CRESST REPORT 851

ASSESSMENT FOR DEEPER LEARNING IN CCSS: PROGRESS REPORT ON THE STATUS OF SMARTER BALANCED AND PARCC ASSESSMENT CONSORTIA

JULY 2013

Joan L. Herman
Rebecca E. Buschang
Deborah La Torre Matrundola
Jia Wang

National Center for Research
on Evaluation, Standards, & Student Testing
UCLA | Graduate School of Education & Information Studies

Assessment for Deeper Learning in CCSS: Progress Report on the Status of Smarter Balanced and PARCC Assessment Consortia

CRESST Report 851

Joan L. Herman, Rebecca E. Buschang, Deborah La Torre Matrundola, and Jia Wang
CRESST/University of California, Los Angeles

July 2013

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported by grant number 2011-7086 from The William and Flora Hewlett Foundation with funding to National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

The findings and opinions expressed in this report are those of the authors and do not necessarily reflect the positions or policies of The William and Flora Hewlett Foundation.

To cite from this report, please use the following as your APA reference: Herman, J. L., Buschang, R. E., La Torre Matrundola, D., & Wang, J. (2013). *Assessment for deeper learning in CCSS: Progress report on the status of Smarter Balanced and PARCC Assessment Consortia* (CRESST Report 851). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

Abstract 1
Introduction 1
Assessment Development 1
Smarter Balanced Assessment Development 2
PARCC Assessment Development 3
Achievement Level Descriptions 4
Accessibility and Accommodations Guidelines 8
Technology Readiness 10
Comparing the Smarter Balanced and PARCC Content Specifications 12
Expectations for Deeper Learning Representation 15
CCSS Depth of Knowledge Designations: English Language Arts 16
CCSS Depth of Knowledge Designations: Mathematics 17
Depth of Knowledge Representation in Respected National and International Assessments 19
Summary 22
Conclusion and Recommendations 22
References 25
Appendix: Smarter Balanced and PARCC Content Specifications 27

ASSESSMENT FOR DEEPER LEARNING IN CCSS: PROGRESS REPORT ON THE STATUS OF SMARTER BALANCED AND PARCC ASSESSMENT CONSORTIA

Joan L. Herman, Rebecca E. Buschang, Deborah La Torre Matrundola, and Jia Wang
CRESST/University of California, Los Angeles

Abstract

This report examines progress made by the Smarter Balanced Assessment Consortium (Smarter Balanced) and the Partnership for Assessment of Readiness for College and Careers (PARCC) through the spring of 2013. This is done by examining efforts concerning the development of their respective assessments, achievement level descriptors, accessibility and accommodations guidelines, and technology guidelines. This report also projects expectations for deeper learning in the consortia's summative assessments and compares these results to related studies. Results indicate that both consortia made substantial progress in their assessment development over the course of the year. In addition, analyses show a range of depth of knowledge (DOK) expectations across those aligned to the Common Core State Standards (CCSS) as well as respected national and international tests. Based on these analyses, this study recommends benchmarks for deeper learning (DOK3 and DOK4) of 33% of total score points in mathematics, 33% in English language arts (ELA) at the elementary level, and 50% in ELA at the secondary level.

Introduction

This report summarizes CRESST's efforts over the past year to monitor and support the Smarter Balanced Assessment Consortium (Smarter Balanced) and the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment consortium. Briefly we will describe Smarter Balanced's and PARCC's efforts in assessment development, their creation of achievement level descriptors, their efforts to make accessible tests and accommodate special needs, and the tools they've both developed to assess states' technology readiness. In addition, based on available data, we compare the content foci planned for each consortium's summative assessments and recommend preliminary targets for the representation of deeper learning in these tests.

Assessment Development

Both Smarter Balanced and PARCC made progress on their ambitious test development schedules and, in doing so, both added additional oversight to support the rigor of their item and task development processes. While Smarter Balanced remained on schedule in conducting their pilot testing this spring, PARCC slowed down their pilot testing plans to concentrate on item and task development. Both consortia plan field tests during spring 2014.

