed part or all of the CCSS for the content area for which they were writing items and/or performance tasks;
- Submitted a statement of interest that described why they wanted to develop Smarter Balanced items and/or performance tasks as well as their qualifications for doing so;
- Completed training and achieved qualifications through the certification process.

184 educators in K-12 and higher education from 16 member states participated in item writing.

Qualifications for Higher Education Faculty included:

- Current employment with, or recent retirement from, a college or university located within a Smarter Balanced Governing State;
- Having taught developmental and/or entry-level courses in English, composition, mathematics, statistics or a related discipline within the last 3 years;
- Having previously reviewed part or all of the CCSS for the content area in which they are interested in writing items and/or performance tasks;
- Completing training and achieving qualifications through the certification process.

The selected educators were trained on the Consortium's content specifications and the item and task specifications as well as in using the system in which the items were developed. In addition, professional item writers held regular meetings to provide direction and feedback to the educators. Educators, state partners, and assessment vendors developed the items in the Consortium's item pool.

Training

Educators participated in a series of facilitated, online webinars in order to qualify as item writers. To facilitate participation, the Consortium scheduled multiple sessions in different time zones, including evening sessions. In addition to the facilitated sessions, the Consortium provided training modules that offered background on the Consortium, assessment design principles, and detailed information about item and performance task development. All modules were available in three formats: a PowerPoint presentation with notes, a streaming online presentation with narration, and a downloadable audio/video presentation.

The item writers were specifically trained on the Consortium's content and item specifications, stimulus specifications,[20] sensitivity and bias guidelines,[21] and general accessibility guidelines.[22] Training on these specifications and guidelines helped ensure that item writers appealed to the widest possible range of students to demonstrate their knowledge, skills, and abilities in regard to the content. This meant that item writers needed to understand the content for which they were writing items as well as accessibility and sensitivity issues that might hinder students' ability to answer. Item writers were also trained to be aware of issues that might unintentionally bias an item for or against a particular group.

Educator Participation

Consistent with the Consortium process, educators were the primary developers of items. The active involvement of educators was critical to the success of the item writing activities. Educators engage with students on a daily basis, and they understand the ways in which students can demonstrate their knowledge. Their involvement in item writing helped ensure that the assessment system is accurate and efficient, and provides valid evidence of student learning.

State-Managed Item Development

The Consortium invited member states to participate in a separate effort to write items. This voluntary effort, known as State-Managed Item Development, was conducted to build the capacity of states to write items and to support the overall sustainability of the Consortium. To this end, six states (HI, ID, MI, WA, WV, and WY) participated in the state-managed field test item development opportunity. During this opportunity, educators within the six states developed approximately 3,100 items in mathematics and ELA/L across grades 3 through 8 and high school. Many of these items were field tested during the operational test in spring 2015.

Item Reviews

Once items were written, groups of educators reviewed items prior to their pilot test administration in spring 2013 and their field test administration in spring 2014. Items that survived the pilot test were again reviewed prior to their use in the spring 2014 field test.

Accessibility, Bias/Sensitivity, and Content Reviews

Panels of educators reviewed all items, performance tasks, and item stimuli for accessibility, bias/sensitivity, and content. (Item stimuli refer to the reading passages used on the ELA/L assessments or the figures and graphics used on the mathematics assessments.) Prior to the spring 2013 field test, 122 ELA/L educators and 106 mathematics educators reviewed items and performance tasks for accessibility, bias/sensitivity, or content, and 60 educators reviewed the ELA/L stimuli. Prior to the spring 2014 field test, 107 ELA/L educators and 157 mathematics educators from 14 states reviewed items and performance, and 95 educators from 13 states reviewed the ELA/L stimuli.

The educator qualifications for the accessibility, bias/sensitivity, and content reviews were the same as the educator qualifications for item writing except that participants were not required to submit a statement of interest. In addition, it was preferred (but not required) that educators have previous experience reviewing items, tasks, and/or stimuli.

[20] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/EnglishLanguageArtsLiteracy/ELAStimulusSpecifications.pdf

[21] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/Guidelines/BiasandSensitivity/BiasandSensitivityGuidelines.pdf

[22] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/Guidelines/AccessibilityandAccommodations/GeneralAccessibilityGuidelines.pdf

End of Grant Report

During the accessibility reviews, panelists identified issues that may negatively affect a student's ability to access stimuli, items, or performance tasks, or to elicit valid evidence about an assessment target. During the bias and sensitivity review, panelists identified content in stimuli, items, or performance tasks that may negatively affect a student's ability to produce a correct response because of his or her background. The content review focused on developmental appropriateness and alignment of stimuli, items, and tasks to the content specifications and appropriate depths of knowledge. Panelists in the content review also checked the accuracy of the content, answer keys, and scoring materials. Items flagged for accessibility, bias/sensitivity, and/or content concerns were either revised to address the issues identified by the panelists or removed from the item pool.

Data-Related Reviews

The items developed for the item pool were administered during the spring 2013 and spring 2014 pilot tests, and the pilot test data from both administrations were analyzed to examine the statistical quality of the items in the pool. The Consortium established statistical criteria to flag items for possible defects in quality due to content, bias, or accessibility. For example, content-related criteria flagged items for further review if they were extremely difficult or extremely easy. Accessibility-related criteria flagged items that were differentially more difficult for students with disabilities compared to students without disabilities.

Following the spring 2013 pilot, 40 educators participated in the item data review and examined the items for possible content-related issues or accessibility-related issues. Following the spring 2014 pilot, 57 ELA/L educators from 16 states and 30 mathematics educators from 12 states participated in item data review, examining the items for possible content-related issues or accessibility-related issues. At least two educators reviewed each item. These educators were trained via webinars on the flagging criteria and on how to evaluate flagged items. These educators made recommendations on whether to accept the item with no change, revise and re-field test the item, or reject the item from the pool. McGraw-Hill CTB content experts reviewed all items where the reviewers disagreed. In addition, McGraw-Hill CTB content experts and psychometricians reviewed and provided recommendations for all items where both reviewers recommended accepting the item. In each situation, the content expert provided the Consortium with a final recommendation for the item.

The educator qualifications for the item data reviews were the same as the educator qualifications for item writing except that participants were not required to submit a statement of interest.

Item Scoring

For those items requiring human scoring following the spring 2013 pilot, the Consortium engaged 102 participants from 20 states in range finding activities. After the spring 2014 pilot, 104 educators participated in range finding. Range finding improves the consistency and validity of scoring for the assessment. During range finding, the educators focused on the performance tasks for ELA/L and mathematics. In mathematics, educators also reviewed constructed response items for grades 7, 8, and high school. During range finding, the participants reviewed student responses against item rubrics, validated the rubrics' effectiveness, and selected anchor papers to be used by professional scorers during the main scoring event.

The educator qualifications for range finding were the same as the educator qualifications for item writing, except that participants were not required to submit a statement of interest. In addition, it was preferred (but not required) that educators had previous range finding experience.

Composition of the Item Pool[23]

The Consortium developed many different types of items beyond the traditional multiple-choice item. This was done to measure the claims and assessment targets with varying degrees of complexity by allowing students to construct responses rather than simply recognize a correct response. These different item types are listed in Table 3.1 below.

TABLE 3.1
ITEM TYPES FOUND IN THE CONSORTIUM'S ITEM POOL.

ITEM TYPES	ELA/L	MATHEMATICS
Multiple Choice, Single Correct Response	X	X
Multiple Choice, Multiple Correct Response	X	X
Two-part Multiple Choice, with Evidence Responses	X	
Matching Tables	X	X
Hot Text	X	
Drag and Drop		X
Short Text	X	X
Essay	X	
Hot Spot		X
Fill-in Tables		X
Graphing		X
Equation/Numeric		X

Each grade's item pool for the Consortium's test was necessarily large to support the summative and interim assessments[24] being delivered via a computer using adaptive test-delivery technology, commonly called a computer adaptive test or CAT. Unlike a traditional paper-and-pencil test where all students take the same items, students taking the Consortium's CAT take items and tasks targeted to their ability level. This means that the Consortium needed to develop a very large number of items in order to meet the needs of the student population.

In addition to the items for the CAT, the Consortium also developed performance tasks. All students take performance tasks designed to measure a student's ability to integrate knowledge and skills across multiple assessment targets. These performance tasks may also be delivered via the same online assessment delivery system as the CAT.

Tables 3.2 and 3.3 show the total number of CAT items found in the 2014-2015 summative and interim item pools, including both operational items and field test items embedded in the summative test. As the tables show, over 1,200 ELA/L CAT items were developed in each of grades 3 – 8, and 3,697 items were developed for high school. In mathematics, approximately 1,500 items were developed in each of grades 3 – 8, and 2,778 items were developed for high school.

[23] Examples of many of the item types may be found at: http://www.smarterbalanced.org/sample-items-and-performance-tasks/.

[24] Interim assessments are currently fixed form assessments.

TABLE 3.2
TOTAL NUMBER OF ELA/L CAT ITEMS ON THE 2014-2015 SUMMATIVE AND INTERIM ASSESSMENTS BY GRADE

ELA/L				
GRADE	SUMMATIVE	EMBEDDED FIELD TEST	INTERIM	GRAND TOTAL
3	592	660	93	1,345
4	568	637	93	1,298
5	550	646	91	1,287
6	552	634	104	1,290
7	510	632	100	1,242
8	500	609	97	1,206
HS	1,457	2,141	99	3,697
Grand Total	**4,729**	**5,959**	**667**	**11,365**

TABLE 3.3
TOTAL NUMBER OF MATHEMATICS CAT ITEMS ON THE 2014-2015 SUMMATIVE AND INTERIM ASSESSMENTS BY GRADE

MATHEMATICS				
GRADE	SUMMATIVE	EMBEDDED FIELD TEST	INTERIM	GRAND TOTAL
3	855	564	54	1,473
4	834	659	61	1,554
5	828	616	60	1,504
6	749	677	54	1,480
7	680	684	58	1,422
8	649	693	56	1,398
HS	1,781	932	65	2,778
Grand Total	**6,376**	**4,825**	**408**	**11,609**

There were approximately 45 performance tasks per grade developed in each of grades 3 – 8 in both ELA/L and mathematics. In high school, the Consortium created 118 performance tasks for ELA/L and mathematics. Each performance task has multiple associated items: four and six items per performance task in ELA/L and mathematics, respectively. Tables 3.4 and 3.5 show the number of performance tasks on the 2014-2015 summative assessments and the number of performance tasks slated for non-embedded field testing in spring 2016.

TABLE 3.4
TOTAL NUMBER OF ELA/L PERFORMANCE TASKS (PT) DEVELOPED BY GRADE

ELA/L				
GRADE	SUMMATIVE	NON-EMBEDDED FIELD TEST	INTERIM	GRAND TOTAL
3	14	29	4	47
4	19	24	4	47
5	20	21	4	45
6	14	25	3	42
7	19	21	4	44
8	21	22	4	47
HS	24	90	4	118
Grand Total	**131**	**232**	**27**	**390**

TABLE 3.5
TOTAL NUMBER OF MATHEMATICS PERFORMANCE TASKS (PT) DEVELOPED BY GRADE

MATHEMATICS				
GRADE	SUMMATIVE	NON-EMBEDDED FIELD TEST	INTERIM	GRAND TOTAL
3	18	24	6	48
4	19	20	6	45
5	15	28	6	49
6	18	18	6	42
7	16	18	6	40
8	18	20	6	44
HS	17	95	6	118
Grand Total	**121**	**223**	**42**	**386**

Item Pool Expansion
The Consortium will expand the item pool and allow the administration of off-grade-level items during the computer-adaptive portion, under the following three conditions:

- On-grade blueprint coverage minimum is met;
- Student performance is clearly in Level 1 or in Level 4;
- Out-of-grade-level items better meet measurement and content requirements than on-grade-level items.

The expansion of the item pool in the CAT should improve the precision of measurement for those students with very high or very low ability estimates on the test. Only on-grade-level performance tasks will be administered.

Selection of Items for the Operational Item Pool
The statistical quality of the items was again evaluated following the 2014 field test. Items that did not perform well according to established psychometric criteria (for example, item statistics such as difficulty and discrimination) were removed from the item pool. The same psychometric criteria were used to judge items regardless of whether the items were used on the interim assessment or the summative assessment.