Smarter Balanced Assessment Development

Smarter Balanced released sample assessment items and performance tasks, with rubrics when appropriate, for multiple grade levels. These include items incorporating technology enhancements and tools. Item development for the samples is based on content specifications that outline assessment targets with depth of knowledge (DOK) ranges for each grade level in mathematics and English language arts. While Smarter Balanced did not list the DOK for the samples, the content specifications do indicate that all performance tasks are associated with DOK4 and often with DOK3. As such, it is easy to assume that at least the three English language arts (ELA) and three mathematics performance tasks represent a higher DOK.

Preliminary test blueprints for Smarter Balanced were released in November 2012. The blueprints describe the content for ELA and mathematics as well as the number of items or percentage of points, item types, and DOK for each assessment target for the different grade bands (e.g., 3–5, 6–9, and 11). These blueprints will guide development of items and performance tasks, pilot tests, score reporting, standards setting, and research. Test blueprints may be refined after the pilot and field tests. Based on current test blueprints and released sample items, Smarter Balanced performance tasks will encapsulate DOK4 and are likely to include DOK3. The blueprints for the computer adaptive testing (CAT) portion of the end-of-year, summative assessment also includes representation of DOK3. However, because the blueprints contain complex rules for sampling assessment targets for any given individual, it is difficult to calculate the likelihood or intensity of higher DOK levels for any given student.

The blueprints are organized relative to the major claims guiding test development, specify all assessment targets—many of which may be assessed at multiple DOK levels—and then provide probabilities that certain clusters of targets will be sampled. The bottom line is that, given these rules, the range of possible DOK representation is very large. For example, based on the blueprint for the ELA CAT for Grades 3–5, anywhere from 6% to 72% of items for a given individual could be at DOK3 and DOK4. The rules underlying CAT item selection will be further refined in the CAT algorithms, which are still under development. To meet the Hewlett Foundation's deeper learning goals, it will be important that these algorithms take further account of DOK and assure that CAT administrations for all students, regardless of ability level, include adequate representation of higher levels.

Smarter Balanced recently completed pilot testing of their assessment with randomly selected schools from across the consortium, representing students in Grades 3–11. Results from the pilot, which included more than 650,000 students in approximately 5,000 schools and 21 states, will inform the design of the spring 2014 field test, as well as the operational design. For

example, results will be used to fine-tune timing requirements, examine psychometric characteristics of items, and determine generalizability of performance task results. Furthermore, the piloting has already encouraged the consortium to strengthen their quality control procedures for assuring the rigor of assessment items and tasks through the convening of an expert oversight panel and instituting detailed qualification and review procedures for their assessment development contractor and subcontractors.

Because more schools volunteered for the pilot test than could be accommodated in the rigorous design, Smarter Balanced originally planned to involve additional volunteer sites in an unofficial pilot. Instead, the consortium developed and released online practice tests for Grades 3–8 and 11 in both ELA and mathematics in May. These tests include multiple choice, open-ended, and technology-enhanced items in both content areas as well as technology-based performance tasks for ELA. Braille versions are in development with mathematics tests already available through a secure portal. In addition, the consortium plans to release classroom-based and technology-based performance tasks in mathematics, classroom-based performance tasks for ELA, scoring rubrics, and other accommodation and accessibility tools by fall 2013. The release of practice tests to the public instead of only the volunteer sample is intended to increase awareness of the assessments, and provide a wider audience with a comprehensive look at what the assessments will look like (see http://www.smarterbalanced.org/assessments/practice-and-training-tests/). In addition to the pilot test activities, the consortium will continue to have K–12 teachers and higher education faculty develop summative test items, interim test items, and performance tasks through the summer.

PARCC Assessment Development

PARCC released an initial set of prototype items and rubrics for a variety of grades between 3 and 11 in August 2012. These include some end-of-year and performance-based assessment components that require strategic or extended thinking, as required for classification at DOK3 and DOK4. As with Smarter Balanced, development of these items is based on the content specifications for both ELA and mathematics developed by the consortium (see http://www.parcconline.org/samples/item-task-prototypes). Item prototypes and rubrics are periodically being released with an additional set currently in development for release in summer 2013.