For the first operational year (2014-2015), items for both the interim assessment and the summative assessment were drawn from the same item pool. In the future, most items will first be administered on the summative assessment before entering the interim item pool. A handful of items will move directly from the field test to the interim item pool if necessary to meet the content requirements of the blueprints.

DESIGNING THE ONLINE ASSESSMENT SYSTEM: SMARTER BALANCED SYSTEM ARCHITECTURE[25]

The technological framework for the Smarter Balanced online assessment and reporting system is described in the Consortium's system architecture. Much like the decisions that one makes when building the assessment (how many item are needed, what types of tasks are needed to support the desired claims about student performance, etc.), similar decisions are needed to establish the technological specifications for an online assessment system. The Consortium's system architecture describes the rules, standards, and specifications needed for the successful integration of the applications, databases, hardware, data formats, etc. that support the Consortium's vision and purpose.

[25] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/03/SmarterBalanced_ArchitectureReport_120321.pdf

As described in the system architecture, the Consortium created an open-source, cohesive online assessment system that supports interim and summative assessments. This open-source system allows flexibility to member states as they decide the best approach for administering and reporting the Consortium's assessments in their own states. This section will address the two aspects of the open-source comprehensive assessment system that will be most often used by educators and students: assessment delivery and assessment reporting.

Promise of an Open-Source System

From its inception, the Consortium embraced an open-source approach in order to maximize flexibility for our members. Like all open-source software, a single company does not own our open-source assessment system meaning that any entity can use it. This means that member states can partner with the vendor of their choosing in order to administer the Consortium's assessments.

In addition, the Consortium expects that the use of open-source software will encourage innovation and quality improvements to the assessment system. In the traditional assessment market, states must ask vendors to make changes or enhance functionality of the vendor's proprietary assessment systems. Using an open-source system, users and developers will have the opportunity to modify the open-source software in order to add desired functionality. For example, a state may modify the open-source software to add functionality for a teacher to enter observational notes as students take the test. The Consortium must pilot and approve this additional functionality before it can be used for operational assessment.

The Consortium expects that the use of the open-source software will enhance market innovation in the assessment industry. As users and developers start improving the open-source system, vendors will likewise add similar features as well as other features demanded by customers.

The Consortium embraced an open-source approach in order to maximize flexibility for our members.

Another benefit of maximizing the number of vendors that can use the system is to increase competition in the marketplace. The Consortium felt it was important to assure that the already small marketplace of vendors capable of administering statewide assessments not be diminished by the advent of consortium-based testing in the form of a single vendor winning a contract for administering assessments for all states in the consortium.

Finally, the Consortium's comprehensive tool for administering and reporting was built to allow maximum flexibility for our member states. States may use the system in its entirety, use components of the system, or use a completely different system so long as comparability from member to member can be demonstrated.

Assessment Delivery

The assessment delivery system is a set of web applications that manage the registration of students for tests, delivery of those tests to the students, scoring of test items, integration of item scores into an overall test score, and delivery of scores to a data warehouse (whether that warehouse resides at the individual state or at the consortium is up to the individual state). To ensure access by the largest range of students, the Consortium embedded usability and accessibility features as well as accommodations into the assessment delivery system itself.

The assessment delivery system is flexible, meaning states may use it in its entirety, may use components of the open-source system, or they may use a completely different system. The interim assessment and the summative assessment can both be delivered through the Consortium's assessment delivery system. Even though member states are not required to use the open-source system, they must comply with the specifications

and guidelines that the Consortium has developed for its use. To ensure comparable test scores across all students, member states must make certain that their assessment delivery system complies with the Consortium's Implementation Readiness Package.

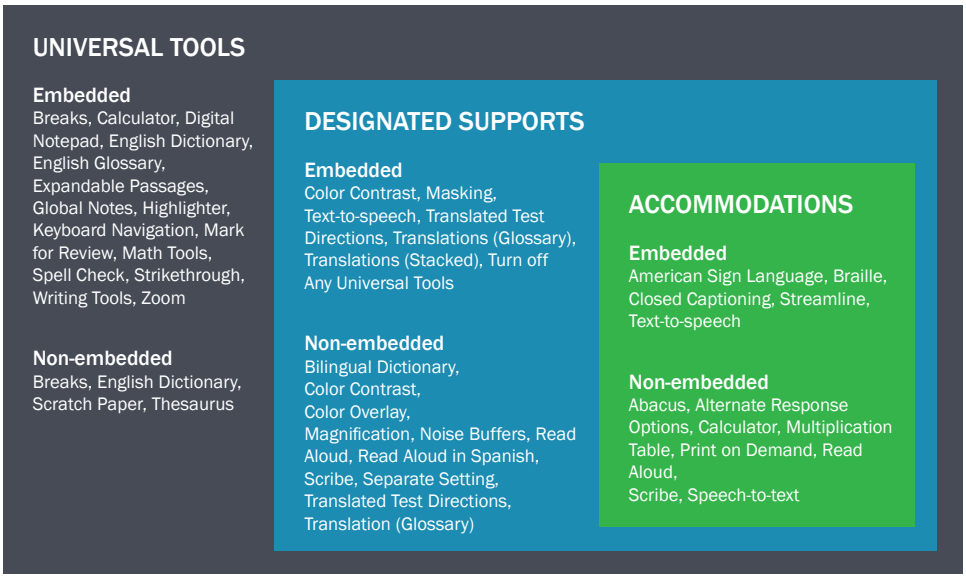
Maximizing Accessibility for All Students

The Consortium strives to provide every student with a positive and productive assessment experience, generating results that are a fair and accurate estimate of each student's achievement. Pulling from rich research literature, member states worked together to create an accessibility framework that specified the tools, supports, and accommodations that would be available on the Consortium's assessments. This framework was built with all students in mind; all students (including English language learners, students with disabilities, and English language learners with disabilities) are expected to participate and perform on state assessments. This group's work became the basis for the Consortium's *Usability, Accessibility,* and *Accommodations Guidelines.*[26]

The Consortium recognizes that the validity of students' assessment results depends on each and every student having appropriate assessment features, universal tools, designated supports, and/or accommodations, when needed. The Consortium has created embedded and non-embedded versions of these accessibility resources, depending on whether they can be delivered through the online assessment delivery system (embedded) or outside of the system (non-embedded). Universal tools, designated supports, and accommodations all yield valid scores when used in the manner specified by the Consortium's guidelines. Figure 3.1 below shows the conceptual model that serves as the basis for the Consortium's Usability, Accessibility, and Accommodations Guidelines

Universal tools, the largest box in Figure 3.1, are available to all students. Some embedded universal tools include digital notepads and highlighters. Scratch paper is an example of a non-embedded universal tool. Students who receive designated supports and/or accommodations may also use the universal tools.[27]

FIGURE 3.1 CONCEPTUAL MODEL UNDERLYING THE CONSORTIUM'S *USABILITY, ACCESSIBILITY,* AND *ACCOMMODATIONS GUIDELINES*



[26] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/08/SmarterBalanced_Guidelines.pdf

[27] See *Usability, Accessibility*, and *Accommodations Guidelines* for a full list of the embedded and non-embedded universal tools, designated supports, and accommodations that are available on the interim and summative assessments.

Designated supports are those features that are available for use by any student for whom the need has been indicated by an educator (or a team of educators along with the parent/guardian and the student). Masking, which blocks off content that is either not immediately needed or distracting, is an example of an embedded designated support. Separate setting (i.e., student is tested in a setting that is different from that used by most students) is an example of a non-embedded designated support.

Accommodations are only available to those students with documentation of the need through a formal plan (i.e., IEP or 504 Plan). Accommodations are changes in procedures or materials that increase equitable access during the assessment. Those students also may use designated supports and universal tools. American Sign Language (ASL) is an example of an embedded accommodation that may be used with ELA/L listening items and mathematics. For students who need this accommodation, test content is translated into an ASL video. An ASL human signer and the signed test content are viewed on the same screen. Use of multiplication tables in grade 4 and above is an example of a non-embedded accommodation that can be used by students with a documented and persistent calculation disability.

Accessibility and Test Registration

Authorized users enter the accessibility resources (in other words, the specific designated supports and/or accommodations) that students need in order to complete the test. For example, the Consortium's assessment text is normally presented in black and white. If a student, however, needs a different color combination in order to read the online text, then an authorized user (e.g., a test administrator) would enter the necessary accessibility resource during the test registration process. To facilitate the decision making regarding which specific universal tools, designated supports, and accommodations a student needs, the Consortium created an Individual Student Assessment Accessibility Profile (ISAAP)[28] process and tool that states can use. For students requiring one or more designated supports and/or accommodations, schools document this need prior to test administration. Furthermore, the ISAAP can include information about universal tools that may need to be turned off for an individual student.

Identifying Needs Prior to Test Administration

By documenting individual student accessibility needs prior to test administration, the state's digital test administration delivery system activates the specified options when the student logs in to an assessment. In this way, the ISAAP supports educators and schools as they focus on each individual student to document the universal tools, designated supports, and accommodations required for valid assessment of that student. The ISAAP is designed to help end users identify the accessibility information that needs to be captured during the test registration process.

Maximizing Accuracy of Student Test Scores

The Consortium's computer adaptive tests allow students to take a test that is targeted to their unique level of performance. With a traditional paper-and-pencil test, all students take the same test that must be capable of measuring student achievement reasonably accurately across a wide range of achievement. A computer adaptive test adjusts the difficulty of the items based on the performance of the student taking the test on all items previously presented to the student. For example, a student who answers an item correctly will receive a more challenging item, while an incorrect answer generates an easier item. So, a higher performing student will encounter a different set of items than a lower performing student.

The use of a computer adaptive test allows the Consortium to provide students with more accurate estimates of their achievement than they would get on a traditional paper-and-pencil test.

[28] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/08/ISAAP-Tool-Instructions-.docx

In order to administer tests targeted to the individual ability level of a wide range of students and to report results from different sets of items on the same scale, two key conditions must be met:

- It is necessary to have a very large item pool within and across grade levels;
- Extensive field testing and psychometric analyses must be done prior to testing to place all of the items on the same scale.

All items in the Consortium's item pool were administered during the spring 2014 field test. The items from the spring 2014 field test were analyzed using rigorous psychometric methods, and poor performing items were removed from the item pool. The remaining items were calibrated using item response theory, an advanced psychometric model. This model allows both students and items to be placed on a common scale.

At the same time, the Consortium's adaptive system ensures that each student takes a test that fulfills the test blueprint. In other words, every student takes the Consortium's mathematics test regardless of the individualized set of items encountered during the testing event. This safeguard guarantees that all students are measured on all assessment targets and claims specified in the ELA/L or mathematics summative blueprints. Because the items are on a common scale and the same test blueprint is used, student scores may be compared to one another regardless of the set of items taken during testing.

The use of a computer adaptive test allows the Consortium to provide students with more accurate estimates of their achievement than they would get on a traditional paper-and-pencil test, especially for those students who are high or low performing in a content area. An important benefit of reducing "floor" and "ceiling" effects is that measures of individual student growth will be much more accurate for very low and very high performing students.

The Consortium created the Implementation Readiness Package to ensure that students receive comparable test scores regardless of the system used to deliver and score the assessments.

Maximizing Flexibility for States

States may direct vendors to engage the open-source system in its entirety for administering the assessment; they may have their vendor use components of the open-source system within the vendor's proprietary system; or they may forgo the open-source system by engaging only a proprietary system.

The open-source system contains components that operate behind the scenes to ensure the successful delivery of the test. These include the component that allows the adaptive functionality of the assessment, the components that machine score test items, the components that send items to human scorers, and the components that assign test scores to students. States may ask vendors to engage only certain components when delivering the Consortium's assessments. For example, a state using a proprietary assessment delivery system may require their vendor to use the adaptive component from the open-source system.

While the Consortium does not require that member states use the open-source system (or any component of the open-source system), all assessment delivery systems must meet the Consortium's standards for implementation readiness. Compliance with the implementation readiness standards ensures that tests are delivered and scored consistently for all students in all member states regardless of the tool used for assessment delivery.

Ensuring Comparability of Student Test Scores

Whatever approach is used, member states are responsible for deploying the assessment delivery system within their own state. Because unique assessment delivery systems were operating in various member states, the Consortium created an

Implementation Readiness Package[29] to ensure that students receive comparable test scores regardless of the system used to deliver and score the assessments. To do this, the Implementation Readiness Package helps vendors show evidence that their various assessment delivery systems are able to correctly do the following:[30]

- Display items on the screen;
- Machine score test items;
- Score tests as a whole;
- Receive and process student registration and accommodation data;
- Receive and process test items;
- Deliver all deidentified student data to the Consortium's data warehouse.

Comparability of Items Scored by Humans

The Consortium has taken several steps to ensure that items requiring human scoring (e.g., essay items) receive comparable scores regardless of the vendor conducting the scoring. As in any assessment program, the quality and comparability of scores across raters is dependent upon the quality of training materials and procedures, the selection of qualified raters, and active monitoring throughout the scoring process.

Training

The Consortium has provided member states' vendors with the same sets of training papers, scoring rubrics, and qualification papers for all items that require human scoring. A set of training papers consists of student work that is used to train raters on how to apply the scoring rubrics. The scoring rubrics provide specific guidance on how to score each item.

Qualification

Once raters are trained, they must pass a qualification process to show that they can rate each qualification paper accurately. Table 3.6 shows the exact agreement standards used by the Consortium for different score-point ranges for both qualification and quality monitoring. Suppose that a rater is scoring an item that is worth zero to one point, then that rater must assign a score to the item's qualification papers that agrees 90% of the time with the previously assigned score.

Monitoring

The Consortium also requires that all vendors monitor inter-rater reliability. Inter-rater reliability refers to the degree to which two raters exactly agree with each other. The Consortium requires that two independent raters score 10% of all papers. Analysis of inter-rater reliability reports is an excellent source to determine if raters are drifting away from scoring guidelines.

Finally, the Consortium requires that all vendors engage in a quality monitoring process during item scoring. For this process, a vendor's supervisory staff select and pre-score a set of validity papers for each item that are representative of the range of possible score points. Validity papers were interspersed with the other papers that raters are scoring. The scores assigned to the validity papers are compared to the approved score, providing information about the accuracy and reliability of the rater.

[29] http://www.smarterapp.org/documents/ImplementationReadinessPackage.pdf

[30] The functionality to check the adaptive algorithm will be added for the 2015-2016 administration.

End of Grant Report

Assessment Data Warehouse and Reporting

In creating the data warehouse and reporting system, the Consortium focused on the need for flexibility as each state has unique concerns regarding student data and unique needs around reporting.

Protecting Personally Identifiable Information

The Consortium, along with our member states, is committed to protecting students' personally identifiable information. The Consortium does not require that member states send personally identifiable information to the data warehouse. Instead, member states may choose to send de-identified student data to the Consortium's data warehouse. With de-identified data, students' personally identifiable information has been removed from the data.

Meeting State Needs for Data Storage

The Consortium requires that member states send de-identified student test data to the Consortium data warehouse; however, the Consortium (as just mentioned) does not need or require states to send students' personally identifiable information with the response data. States may choose to store their personally identifiable student data in the data warehouse managed and hosted by the Consortium or in a data warehouse that they manage and host at a state level depending on their needs and policies. If states opt to store personally identifiable data in the Consortium data warehouse, the privacy of those students' data is rigorously safeguarded.

It is necessary to collect de-identified student response data so that the Consortium's psychometricians may analyze the test data. This allows them to maintain the item bank, examine the test scale, and ensure the proper functioning of the adaptive algorithm. For example, they will calculate various statistics for each item, particularly to flag them for potential issues to be reviewed by the content, bias, and accessibility review committees.

Meeting State Needs for Reporting

The Consortium has designed and deployed a reporting system that states may use. This system provides reports for various levels (e.g., student, school, district, state), and will provide longitudinal reports showing student performance over time. We recognize that member states have unique reporting needs. For this reason, member states may use our reporting system, integrate aspects of it into their own systems, or deploy their own reporting systems.

Reporting Student Results

States have the option of not reporting on a common metric with common achievement standards. All states can report student scores using the Consortium's vertical scale in ELA/L and in mathematics. A vertical scale means that the grade-level tests are reported in the same metric so that scores from one grade level can be compared to another grade level. This allows stakeholders to examine student growth toward college and career readiness as students move from grade to grade. The vertical scale was built using data collected during the spring 2014 field test. This scale will be re-examined and, if necessary, adjusted using data from the spring 2015 operational test.

In addition, all states will use a common set of achievement levels when reporting student scores. In other words, all states have agreed to use the same set of cut scores at each grade when separating students into achievement levels 1, 2, 3, and 4.

Achievement Level Setting

The Consortium developed a multi-step process, called an Achievement Level Setting, to establish the cut scores that separate students into achievement levels based on their performance on the summative assessment. Through this four-step process, achievement levels were established in ELA/L and mathematics across grades 3 – 8 and high school. The Achievement Level Setting process was guided by the achievement level descriptors created in 2012 (see Chapter 2).

Step 1: Online Panel

In order to maximize participation, the Consortium invited educators, parents, and other concerned citizens from member states to participate in an online achievement level setting. During the online session, thousands of teachers and other interested parties independently reviewed test questions and recommended the level of performance required for students to be considered on-track toward college and career readiness. In other words, the online panel only made recommendations in regard to Level 3.

Step 2: In-Person Panel

The in-person panel allowed teams of educators and other stakeholders nominated by member states to deliberate and recommend cut scores for all four achievement levels: Level 1, Level 2, Level 3, and Level 4. Separate grade-level panels, consisting of approximately 30 members each, were convened to recommend cut scores for ELA/L and mathematics. Member states nominated all panelists, which included teachers and administrators, higher education faculty, business and community leaders, and parents.

Step 3: Cross-Grade Review

Following the in-person achievement level setting, a subset consisting primarily of educators from the in-person panels met to review the achievement levels recommended in Steps 1 and 2. Separate cross-grade panels were convened for ELA/L and for mathematics. The purpose of the cross-grade review was to ensure that the achievement levels were appropriately aligned across grades and would accurately reflect student progress from year to year. The panelists at the cross-grade review examined the earlier recommendations and suggested changes that would improve cross-grade alignment of the achievement levels.

Step 4: State Approval

The final recommendations were reviewed, adjusted, and then endorsed by the member states. Higher education leaders participated in the decisions regarding grade 11 achievement levels to ensure they reflect the expectations of colleges and universities. An independent Technical Advisory Committee and an expert auditor (Gregory Cizek, a nationally-recognized expert in standard setting) certified that the multi-step process was appropriately implemented. Each member state must decide whether to adopt the Consortium's achievement levels or set their own achievement levels.

CHAPTER 4. POPULATION SERVED

The Consortium's 2014 field test marked the largest field test administration of a K-12 assessment in the United States. This chapter examines the population that took the spring 2014 field test and the expected population[31] who took the spring 2015 operational test.

SPRING 2014 FIELD TEST

From March 25-June 6, 2014,[32] the Consortium administered our field test to 4.2 million students in approximately 16,500 schools across the 21 Governing States and the U.S. Virgin Islands. The field test was administered to students in grades 3 – 8 and 11, plus a small number of students in grades 9 and 10. Some member states administered the field test to all students in their state while other member states administered the field test to a sample of their students (with a minimum target of 10 percent of students for each content area).

The spring 2014 field test was a major step forward in the Consortium's development of the next generation of assessments. The large-scale field test allowed us to evaluate the performance of the more than 19,000 items and performance tasks in our item pool. The Consortium used the field test data to examine the items to understand which performed well and which needed to be improved. This information will inform future item writing efforts so that all test questions meet the consortium's needs.

The information collected during the spring 2014 field test also informed the October 2014 standard setting. As explained in Chapter 3, groups of educators set preliminary achievement standards, which indicate if students are on track to achieve content college and career readiness in ELA/L and mathematics. The preliminary achievement standards will be examined once the data from spring 2015 operational test are available.

For schools, districts and states, the field test provided an opportunity to make sure that technology systems and administration logistics were ready for the operational test in spring 2015. For students and teachers, the field test presented an opportunity to interact with the Consortium's online assessment system and with the new item types that comprise the next-generation assessments.

From March 25-June 6, 2014, the Consortium administered our field test to 4.2 million students in approximately 16,500 schools across the 21 Governing States and the U.S. Virgin Islands.

Designed for All Students

The Consortium worked to provide all students with a positive and productive field test experience. As described in Chapter 3, the Consortium created Usability, Accessibility, and Accommodations Guidelines that detail the assessment features students can use during testing to ensure that all students can access the assessment's content. In addition, the Consortium created the Individual Student Assessment Accessibility Profile (ISAAP) process and tool to ensure that educators select the accessibility supports needed by each student during the interim and summative assessments.

The field test was open to all eligible students in member states' schools and districts selected to participate in the assessment. The only students not eligible for the field test were those who would not ordinarily participate in the general state assessment;

[31] The actual data will not be available until after the completion of this report.

[32] An optional makeup week was available June 9-13, 2014.

that is those students who participate in an alternate assessment of alternate achievement standards (the "1%" assessment). Also, English language learners who were enrolled for the first year in a U.S. school could be excused from participation in the assessment.

State Participation

As stated above, students from all Governing States as well as the U.S. Virgin Islands participated in the field test. The following Governing States administered the field test to all eligible students: California, Idaho, Montana, and South Dakota. Table 4.1 shows the number of students who participated in the field test by state.

Psychometric Analysis

The quality of the field test data was critical to the success of the psychometric analysis of the Consortium's assessments. More students participated in the field test than were needed for psychometric analysis; therefore a sample of student responses was used from those who participated in the field test. To the degree possible, this sample was selected so that it was more representative of the total Consortium student population than the total sample that participated in the field test. In the few cases of demographic variables where it was not possible to obtain a sample representative of the total Consortium population, the sample was weighted statistically so that it was representative of the total Consortium student population. Table 4.2 shows the percentage of students in the Consortium by state.

It is important to note that the samples are representative of the Consortium, not individual states. For example, California contributes about 36% of all students to the Consortium. The remaining columns show the percentage of students sampled from each state by grade. In those cases where the percentage of students in California is over or under 36%, the results were statistically weighted to achieve representativeness.

The careful construction of the field test sample allowed the Consortium to conduct various types of psychometric analyses on the data and to be assured that the results of these analyses were representative of the entire Consortium student population. In some cases, it was necessary to oversample some groups so that specific analyses could be conducted. For example, it was necessary to oversample Native American students in order to empirically determine if test questions performed differently for this group than they did for other groups.

The final Smarter Balanced vertical scale was established, and the quality of the item pools was evaluated using data sampled from the 2014 field test. In addition, the quality of the item pools was evaluated using the data from the 2013 pilot test. It was necessary to evaluate a sufficiently large number of items to ensure that the item pools were large enough to support both the operational interim and summative assessments. The field test data was also used to evaluate the procedures established for the test administrations and computer delivery system as well as to evaluate the appropriateness of the assessment accommodations.

TABLE 4.1
TOTAL NUMBER OF STUDENTS COMPLETING THE ONLINE FIELD TEST

	TOTAL STUDENTS TESTED
Total	4,200,799
California	3,114,871
Connecticut	263,613
Delaware	13,772
Hawaii	20,259
Idaho	161,696
Iowa	6,725
Maine	16,788
Michigan	76,173
Missouri	27,509
Montana	72,783
Nevada	37,446
New Hampshire	11,177
North Carolina	2,124
North Dakota	5,662
Oregon	25,073
South Carolina	9,075
South Dakota	70,426
US Virgin Islands	2,584
Vermont	4,735
Washington	186,790
West Virginia	15,705
Wisconsin	52,068
Wyoming	3,745

TABLE 4.2
SAMPLE SIZES (IN PERCENTAGES) AT EACH GRADE LEVEL FOR EACH SMARTER MEMBER STATE FOR CALIBRATING ALL ITEMS

STATE	PERCENTAGE OF CONSORTIUM	GRADE 3		GRADE 4		GRADE 5		GRADE 6		GRADE 7		GRADE 8		HIGH SCHOOL	
		ELA/L	Math	ELA/L	Math	ELA/L	Math	ELA/L	Math	ELA/L	Math	ELA/L	Math	ELA/L	Math
California	36.4	38.6	30.0	36.1	25.8	27.5	24.2	36.1	53.9	41.6	60.2	36.7	62.1	62.7	61.5
Connecticut	3.2	14.7	17.1	15.0	21.5	19.7	21.4	17.4	11.6	15.6	9.6	18.1	8.4	12.1	12.1
Delaware	0.7	0.6	0.4	0.5	0.5	0.3	0.6	0.5	0.2	0.6	0.1	0.6	0.4	0.6	0.6
Hawaii	1.0	2.5	1.7	1.7	1.1	1.4	1.3	0.8	0.6	1.0	0.8	0.6	0.9	0.5	0.4
Idaho	1.6	4.8	7.1	5.2	9.3	6.8	10.7	5.5	2.7	5.2	1.4	5.9	1.7	7.1	8.4
Iowa	2.9	0.7	2.1	0.0	0.1	0.6	0.1	0.3	0.1	0.2	0.5	0.0	0.0	0.0	0.0
Kansas*	2.8														
Maine	1.1	0.8	1.0	1.1	1.2	0.9	1.0	0.7	0.6	0.8	0.8	0.6	0.4	0.3	0.3
Michigan	9.2	5.6	5.0	4.9	3.9	5.5	3.5	4.2	3.0	3.7	2.6	3.8	2.8	3.4	3.8
Missouri	5.3	1.5	0.9	1.4	0.7	0.5	0.5	1.2	0.6	0.2	0.5	1.0	0.9	1.0	0.9
Montana	0.8	2.9	3.8	3.5	5.4	4.5	5.7	3.3	1.5	3.3	0.6	3.8	0.8	3.2	3.1
Nevada	2.5	2.3	2.3	2.2	1.7	2.2	1.9	2.3	1.1	1.9	1.5	2.1	1.6	1.4	1.9
New Hampshire	1.1	0.5	0.7	0.6	0.4	0.4	0.4	0.3	0.3	0.6	0.4	0.3	0.3	0.2	0.3
North Carolina	8.6	0.0	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.0	0.1	0.1
North Dakota	0.6	0.4	0.3	0.2	0.3	0.3	0.2	0.3	0.1	0.5	0.1	0.2	0.2	0.2	0.2
Oregon	3.3	1.8	2.0	1.7	1.6	1.8	1.7	1.7	1.7	1.9	1.5	1.4	1.7	0.5	0.6
South Carolina	4.2	0.5	0.2	0.3	0.2	0.5	0.2	0.2	0.0	0.2	0.0	0.0	0.0	0.4	0.4
South Dakota	0.7	4.0	4.2	6.0	2.8	5.1	4.0	5.9	3.4	5.4	3.5	5.7	2.2	3.2	3.1
Vermont	0.6	0.5	0.4	0.4	0.4	0.4	0.3	0.5	0.2	0.4	0.2	0.4	0.2	0.3	0.4
US Virgin Islands		0.0	0.0	0.5	0.8	0.0	0.0	0.0	0.0	0.7	0.7	0.0	0.0	0.1	0.1
Washington	6.0	12.0	16.1	14.1	18.8	17.5	18.4	14.7	16.7	12.6	12.7	15.4	13.3	1.7	0.9
West Virginia	1.6	0.9	0.8	0.8	0.6	0.3	1.0	1.0	0.5	0.6	0.4	0.5	0.5	0.6	0.7
Wisconsin	5.0	4.2	3.9	3.4	2.7	3.3	2.6	2.7	1.1	2.7	1.9	2.9	1.8	0.0	0.0
Wyoming	0.5	0.2	0.2	0.3	0.2	0.3	0.1	0.4	0.2	0.0	0.0	0.0	0.1	0.1	0.1
Sample Size	–	85,889	95,143	94,915	109,441	88,293	108,412	93,536	117,691	93,431	117,049	98,433	116,459	261,405	262,111

* Kansas left the Consortium after this sampling design was completed.

Under-Represented Populations

Because one of the goals of the psychometric analyses was to evaluate the appropriateness of the assessment accommodations, it was important to ensure that traditionally under-represented populations were adequately represented in the target sample. These populations include students with disabilities, English language learners, and English language learners with disabilities. Tables 4.3 and 4.4 show the percentage of the population and the sample for various demographic groups in ELA/L and mathematics. These tables show that the percentage of traditionally under-represented students in the sample was very close to the percentage of those groups of students in the population.

In selecting the sample of the under-represented students, the Consortium was careful to ensure that the sample adequately represented students with specific disabilities:

- Autism;
- Deaf-Blindness;
- Emotional Disturbance;
- Hearing Impairment;
- Multiple Disabilities;
- Orthopedic Impairment;
- Other Health Impairment;
- Specific Learning Disability;
- Traumatic Brain Injury;
- Visual Impairment;
- Intellectual Disability;
- Speech or Language Disabilities;
- Developmental Delay.

TABLE 4.3
STUDENT DEMOGRAPHIC CHARACTERISTICS (IN PERCENTAGES) FOR CALIBRATING ALL ELA/L ITEMS

DEMOGRAPHIC GROUPS	GRADE 3		GRADE 4		GRADE 5		GRADE 6		GRADE 7		GRADE 8		HIGH SCHOOL	
	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample
Male	51.0	51.4	51.1	51.4	51.0	51.2	51.1	51.2	51.1	51.4	51.1	51.6	51.3	51.2
Female	48.5	48.6	48.5	48.6	48.5	48.8	48.5	48.8	48.5	48.6	48.5	48.4	48.7	48.8
American Indian or Alaska Native	1.1	2.8	1.1	3.2	1.1	3.1	1.1	3.3	1.1	3.2	1.1	3.2	1.0	1.9
Asian	6.5	7.3	6.7	7.5	6.7	7.1	6.6	6.8	6.5	7.6	6.7	6.9	6.1	8.2
Native Hawaiian or other Pacific Islander	0.8	1.5	0.8	1.1	0.8	1.2	0.8	0.9	0.7	0.8	0.7	0.9	0.7	0.9
Hispanic or Latino	28.7	30.2	28.0	28.4	27.8	28.6	27.4	28.8	26.9	32.8	26.6	27.8	26.7	30.3
Black or African American	10.7	10.0	10.6	9.3	10.8	10.2	11.1	10.4	11.4	9.9	11.4	10.6	11.8	9.9
White or Caucasian	48.7	54.1	49.4	56.6	49.6	58.8	49.9	57.3	50.3	52.3	50.6	57.3	50.2	50.3
Two or More Races	3.6	4.1	3.4	4.2	3.3	4.0	3.2	3.9	3.1	3.5	3.0	3.4	2.7	3.2
Individualized Education Program	11.4	10.3	12.3	10.9	12.5	11.5	12.1	11.1	11.7	10.4	11.5	10.4	10.4	8.1
Limited English Proficient	18.0	16.6	15.3	13.6	12.6	11.1	9.8	9.7	8.7	9.7	7.8	7.4	7.1	6.2
Economic Disadvantaged	55.4	53.4	55.3	51.9	54.6	50.6	54.2	51.1	53.1	52.8	51.9	48.6	48.6	46.2

TABLE 4.4
STUDENT DEMOGRAPHIC CHARACTERISTICS (IN PERCENTAGES) FOR CALIBRATING ALL MATHEMATICS ITEMS

DEMOGRAPHIC GROUPS	GRADE 3		GRADE 4		GRADE 5		GRADE 6		GRADE 7		GRADE 8		HIGH SCHOOL	
	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample
Male	51.0	51.2	51.1	51.3	51.0	51.4	51.1	51.1	51.1	51.5	51.1	51.2	51.3	50.9
Female	48.5	48.8	48.5	48.7	48.5	48.6	48.5	48.9	48.5	48.5	48.5	48.8	48.7	49.1
American Indian or Alaska Native	1.1	3.0	1.1	4.4	1.1	4.3	1.1	2.1	1.1	1.9	1.1	2.0	1.0	2.0
Asian	6.5	7.0	6.7	7.0	6.7	6.9	6.6	8.2	6.5	8.3	6.7	8.2	6.1	8.5
Native Hawaiian or other Pacific Islander	0.8	1.4	0.8	1.1	0.8	1.1	0.8	0.9	0.7	1.0	0.7	0.9	0.7	0.9
Hispanic or Latino	28.7	31.3	28.0	29.0	27.8	27.7	27.4	33.7	26.9	36.2	26.6	33.0	26.7	31.5
Black or African American	10.7	9.8	10.6	9.7	10.8	8.5	11.1	7.9	11.4	8.3	11.4	8.0	11.8	9.9
White or Caucasian	48.7	56.2	49.4	59.3	49.6	61.4	49.9	50.7	50.3	46.9	50.6	49.7	50.2	49.0
Two or More Races	3.6	3.8	3.4	4.4	3.3	4.4	3.2	4.4	3.1	3.7	3.0	3.9	2.7	3.2
Individualized Education Program	11.4	10.4	12.3	11.1	12.5	11.3	12.1	11.1	11.7	10.5	11.5	10.2	10.4	8.0
Limited English Proficient	18.0	17.6	15.3	13.3	12.6	10.5	9.8	11.5	8.7	11.0	7.8	9.4	7.1	7.2
Economic Disadvantaged	55.4	53.1	55.3	52.1	54.6	50.0	54.2	52.2	53.1	53.3	51.9	49.9	48.6	46.4

SPRING 2015 OPERATIONAL TEST

The Consortium's assessments were administered operationally for the first time in spring 2015 in 18 states, three BIE schools, and one territory. Table 4.5 shows the states that participated in the assessment, the grades that they tested, and the projected total number of students in each state that took the spring 2015 operational test.

TABLE 4.5
STATES THAT ADMINISTERED THE SPRING 2015 OPERATIONAL TEST

	GRADES TO BE TESTED							PROJECTED NUMBER OF STUDENTS[33]	
Participating State	3	4	5	6	7	8	HS	Grades 3 - 8	HS
Bureau of Indian Education	X	X	X	X	X	X	X	104	123
California	X	X	X	X	X	X	X	2,480,754	487,113
Connecticut	X	X	X	X	X	X	X	240,554	41,482
Delaware	X	X	X	X	X	X	X	61,000	8,700
Hawaii	X	X	X	X	X	X	X	80,300	12,400
Idaho	X	X	X	X	X	X	X	133,550	172,370
Maine	X	X	X	X	X	X	X	81,101	14,634
Michigan*	X	X	X	X	X	X	X	684,000	114,000
Missouri**	X	X	X	X	X	X		420,000	
Montana	X	X	X	X	X	X	X	65,857	10,256
Nevada	X	X	X	X	X	X		213,876	
New Hampshire	X	X	X	X	X	X	X	83,701	14,050
North Dakota	X	X	X	X	X	X	X	50,000	8,000
Oregon	X	X	X	X	X	X	X	250,900	42,100
South Dakota	X	X	X	X	X	X	X	61,600	9,670
US Virgin Island	X	X	X	X	X	X	X	6,950	6,400
Vermont	X	X	X	X	X	X	X	38,904	7,041
Washington	X	X	X	X	X	X	X	469,050	244,710
West Virginia	X	X	X	X	X	X	X	120,367	103,350
Wisconsin***	X	X	X	X	X	X		367,892	

** Michigan administered an assessment named the Michigan Student Test of Education Progress (M-STEP) composed of Smarter Balanced test questions as well as Michigan-developed field test items. It leveraged the Smarter Balanced scale to create its own scale.*

***Missouri administered the full assessment in Grades 5 and 8. They administered only the CAT portion in Grades 3, 4, 6, and 7. Missouri did not administer the high school test.*

****Wisconsin did not administer the high school test.*

The demographics of these students are expected to closely align to the demographics of the students in the spring 2014 field test. The sample for the field test was carefully constructed to mirror the Consortium's student population. This was important so that the results of the field test could be generalized to the total population and they could be applied during the spring 2015 operational test.

[33] Numbers are from Smarter Projected Student Counts 2014_09_15.docx.

CHAPTER 5.
OUTREACH AND COMMUNICATION

The Consortium serves a range of audiences, including state education agencies, students, parents, educators, district and school leaders, higher education faculty and leaders, business and community leaders, policy makers, and the general public. The Consortium uses various communication and outreach strategies to reach stakeholder groups. Some strategies are common across groups, including communication through online/print media or through social media. Other strategies are tailored to specific groups (for example, newsletters targeted to teachers).

GENERAL PUBLIC

The Consortium has set up several communication and outreach tools with which we communicate directly with our stakeholder community. We use these tools to tell stakeholders important information regarding the assessments. The Consortium plans to use these tools to communicate the results of evaluations regarding the validity, reliability, and fairness of our assessments to our various stakeholder groups.

Smarter Balanced Website

The Smarter Balanced Assessment Consortium's main website (http://www.smarterbalanced.org/) serves as a primary source of information for all stakeholder groups. The Consortium's latest news is communicated on the main landing page. From this landing page, users can navigate to areas targeted to their particular group: K-12 Education, Higher Education, and/or Parents & Students. The website also maintains information on upcoming events and other resources related to K-12 education.

From September 2013 to November 2014, an average of 52,321 unique visitors (standard deviation: 21,343 unique visitors) visited smarterbalanced.org each month. The number of unique visitors ranged from a low of 23,837 in January 2014 to a high of 109,777 in November 2014. A Unique visitor is defined as a person who visits the website at least once in a month.[34]

Materials for State Use

Smarter Balanced has created an array of communication materials in both English and Spanish to support state communication efforts. These include:

- One- and two-page fact sheets for key audiences (parents, teachers, higher education) and on key issues (adaptive testing, keyboarding required of younger students, etc.);
- Comprehensive communication toolkits for important events such as the pilot test, field test, and new achievement levels that include an array of customizable resources for use at the state, district, and school levels;
- Customizable presentations for use by state education agency leaders, district superintendents, and school principals.

Webinars

The Consortium provides live training and professional development opportunities through webinars. These webinars are recorded and housed on the main Smarter Balanced website, YouTube, and/or SchoolTube. The webinars cover a range of topics targeted to various audiences.

[34] For the number and definition of unique visitors, see: https://siteanalytics.compete.com/smarterbalanced.org/?gateway=1#.VGz1UlfF85E.

End of Grant Report

The webinars are an efficient way of reaching a large audience. Those attending the live webinar have opportunities to ask questions. The Consortium hosts webinars as key materials are released or as key information is needed. For example, the Consortium hosted a webinar that introduced the achievement level descriptors for ELA/L and for mathematics and explained the process used to create the achievement level descriptors. This webinar was targeted to staff from our member states, K-12 faculty, and higher education faculty.

Online/Print Media

The Consortium provides press releases to relevant online/print media, including the Associated Press and Education Week. Education Week has published 171 news stories, 353 news blogs, and 62 commentaries or opinion blogs about the Consortium in the last five years.

Social Media

The Consortium maintains an active presence on Twitter (@SmarterBalanced), where we have 14,800 followers. The Consortium regularly tweets Smarter Balanced news, CCSS-related events, and other pertinent happenings to their followers.

K-12 EDUCATORS AND ADMINISTRATORS

The active involvement and support of educators and administrators are crucial to the ultimate success of the Consortium. For this reason, the Consortium has targeted communication and outreach specifically for this community.

The active involvement and support of educators and administrators are crucial to the ultimate success of the Consortium.

Newsletters

The Consortium sends out a bi-monthly e-newsletter, Smarter News, which is targeted to teachers. This e-newsletter focuses on the tools that the Consortium is creating for teachers and features guest columns by teachers. It is sent to 20,000 subscribers. The Consortium also sends out a weekly e-newsletter to approximately 600 actively engaged constituents. This newsletter, which features a personal message from the executive director, provides a quick update on key consortium activities.

Achievement Level Descriptors and Information About Achievement Level Setting

The Consortium has developed detailed achievement level descriptors describing the knowledge and skills characterizing student performance at each achievement level at each grade level. This level of information is designed to support K-12 educators in making necessary connections between the test, the content standards, and their curriculum and instruction. The K-12 educator community was an integral part of the process of creating the achievement level descriptors (see Chapter 2) and setting the cut scores (see Chapter 3). This community brought much needed expertise to both sets of workshops. The participants in these workshops have experience implementing the CCSS and working with students, so they were able to make informed decisions about the types of content and knowledge that students should know in order to be college and career ready.

Item Writing and Reviews

Hundreds of K-12 educators took part in writing and reviewing items. The active involvement of educators in the item development and review process helps to ensure that the assessment system provides valid evidence of student learning. Through this process, educators and states build the capacity to support the overall sustainability and success of the assessment system.

HIGHER EDUCATION

The Consortium established an extensive infrastructure to build support for the assessment among higher education faculty and leaders. In addition, the Consortium invited representatives from higher education to participate in establishing achievement level descriptors, setting standards, and reviewing items.

Regional Advisors

The Consortium regional advisors successfully built an infrastructure of higher education leads in all member states. The Consortium selected the regional advisors because of their established careers in higher education and name recognition in their region. The regional advisors worked with higher education state leads to facilitate the necessary policy development and build support for the Consortium. To do this, they held monthly phone calls and webinars, and routinely visited state leads in their regions.

College Readiness Definition and Policy Framework

Higher education and K-12 state leads as well as higher education faculty and K-12 teachers created the Consortium's definition of college content readiness as well as our policy framework for the Grade 11 achievement levels. This joint work represented one of our earliest efforts to involve both higher education faculty and K-12 teachers in discussions of what it means to be college and career ready. These joint efforts facilitated the connection between K-12 and higher education.

Achievement Level Descriptors and Achievement Level Setting

As with K-12, the Consortium invited higher education faculty and leaders to establish achievement level descriptors (see Chapter 2) and to recommend cut scores during the standard setting (see Chapter 3). Faculty from community colleges, four-year colleges, and four-year universities were represented. The higher education faculty brought necessary expertise to the conversation on what it means to be content ready for college.

Career Readiness Task Force

The Consortium created a task force to provide advice and direction on career readiness. This 16-member task force included high school educators and leaders specializing in career technical education, faculty and leaders from community and technical colleges, and state-level staff specializing in career and technical education. The work of this task force resulted in the Smarter Balanced Career Readiness Frameworks. Member states can use these frameworks[35] to provide additional guidance and information to students, parents, teachers and counselors as they chart a path toward meeting each student's career goals and consider the implications of a student's summative ELA/L and mathematics test scores.

Grant-funded Initiatives

Many consortium states took part in grant-funded initiatives led by organizations such as the Lumina Foundation, the William and Flora Hewlett Foundation, the Bill & Melinda Gates Foundation, and the Carnegie Corporation of New York. These states are leading the way for other member states in the adoption of Consortium assessments to inform decisions about student placement in credit-bearing college courses.

The Consortium established an extensive infrastructure to build support for the assessment among higher education faculty and leaders.

STATE PARTNERS

Member states are the Consortium's primary partners in this endeavor to create a comprehensive assessment system. Extensive efforts are ongoing to engage in meaningful communication with our member states.

Executive Director Weekly Messages

State K-12 and higher education leads each receive weekly messages from the Consortium's executive director, Tony Alpert, that summarize key activities, requests, and deadlines for each group.

[35] http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/02/Understanding-and-Using-the-Career-Readiness-Frameworks-V2-Public-Review.pdf

State Toolkit

The Consortium provided member states with a comprehensive State Toolkit for Communications and Implementation that includes modules on describing the state context for Smarter Balanced, developing the case for Smarter Balanced, and implementation planning. Through the toolkit, the Consortium provided member states with generic tools that could be customized to state-specific situations. The toolkit provided step-by-step guidance on how to implement the resources within a state.

Work Groups

Although not an official communication tool, the work groups served as an efficient way to involve large numbers of state personnel in the test design process. The Consortium launched the work groups in the fall of 2010. Representatives from member states comprised the work groups, which oversaw the development and implementation of the Consortium's work. Governing States were required to serve on two different work groups. Advisory States were able to serve on work groups, but they were not required to serve. Member states nominated representatives for the work groups, and the Consortium's Executive Committee assigned representatives to the work groups. All work group members were employees of state education agencies.

The number and type of work groups evolved over the course of the project. The final work groups were:

- Technology and Reporting;
- Item Development and Performance Tasks;
- Test Development & Validation;
- Test Administration and Student Access;
- Formative Assessment Practices and Professional Learning.

The work groups met regularly through virtual meetings, and they provided regular updates to the Executive Committee.

Collaboration Conferences

The Consortium held six collaboration conferences attended by Consortium staff and member states, as well as vendors and consultants working on Consortium contracts. Two conferences were held in 2012, one in 2013, and three in 2014. These conferences allowed the Consortium to convene K-12 State Leads and Higher Education Leads to discuss key Consortium decisions. These conferences also allowed time for the State Chiefs to meet and vote in public session on important decisions. Finally, these meetings allowed Consortium work groups and vendors to meet in person to discuss and resolve pressing contract issues.

CHAPTER 6. SUSTAINING THE WORK OF THE SMARTER BALANCED CONSORTIUM

In March 2013, Smarter Balanced member states voted to approve a post-grant sustainability plan designed to not only create a smooth transition between the end of the grant period and the operational administration of the assessment during the 2014-2015 school year, but also create the infrastructure for the ongoing sustainability of the Consortium and enhancement of the assessment system. The plan was the culmination of a two-year effort of the externally funded and state-led Sustainability Task Force established in 2011. Following the end of the grant period in 2014, the Consortium co-located with the Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the Graduate School of Education and Information Studies at the University of California, Los Angeles (UCLA).

Smarter Balanced remains a state-led organization with major budget, policy, and governance decisions resting with the governing members. Each Smarter Balanced governing member has executed a memorandum of understanding with UCLA to provide shared services to all member states (e.g., project management, maintenance of and enhancements to the assessment content and technical systems, and on-going validity and other research studies). The governing members (either individually or in groups) contracted with one or more organizations to administer the Smarter Balanced assessments to their students beginning in the 2014-2015 school year.

Smarter Balanced remains a state-led organization with major budget, policy, and governance decisions resting with the governing members.

The decision to house the consortium within a state university, in general, and to co-locate with CRESST, in particular, reflects the intent of the Consortium not only to sustain the work completed during the grant period, but also to continue to enhance the assessment system. UCLA offers faculty expertise, research support and a full array of administrative services that the Consortium requires for ongoing improvement.

2014-2015 ADMINISTRATION

All governing members secured contracts with their own assessment contractor(s) to administer the Smarter Balanced assessment to their students during the 2014-2015 school year (see Table 6.1). Some members entered into contracts individually while others formed multi-state groups. There is also variation among members regarding the length of contracts and in the contractors selected to support the administration of the assessments. The ability to independently procure services for administration, scoring, and reporting (rather than a single contract covering the entire Consortium) has resulted in continued marketplace competition as intended.

TABLE 6.1
SUMMARY OF STATES 2014-2015 ASSESSMENT CONTRACTS

GOVERNING MEMBER	INDIVIDUAL (I) OR GROUP (G) CONTRACTS	CONTRACTOR
California	I	ETS
Connecticut	G (CT, NH, VT)	AIR
Delaware		AIR
Hawaii	G (HI, ID, SD, WA, WV, VI)	AIR
Idaho	G (HI, ID, SD, WA, WV, VI)	AIR
Maine	I	AIR
Michigan	I	DRC
Missouri	I	CTB
Montana	I	Measured Progress
Nevada	I	Measured Progress
New Hampshire	G (CT, NH, VT)	AIR
North Dakota	I	Measured Progress
Oregon		AIR
South Dakota	G (HI, ID, SD, WA, WV, VI)	AIR
US Virgin Islands	G (HI, ID, SD, WA, WV, VI)	AIR
Vermont	G (CT, NH, VT)	AIR
Washington	G (HI, ID, SD, WA, WV, VI)	AIR
West Virginia	G (HI, ID, SD, WA, WV, VI)	AIR
Wisconsin	I	ETS

THE SUSTAINABILITY MODEL

In addition to the decision to house the Consortium at UCLA, the Smarter Balanced sustainability plan established key principles to guide the organizational structure of the Consortium. The plan identified the shared services and supports that would be provided by the Consortium along with those the states would be responsible for providing. The three key principles that guided the Consortium were:

- Retain member led governance of the Consortium, with only minor changes to the current governance structure;
- Share state ownership of the item pool, digital library, and other intellectual property of the consortium;
- Establish a business model that retains member autonomy.

That business model requires Smarter Balanced to provide each member with the services necessary to maintain the quality and comparability of the assessment system. At the same time, members retain the autonomy and flexibility to independently acquire assessment management, administration, and scoring services that best fit their individual needs and requirements. In addition, as described above, the Consortium's affiliation with UCLA provides the administrative support needed to sustain the Consortium in an efficient manner that does not require hiring staff to provide those services.

Tables 6.2 and 6.3 provide a summary list of those services that are being provided by the Consortium and the services being independently procured by the states.

The Consortium remains committed to an organization in which members jointly govern decisions affecting the maintenance and improvement of the system, and have independent authority to procure services for the administration, scoring, and reporting of the system.

TABLE 6.2
SERVICES PROVIDED BY THE CONSORTIUM TO MEMBER STATES

SUMMATIVE & INTERIM ASSESSMENTS	• Develop, calibrate and evaluate the quality of items • Ensure integrity of blueprint and scale • Provide necessary Peer Review assurances for federal accountability • Develop and release the Smarter Balanced version of the test administration platform (on an annual basis) • Develop and implement a process to ensure the proper administration of the Smarter Balanced system by states and their assessment contractors • Produce materials and processes to maintain consistency across states (e.g., training, administration manuals, accommodations procedures, etc.) • Produce a standardized reporting system for assessment results, and deploy that system if requested • Host a data warehouse for maintaining and enhancing the system with student data from the various states, either in de-identified or personally identifiable form, depending on state policy • Supply student results to the state level (if requested), and provide access to a reporting system • Conduct research studies in support of the Smarter Balanced validity framework and the use of effective accommodations and supports for students • Design paper & pencil forms for three years as states transition to technology-based assessments
DIGITAL LIBRARY	• Develop and maintain digital library application • Centrally host digital library application • Facilitate development and review of digital library materials • Regular review and evaluation of user needs
MEMBER SERVICES	• Provide general communication tools & templates • Provide “Tier-1” help desk support for State Assessment Directors and Chiefs • Maintain state-led governance system

TABLE 6.3
SERVICES INDEPENDENTLY PROCURED BY MEMBERS

SUMMATIVE & INTERIM ASSESSMENTS	• Deliver the assessment • Host the test administration platform • Provide help desk services to users for test administration • Provide training at the local level on the assessment administration procedures • Score operational items, tasks, and tests • Deploy a reporting system, if the state does not request that Smarter Balanced deploy its system • Produce any special reports to comply with state-specific accountability requirements • Produce and distribute any paper & pencil forms • Manage the coherent flow of institution, teacher, and student data, including the following: - Maintain unique, high-quality student identifier - Register students for testing and deliver the appropriate tests to each student - Reconcile student records - Deliver student data sets to Districts - Manage transmission of grade 11 scores to institutions of higher education • Transmit item score data, total score data, and necessary demographics to the Consortium data warehouse for purposes of maintaining and enhancing the system[36]
DIGITAL LIBRARY	• Establish and maintain user permissions • Engage with Smarter Balanced in the development of digital library materials
MEMBER SERVICES	• Communicate with the legislature and in-state stakeholders • Serve as the primary point of contact for districts, principals, teachers, parents, and other primary users

[36] Depending on state policy, members can do so without submitting individually identifiable data and host their own data warehouse, or states can request that the Consortium host the data warehouse.

CHAPTER 7. INTEROPERABILITY OF TECHNOLOGY-BASED ASSESSMENTS

INTEROPERABILITY

The Smarter Balanced vision for the operational assessment, which began in the 2014-2015 school year, calls for members to individually (or in small groups) contract with organizations to administer and score the assessments. In July 2014, Smarter Balanced released a brief, entitled *Smarter Balanced Technology Implementation Readiness*,[37] which describes the entire assessment platform, the components for which states are responsible, and the resources that are available to support deployment and evaluation of readiness for each state's assessment delivery system. The brief, in conjunction with the *Implementation Readiness Package* (IRP), supports states in their self-assessment of readiness to administer the Smarter Balanced Assessments.

A related aspect of interoperability is specifying the technology standards for items and tasks so that the items can be rendered properly on the assessment delivery system. During Year 3, Smarter Balanced continued to develop the technology standards for items and tasks. In November 2013, the consortium determined that it would develop two formats to support item development. The first, which the consortium has been using in item development, is the SmarterApp Item Format. It was developed by the consortium to match the Smarter Balanced item specifications and accessibility guidelines and has been used to deliver items for the pilot test in spring 2013 and for the field test in spring 2014. The second is a set of compatible technical standards that are a profile of the Smarter Balanced Item Format within the IMS Accessible Portable Item Protocol standards (APIP). This profile will enhance the current APIP standards to account for newer, more complex item types. This will help support the Smarter Balanced vision by providing an option for states to contract with organizations that currently use the APIP format. In Year 4, the consortium will publish the technology standards.

In Year 4, the Consortium published the technology specifications on SmarterApp.org. In addition, the Consortium formed a committee to work with IMS Global to extend the IMS QTI to support the Consortium's new item types. Finally, the Consortium has begun a research project on accessible rendered items (ARI) that will allow the faster development of new item types. Currently, the Consortium projects that it takes four years to develop new item types using available development protocols. Preliminary results from the ARI research are promising, showing that the time needed to develop new item types can be greatly reduced through the use of pre-rendered items. The Consortium will continue to investigate this line of research as the project moves into its operational phase.

Smarter Balanced Implementation Readiness Package

States are responsible for deploying and operating the Smarter Balanced Assessment Delivery System, whether it be the Consortium's open-source system, a vendor's proprietary system, or some combination of both. To aid with this effort, the Consortium has supported the development of the Implementation Readiness Package (IRP). The IRP consists of a collection of "software test harnesses" designed to assist states with their assessment delivery system compliance testing. (A software test harness is an

[37] http://www.smarterapp.org/documents/Technology_Implementation_Readiness_v1-1.pdf

automation tool that consists of software and input data, and it analyzes the expected output with the actual output.) Through the IRP, the Consortium has defined the requirements for assessment delivery systems to deliver assessment items with authenticity, score items and tests properly, and deliver de-identified or identifiable assessment results to the Smarter Balanced Data Warehouse. The IRP allows vendors to self-test their assessment delivery systems.

Each module of the IRP is designed to test one or more specific capabilities, and to provide feedback on compliance status. Table 7.1 below provides a brief description of the planned features for the IRP.

TABLE 7.1
PLANNED FEATURES OF THE IMPLEMENTATION READINESS PACKAGE, VERSION 2.0

IRP MODULE	DESCRIPTION	INTERFACE	FUNCTIONAL REQUIREMENTS TESTED
A	Compliance of test item delivery to the student	*IRP Version 2.X* • Web Service to send requests (as a sample student) for test items, and to receive test item (response)	a. Correctness of item rendering (display) and associated interaction on screen (human visual test) b. Correctness of test/test items delivered to student, given Student Data (provided by the state), Eligibility Requirements, and test data contained in the Test Package c. Capability to receive student responses to assessment items
B	Compliance of receipt of Student Registration data, and Student Accommodations data	*IRP Version 2.X* • Web service to deliver Sample Student Data (function will be similar to authentic retrieval of Student Data from State Data System)	a. Receipt verification of School Data, Student Registration data, and Student Accommodations data
C	Compliance of delivery of fully scale scored XML documents to the Data Warehouse	*IRP Version 2.X* • Web service to receive TDS Output xml file(s) from Vendor (will simulate API to Data Warehouse)	a. Capability to score student responses to individual assessment items b. Compliance with TDS Output file XML structure c. Correctness of data: - Student information - Item scoring - Test Scoring
D	Analysis of Adaptive Engine	*IRP Version 2.X* • Web Service to interface with vendor system, involving student responses to assessment items and resulting adaptations	a. Efficiency and accuracy of vendor adaptive engine analyzed and compared against "gold standard" adaptive engine using known psychometric properties

SmarterApp.org

In 2014, the Consortium launched the SmarterApp.org website. SmarterApp is a community of organizations devoted to collaboration on an open software suite for the support of educational assessment. At its launch, the SmarterApp website was a repository of specifications related to the SmarterApp assessment delivery platform. The collection of materials will progressively grow into a complete set of specifications and, eventually, the full source code set. The site will also grow to include feedback, discussion and eventually full collaboration on the code base. The initial participants in SmarterApp are organizations involved in the delivery of assessments to the Smarter Balanced member states. However, the software suite is suitable for a variety of other assessment subjects and contexts. All interested parties are invited to participate in developing or deploying the SmarterApp solutions. As of December 2014, SmarterApp includes information about Smarter Balanced systems architecture, technology specifications and related documents. SmarterApp also contains links to source code repositories for key components of the Smarter Balanced assessment system related to assessment construction (e.g., test authoring, test item bank), assessment delivery (e.g., secure browser, item scoring modules, item renderer), shared services (e.g., portal, permissions, core standards), software build and administration (e.g., administrative, shared build component), and the Technology Readiness Tool (TRT).

CHAPTER 8.
DATA PRIVACY AND SECURITY

The Consortium is committed to protecting the privacy, security, and confidentiality of student data. In September 2013, the Governing States adopted the policy that, "Each member state retains control of its student-level data." In 2015, the Consortium worked with members to develop and approve student data privacy agreements consistent with the member's state laws and policies. In addition to these agreements, each member maintains its own data security guidelines. In this section, we discuss the Consortium's policies on data privacy and data security.

DATA PRIVACY

Smarter Balanced follows industry best practices for data privacy and adheres to all applicable federal laws and regulations that safeguard education records, including the Family Educational Rights and Privacy Act (FERPA). In this section, we discuss the types of personally identifiable information that will and will not be collected as well as the organizations with whom we will and will not share information.

Personally Identifiable Information

The Consortium will collect and maintain a minimum amount of student information necessary to ensure that the assessments are accurate and fair. The Consortium will never collect extraneous personally identifiable information, such as Social Security numbers, home addresses, or medical information. We will not release nor will we sell personally identifiable information.

Information Collected

In order to make sure that the assessments are accurate and fair for all students, states will submit the following information to Smarter Balanced:

- An identification number (the Consortium recommends that this be different from the state's official unique student identifier so that only the state can link back to a student's official education record);
- Race/ethnicity, gender, grade level, school attended;
- Student eligibility for English language development services, or special education services provided to the student;
- Student eligibility for Title I compensatory programs;
- Smarter Balanced test scores, achievement levels, and responses to test questions.

If a state elects to have the Consortium generate reports of student assessment results, the consortium will also collect student name and date of birth. States may choose to manage that function through another organization. In either case, states can impose guidelines to safeguard those data.

Information Not Collected

The Consortium will never collect identifying information (such as student name and date of birth) unless specifically directed to do so by a member state. Further, they will not collect information unnecessary to the assessment system, such as:

- Names of parents;
- Parent or student email address;
- Telephone numbers;
- Parent or student Social Security numbers;

- Parent or student addresses;
- Parent or student medical information.

U.S. Department of Education

The Consortium will not share student-level information with the U.S. Department of Education. Further, state reporting to the U.S. Department of Education will not increase or change among Consortium members. Members will only share the same information required by the No Child Left Behind Act of 2001.

In addition, Smarter Balanced will comply with UCLA policies regarding the use of data, FERPA, and state and local requirements regarding privacy. FERPA, along with three other federal statutes, prohibits the federal government from establishing a database with personally identifiable student information.

On January 23, 2014, State Chiefs from all states involved in a consortia assessment wrote a joint letter to Secretary of Education Arne Duncan confirming that neither consortium will share personally identifiable information about K-12 students with the U.S. Department of Education or any federal agency. On March 13, 2014, Secretary Arne Duncan responded to the State Chiefs and confirmed that the two consortia are not required to share student-level information with the Department of Education or any other federal agency.

Research Requests

The Consortium established a method by which data could be made available for research. Toward that end, Consortium members determined that data may be made available only upon approval by each member state from which data is requested. Each member state will review such requests monthly. If such a request is approved by member states, the Consortium will provide a data file where individual students cannot be identified.

Smarter Balanced will comply with UCLA policies regarding the use of data, the Family Educational Rights and Privacy Act (FERPA), and state and local requirements regarding privacy.

DATA SECURITY

The Consortium follows industry best practice for securing the storage and transmission of student records. These measures include, but are not limited to:

- All student data are encrypted when in transit and when stored in an accessible data system;
- Data backups are encrypted and physically secured;
- Audit logs of user activity are kept including management of accounts, granting of permissions, proctoring of exams and generation of reports;
- Smarter Balanced never accepts social security numbers for any individual;
- Data centers are physically secured against unauthorized access;
- Smarter Balanced has hired a third party to perform a security audit on the Smarter Balanced Data Warehouse to ensure the student data is properly protected.

CHAPTER 9.
FINAL FISCAL REPORT

TABLE 9.1
OVERALL BUDGET SUMMARY (AGGREGATE OF LEVEL 1 MODULES, LEVEL 2 MODULES, AND SUPPLEMENTAL AWARD)

Budget Category	Approved Budget[1]	Grant award – June 30, 2011	July 1, 2011 – June 30, 2012	July 1, 2012 – June 30, 2013	July 1, 2013 –June 30, 2014	July 1, 2014 –March 31, 2015	TOTAL FOR THE GRANT
1. Personnel	$1,906,576	$69,084	$304,556	$556,194	$671,648	$274,818	$1,876,300
2. Fringe Benefits	$548,388	$15,391	$73,134	$137,838	$164,024	$146,298	$536,685
3. Travel	$1,804,327	$5,932	$108,266	$408,938	$348,645	$340,900	$1,212,681
4. Equipment	$0	$0	$0	$0	$0	$0	$0
5. Supplies	$80,318	$3,425	$29,739	$25,680	$10,231	$4,780	$73,855
6. Contractual	$170,532,389	$2,257,788	$13,906,773	$33,124,379	$79,009,548	$42,164,379	$170,462,867
7. Training Stipends	$0	$0	$0	$0	$0	$0	$0
8. Other	$0	$0	$0	$0	$0	$0	$0
9. Total Direct Costs (Lines 1-8)	$174,871,998	$2,351,620	$14,422,467	$34,253,028	$80,204,096	$42,931,175	$174,162,386
10. Indirect Costs	$977,541	$17,381	$99,878	$192,572	$323,963	$167,317	$801,111
11. Total Costs funded by the Race to the Top Assessment grant (Lines 9-10)	$175,849,539	$2,369,001	$14,522,345	$34,445,600	$80,528,059	$43,098,492	$174,963,497
12. Additional federal funds from other grants used for consortium work	$0	$0	$0	$0	$0	$0	$0
13. Non-federal funds used for consortium work	$3,969,367	$2,704	$714,456	$1,051,858	$1,218,558	$629,509	$3,617,085
14. Total cost of the consortium's work (sum of lines 12, 13, and 14)	$179,818,906	$2,371,705	$15,236,800	$35,497,459	$81,746,617	$43,728,002	$178,580,582

Note: *Line 11 should equal the total amount awarded to the consortium (including the original and supplemental award).*

[1] Amounts are based upon the Governing States approved budget from June 2012, and any subsequent approved budget adjustment that occurred from June 2012 through the end of the grant period.

TABLE 9.2
BUDGET SUMMARY BY MODULE, LEVEL 1 MODULES

Budget Module	Approved Budget[1]	Grant award – June 30, 2011[2]	July 1, 2011 – June 30, 2012[2]	July 1, 2012 – June 30, 2013[2]	July 1, 2013 –June 30, 2014[2]	July 1, 2014 –March 31, 2015[2]	TOTAL FOR THE GRANT
1. Governance	$14,833,470	$1,575,240	$3,544,793	$4,016,128	$4,914,075	$4,117,893	$18,168,130
2. Assessment Design	$101,672,951	$0	$6,298,976	$18,478,408	$52,680,923	$21,920,379	$99,378,686
3. System Design	$106,500	$0	$54,000	$45,000	$0	$7,500	$106,500
4. Research & Evaluation	$9,886,735	$0	$756,543	$1,374,819	$3,811,672	$3,762,473	$9,705,506
5. Professional Capacity & Outreach	$15,326,875	$776,379	$2,063,639	$936,095	$10,274,741	$1,259,321	$15,310,176
6. Technology	$31,705,874	$0	$1,704,515	$9,351,577	$8,491,127	$11,860,716	$31,407,935
7. Higher Education Engagement	$1,339,593	$0	$0	$51,001	$31,559	$2,893	$85,453
8. Total of Level 1 Budget Modules (sum of lines 1-7)	$174,871,998	$2,351,620	$14,422,467	$34,253,028	$80,204,096	$42,931,175	$174,162,386
9. Indirect Costs	$977,541	$17,381	$99,878	$192,572	$323,963	$167,317	$801,111
10. Total Costs funded by the Race to the Top Assessment grant (Lines 8-9)	$175,849,539	$2,369,001	$14,522,345	$34,445,600	$80,528,059	$43,098,492	$174,963,497

Note: *Because the activities included in the Supplemental Award may, by definition, supplement the Module 1 activities, the consortium may have included some of the Supplemental Award funding in the corresponding Module 1 activity above for the purposes of effectively tracking expenditures.*

[1] Amounts are based upon the Governing States approved budget from June 2012, and any subsequent approved budget adjustment that occurred from June 2012 through the end of the grant period.

[2] Expended amounts represent all payments WA has completed. Reimbursements made to WestEd for funds spent on behalf of the Consortium are recorded against the WestEd contract (under Governance) and are not recorded against the actual budget category for which the expenditures were incurred. This may cause variances to appear when comparing budgeted expenditures to actual expenditures. The Consortium's internal reporting system provides an integrated view of spending across Washington OSPI and WestEd and represents an alignment of all expenditures to the categories against which they were incurred.

TABLE 9.3
BUDGET SUMMARY BY MODULE, LEVEL 2 MODULES

Budget Module	Approved Budget[1]	Grant award – June 30, 2011	July 1, 2011 – June 30, 2012	July 1, 2012 – June 30, 2013	July 1, 2013 –June 30, 2014	July 1, 2014 –March 31, 2015	TOTAL FOR THE GRANT
1. Translation	***Translation budgets and expenditures are reported as part of the Assessment Design Budget Module, Table 2.***						
2. Total of Level 2 Budget Modules							

TABLE 9.4
BUDGET SUMMARY BY MODULE, SUPPLEMENTAL BUDGET MODULES

Budget Module	Approved Budget	Grant award – June 30, 2011	July 1, 2011 – June 30, 2012	July 1, 2012 – June 30, 2013	July 1, 2013 –June 30, 2014	July 1, 2014 –March 31, 2015	TOTAL FOR THE GRANT
1. Curriculum Materials	***Supplemental budgets and expenditures are reported in alignment with their corresponding budget module listed in Table 2.***						
2. Formative Processes and Tools/Professional Development							
3. Implementation of Common Core Systems							
4. Communications							
5. Aligning Assessments to Readiness							
6. Support for Technology Transitions							
7. Technical Assistance Meetings							
8. PMP Management							
9. Unduplicated Total of Supplemental Budget Modules (sum of lines 1-8)							

Note: Because the activities included in the Supplemental Award may, by definition, supplement the Module 1 activities, the consortium may have included some of the Supplemental Award funding in the corresponding Module 1 activity in Table 2. As a result, the total in this table may be less than the Supplemental Award provided to the consortium.

The grand total of tables 2, 3, and 4 must equal the total amount awarded to the consortium (including the original and supplemental award) – line 11 from table 1.

TABLE 9.5
MODULE-LEVEL BUDGET: *GOVERNANCE* U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Category	Approved Budget[1]	Grant award – June 30, 2011[2]	July 1, 2011 – June 30, 2012[2]	July 1, 2012 – June 30, 2013[2]	July 1, 2013 –June 30, 2014[2]	July 1, 2014 –March 31, 2015[2]	TOTAL FOR THE GRANT
1. Personnel	$783,548	$69,084	$175,362	$187,246	$219,367	$102,213	$753,272
2. Fringe Benefits	$299,159	$15,391	$43,229	$49,726	$57,306	$121,803	$287,455
3. Travel	$1,123,155	$4,553	$79,096	$312,246	$266,822	$153,132	$815,848
4. Equipment	$0	$0	$0	$0	$0	$0	$0
5. Supplies	$80,318	$3,425	$29,739	$25,680	$10,231	$4,780	$73,855
6. Contractual	$12,547,290	$1,482,788	$3,217,367	$3,441,230	$4,360,349	$3,735,965	$16,237,699
7. Training Stipends	$0	$0	$0	$0	$0	$0	$0
8. Other	$0	$0	$0	$0	$0	$0	$0
9. Total Direct Costs (Lines 1-8)	$14,833,470	$1,575,240	$3,544,793	$4,016,128	$4,914,075	$4,117,893	$18,168,130
10. Indirect Costs[3]	NA	NA	NA	NA	NA	NA	NA
11. Total Costs (Lines 9-10)	$14,833,470	$1,575,240	$3,544,793	$4,016,128	$4,914,075	$4,117,893	$18,168,130

Note: *This table should include RTTA grant funds only. Make as many copies of this Module-Level Budget Table as needed. The consortium should submit a separate table for each module listed in tables 2, 3, and 4.*

[1] Amounts are based upon the Governing States approved budget from June 2012, and any subsequent approved budget adjustment that occurred from June 2012 through the end of the grant period.

[2] Expended amounts represent all payments WA has completed. Reimbursements made to WestEd for funds spent on behalf of the Consortium are recorded against the WestEd contract (under Governance) and are not recorded against the actual budget category for which the expenditures were incurred. This may cause variances to appear when comparing budgeted expenditures to actual expenditures. The Consortium's internal reporting system provides an integrated view of spending across Washington OSPI and WestEd and represents an alignment of all expenditures to the categories against which they were incurred.

[3] Washington indirect costs are not captured by module. Refer to Table 1 for a total of the indirects incurred.

TABLE 9.6
MODULE-LEVEL BUDGET: *ASSESSMENT DESIGN* U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Category	Approved Budget[1]	Grant award – June 30, 2011[2]	July 1, 2011 – June 30, 2012[2]	July 1, 2012 – June 30, 2013[2]	July 1, 2013 –June 30, 2014[2]	July 1, 2014 –March 31, 2015[2]	TOTAL FOR THE GRANT
1. Personnel	$0	$0	$0	$0	$0	$0	$0
2. Fringe Benefits	$0	$0	$0	$0	$0	$0	$0
3. Travel	$397,710	$0	$8,558	$5,793	$5,511	$162,015	$181,878
4. Equipment	$0	$0	$0	$0	$0	$0	$0
5. Supplies	$0	$0	$0	$0	$0	$0	$0
6. Contractual	$101,275,241	$0	$6,290,417	$18,472,615	$52,675,412	$21,758,364	$99,196,808
7. Training Stipends	$0	$0	$0	$0	$0	$0	$0
8. Other	$0	$0	$0	$0	$0	$0	$0
9. Total Direct Costs (Lines 1-8)	$101,672,951	$0	$6,298,976	$18,478,408	$52,680,923	$21,920,379	$99,378,686
10. Indirect Costs[3]	NA	NA	NA	NA	NA	NA	NA
11. Total Costs (Lines 9-10)	$101,672,951	$0	$6,298,976	$18,478,408	$52,680,923	$21,920,379	$99,378,686

[1] Amounts are based upon the Governing States approved budget from June 2012, and any subsequent approved budget adjustment that occurred from June 2012 through the end of the grant period.

[2] Expended amounts represent all payments WA has completed. Reimbursements made to WestEd for funds spent on behalf of the Consortium are recorded against the WestEd contract (under Governance) and are not recorded against the actual budget category for which the expenditures were incurred. This may cause variances to appear when comparing budgeted expenditures to actual expenditures. The Consortium's internal reporting system provides an integrated view of spending across Washington OSPI and WestEd and represents an alignment of all expenditures to the categories against which they were incurred.

[3] Washington indirect costs are not captured by module. Refer to Table 1 for a total of the indirects incurred.

TABLE 9.7
MODULE-LEVEL BUDGET: *SYSTEM DESIGN* U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Category	Approved Budget[1]	Grant award – June 30, 2011[2]	July 1, 2011 – June 30, 2012[2]	July 1, 2012 – June 30, 2013[2]	July 1, 2013 –June 30, 2014[2]	July 1, 2014 –March 31, 2015[2]	TOTAL FOR THE GRANT
1. Personnel	$0	$0	$0	$0	$0	$0	$0
2. Fringe Benefits	$0	$0	$0	$0	$0	$0	$0
3. Travel	$0	$0	$0	$0	$0	$0	$0
4. Equipment	$0	$0	$0	$0	$0	$0	$0
5. Supplies	$0	$0	$0	$0	$0	$0	$0
6. Contractual	$106,500	$0	$54,000	$45,000	$0	$7,500	$106,500
7. Training Stipends	$0	$0	$0	$0	$0	$0	$0
8. Other	$0	$0	$0	$0	$0	$0	$0
9. Total Direct Costs (Lines 1-8)	$106,500	$0	$54,000	$45,000	$0	$7,500	$106,500
10. Indirect Costs[3]	NA	NA	NA	NA	NA	NA	NA
11. Total Costs (Lines 9-10)	$106,500	$0	$54,000	$45,000	$0	$7,500	$106,500

[1] Amounts are based upon the Governing States approved budget from June 2012, and any subsequent approved budget adjustment that occurred from June 2012 through the end of the grant period.

[2] Expended amounts represent all payments WA has completed. Reimbursements made to WestEd for funds spent on behalf of the Consortium are recorded against the WestEd contract (under Governance) and are not recorded against the actual budget category for which the expenditures were incurred. This may cause variances to appear when comparing budgeted expenditures to actual expenditures. The Consortium's internal reporting system provides an integrated view of spending across Washington OSPI and WestEd and represents an alignment of all expenditures to the categories against which they were incurred.

[3] Washington indirect costs are not captured by module. Refer to Table 1 for a total of the indirects incurred.

TABLE 9.8
MODULE-LEVEL BUDGET: *RESEARCH & EVALUATION* U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Category	Approved Budget[1]	Grant award – June 30, 2011[2]	July 1, 2011 – June 30, 2012[2]	July 1, 2012 – June 30, 2013[2]	July 1, 2013 –June 30, 2014[2]	July 1, 2014 –March 31, 2015[2]	TOTAL FOR THE GRANT
1. Personnel	$375,695	$0	$100,690	$120,000	$123,708	$31,298	$375,695
2. Fringe Benefits	$87,366	$0	$23,021	$27,247	$30,157	$6,940	$87,366
3. Travel	$49,564	$0	$15,086	$8,267	$16,488	$4,152	$43,993
4. Equipment	$0	$0	$0	$0	$0	$0	$0
5. Supplies	$0	$0	$0	$0	$0	$0	$0
6. Contractual	$9,374,111	$0	$617,747	$1,219,305	$3,641,318	$3,720,083	$9,198,453
7. Training Stipends	$0	$0	$0	$0	$0	$0	$0
8. Other	$0	$0	$0	$0	$0	$0	$0
9. Total Direct Costs (Lines 1-8)	$9,886,735	$0	$756,543	$1,374,819	$3,811,672	$3,762,473	$9,705,506
10. Indirect Costs[3]	NA	NA	NA	NA	NA	NA	NA
11. Total Costs (Lines 9-10)	$9,886,735	$0	$756,543	$1,374,819	$3,811,672	$3,762,473	$9,705,506

[1] Amounts are based upon the Governing States approved budget from June 2012, and any subsequent approved budget adjustment that occurred from June 2012 through the end of the grant period.

[2] Expended amounts represent all payments WA has completed. Reimbursements made to WestEd for funds spent on behalf of the Consortium are recorded against the WestEd contract (under Governance) and are not recorded against the actual budget category for which the expenditures were incurred. This may cause variances to appear when comparing budgeted expenditures to actual expenditures. The Consortium's internal reporting system provides an integrated view of spending across Washington OSPI and WestEd and represents an alignment of all expenditures to the categories against which they were incurred.

[3] Washington indirect costs are not captured by module. Refer to Table 1 for a total of the indirects incurred.

TABLE 9.9
MODULE-LEVEL BUDGET: *PROFESSIONAL CAPACITY & OUTREACH* U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Category	Approved Budget[1]	Grant award – June 30, 2011[2]	July 1, 2011 – June 30, 2012[2]	July 1, 2012 – June 30, 2013[2]	July 1, 2013 –June 30, 2014[2]	July 1, 2014 –March 31, 2015[2]	TOTAL FOR THE GRANT
1. Personnel	$747,333	$0	$28,504	$248,948	$328,573	$141,308	$747,333
2. Fringe Benefits	$161,863	$0	$6,884	$60,864	$76,560	$17,556	$161,863
3. Travel	$77,544	$1,379	$4,268	$12,899	$23,592	$18,707	$60,845
4. Equipment	$0	$0	$0	$0	$0	$0	$0
5. Supplies	$0	$0	$0	$0	$0	$0	$0
6. Contractual	$14,340,135	$775,000	$2,023,983	$613,385	$9,846,016	$1,081,751	$14,340,135
7. Training Stipends	$0	$0	$0	$0	$0	$0	$0
8. Other	$0	$0	$0	$0	$0	$0	$0
9. Total Direct Costs (Lines 1-8)	$15,326,875	$776,379	$2,063,639	$936,095	$10,274,741	$1,259,321	$15,310,176
10. Indirect Costs[3]	NA	NA	NA	NA	NA	NA	NA
11. Total Costs (Lines 9-10)	$15,326,875	$776,379	$2,063,639	$936,095	$10,274,741	$1,259,321	$15,310,176

[1] Amounts are based upon the Governing States approved budget from June 2012, and any subsequent approved budget adjustment that occurred from June 2012 through the end of the grant period.

[2] Expended amounts represent all payments WA has completed. Reimbursements made to WestEd for funds spent on behalf of the Consortium are recorded against the WestEd contract (under Governance) and are not recorded against the actual budget category for which the expenditures were incurred. This may cause variances to appear when comparing budgeted expenditures to actual expenditures. The Consortium's internal reporting system provides an integrated view of spending across Washington OSPI and WestEd and represents an alignment of all expenditures to the categories against which they were incurred.

[3] Washington indirect costs are not captured by module. Refer to Table 1 for a total of the indirects incurred.

TABLE 9.10
MODULE-LEVEL BUDGET: *TECHNOLOGY* U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Category	Approved Budget[1]	Grant award – June 30, 2011[2]	July 1, 2011 – June 30, 2012[2]	July 1, 2012 – June 30, 2013[2]	July 1, 2013 –June 30, 2014[2]	July 1, 2014 –March 31, 2015[2]	TOTAL FOR THE GRANT
1. Personnel	$0	$0	$0	$0	$0	$0	$0
2. Fringe Benefits	$0	$0	$0	$0	$0	$0	$0
3. Travel	$31,196	$0	$1,258	$18,732	$4,673	$0	$24,663
4. Equipment	$0	$0	$0	$0	$0	$0	$0
5. Supplies	$0	$0	$0	$0	$0	$0	$0
6. Contractual	$31,674,678	$0	$1,703,258	$9,332,845	$8,486,453	$11,860,716	$31,383,272
7. Training Stipends	$0	$0	$0	$0	$0	$0	$0
8. Other	$0	$0	$0	$0	$0	$0	$0
9. Total Direct Costs (Lines 1-8)	$31,705,874	$0	$1,704,515	$9,351,577	$8,491,127	$11,860,716	$31,407,935
10. Indirect Costs[3]	NA	NA	NA	NA	NA	NA	NA
11. Total Costs (Lines 9-10)	$31,705,874	$0	$1,704,515	$9,351,577	$8,491,127	$11,860,716	$31,407,935

[1] Amounts are based upon the Governing States approved budget from June 2012, and any subsequent approved budget adjustment that occurred from June 2012 through the end of the grant period.

[2] Expended amounts represent all payments WA has completed. Reimbursements made to WestEd for funds spent on behalf of the Consortium are recorded against the WestEd contract (under Governance) and are not recorded against the actual budget category for which the expenditures were incurred. This may cause variances to appear when comparing budgeted expenditures to actual expenditures. The Consortium's internal reporting system provides an integrated view of spending across Washington OSPI and WestEd and represents an alignment of all expenditures to the categories against which they were incurred.

[3] Washington indirect costs are not captured by module. Refer to Table 1 for a total of the indirects incurred.

TABLE 9.11
MODULE-LEVEL BUDGET: *HIGHER EDUCATION ENGAGEMENT* U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Category	Approved Budget[1]	Grant award – June 30, 2011[2]	July 1, 2011 – June 30, 2012[2]	July 1, 2012 – June 30, 2013[2]	July 1, 2013 –June 30, 2014[2]	July 1, 2014 –March 31, 2015[2]	TOTAL FOR THE GRANT
1. Personnel	$0	$0	$0	$0	$0	$0	$0
2. Fringe Benefits	$0	$0	$0	$0	$0	$0	$0
3. Travel	$125,158	$0	$0	$51,001	$31,559	$2,893	$85,453
4. Equipment	$0	$0	$0	$0	$0	$0	$0
5. Supplies	$0	$0	$0	$0	$0	$0	$0
6. Contractual	$1,214,435	$0	$0	$0	$0	$0	$0
7. Training Stipends	$0	$0	$0	$0	$0	$0	$0
8. Other	$0	$0	$0	$0	$0	$0	$0
9. Total Direct Costs (Lines 1-8)	$1,339,593	$0	$0	$51,001	$31,559	$2,893	$85,453
10. Indirect Costs[3]	NA	NA	NA	NA	NA	NA	NA
11. Total Costs (Lines 9-10)	$1,339,593	$0	$0	$51,001	$31,559	$2,893	$85,453

Note: WestEd is the fiscal agent for the Higher Education Funds and thus any payments made by Washington OSPI to reimburse WestEd's expenditures are recorded against the WestEd PMP contract in Governance.

[1] Amounts are based upon the Governing States approved budget from June 2012, and any subsequent approved budget adjustment that occurred from June 2012 through the end of the grant period.

[2] Expended amounts represent all payments WA has completed. Reimbursements made to WestEd for funds spent on behalf of the Consortium are recorded against the WestEd contract (under Governance) and are not recorded against the actual budget category for which the expenditures were incurred. This may cause variances to appear when comparing budgeted expenditures to actual expenditures. The Consortium's internal reporting system provides an integrated view of spending across Washington OSPI and WestEd and represents an alignment of all expenditures to the categories against which they were incurred.

[3] Washington indirect costs are not captured by module. Refer to Table 1 for a total of the indirects incurred.

APPENDIX A.
LIST OF THE CONSORTIUM'S RFPS AND CONTRACTED VENDORS

RFP ID	RFP NAME	CONTRACTED VENDOR
RFP-00	Project Management Partner	*WestEd*
RFP-01	Communications Services Provider	*GMMB*
RFP-02	IT Readiness	*Pearson*
RFP-03	IT Systems Architecture	*Measured Progress*
RFP-04	Item Specifications	Measured Progress/ETS
RFP-05	Psychometric Services	ETS
RFP-06	Accessibility and Accommodations Policy Guidelines	Measured Progress
RFP-07	Item Authoring/Item Pool	Pacific Metrics
RFP-08	Item/Task Materials Development	Measured Progress
RFP-09	Test and CAT Specifications	ETS
RFP-11/18/20	Test Engine Development/Test Engine Delivery/CAT Simulations	AIR
RFP-12	Initial Achievement Level Descriptors	CTB/McGraw-Hill
RFP-13	Translations	Tri-Lin
RFP-14	Item/Task Writing/Review for Pilot	CTB/McGraw-Hill
RFP-15	Report Development	Amplify Insight (f/k/a Wireless Generation)
RFP-16/17	Item/Task Development for Field Test and Scoring for Pilot and Field Test	CTB/McGraw-Hill
RFP-19a	Test Administration for Pilot Test	AIR
RFP-19b	Test Administration for Field Test	ETS
RFP-21	Standard Setting and Communication Services	Measurement Incorporated
RFP-23	Formative Assessment	Amplify Insight (f/k/a Wireless Generation)
RFP-24	Sustainability Cost Modeling	Assessment Solutions Group