PARCC also recently released their ELA and mathematics assessment blueprints for their mid-year, end-of-year, and performance-based assessments. For ELA, the blueprints currently outline the number of text passages, items, and the maximum points possible for each task type for each of these assessments. Although cognitive complexity is not yet specified on the public

blueprints, PARCC plans to integrate this concern in part by using existing published texts that are content rich and challenging. This approach fits with the consortium's belief that text and task complexity are integrated rather than individual structures.

For mathematics, the blueprints outline eligible test content relative to specific evidence statements and corresponding mathematical practices that may be addressed in each grade-level assessment. The blueprints also specify the number of Type I, Type II, and Type III items, and the points associated with each item that are to constitute each test form. Type I items, which will be used on both the end-of-year and performance-based components, focus on students being able to use concepts, skills, and procedures, and draw principally on selected response and technology-enhanced items that can be automatically scored. In contrast, the other two types of items, which will be included solely in the performance-based assessments, ask students to express their mathematical reasoning (Type II) or model real-world contexts or scenarios (Type III). These last two types should address DOK3 and DOK4.

PARCC recently began small-scale item research and is preparing for a field test in spring 2014. This research included a trial run of items and questions with 2,300 students in six states which was completed earlier this year, with an additional 4,800 students in four states slated to participate during summer 2013. A new management consultant has been brought in to oversee the quality and progress of item and task development. The consortium intends to finalize their participation guidelines and selection of schools for the field test this summer (see http://www.parcconline.org/field-test), and release their revised technology requirements and administration manual in the fall and winter. Furthermore, practice tests of representative items at each grade level will be made available to all PARCC state students in spring 2014.

Achievement Level Descriptions

Both Smarter Balanced and PARCC have released documentation describing their grade-level performance expectations for being designated "on track" or "college-ready" in ELA and mathematics. Smarter Balanced uses the term achievement level descriptors (ALDs) to describe these expectations, and PARCC uses the term performance level descriptors (PLDs). Both consortia define their performance expectations to indicate relative readiness for college and career (see http://www.smarterbalanced.org/achievement-level-descriptors-and-college-readiness and http://www.parcconline.org/plds).

Smarter Balanced drew on their content specifications to develop the ALDs and are using them to guide different phases of test development. The initial public drafts were released in October 2012, revised based on public comment, and approved by the Governing States in April 2013. Table 1 describes the use, purpose, and audience for the four types of ALDs. The

descriptors will also communicate to parents, teachers, and other stakeholders about the general level at which students are performing relative to college readiness. The ALDs have four levels of performance, each including a keyword (i.e., minimal, partial, adequate, and thorough) that is consistent across the different types of ALDs. Currently the consortium anticipates using Level 3 as the minimum level indicating students are on track or content ready for college. While Smarter Balanced does not list DOK levels, they are implied in the descriptors through the specification of increasingly complex skills across the levels and, in regards to ELA, the ability to read increasingly complex texts (i.e., low, moderate, moderate-to-high, and unusually high).

Smarter Balanced notes that the ALDs may be revised further based on field test results in early 2014. Next steps for Smarter Balanced include using the ALDs in the validation of cut-scores differentiating each level, especially for the Grade 11 scores, and encouraging postsecondary institutions to use these scores as evidence of readiness for credit-bearing courses. In addition, the consortium is working to develop operational definitions and a framework for career readiness in contrast to college readiness.

Table 1

Types of Smarter Balanced Achievement Level Descriptors

Type	Use	Purpose	Audience
Policy	Test development and conceptualization	Set tone for the rigor of performance standards expected by sponsoring agency	Policymakers
Range (Content)	Item-writing guidance	Define the range of possible performance for students within a specific achievement level in order to reflect expected learning progressions for a content area at a specific grade level	Item writers and test developers
Threshold	Cut-score recommendation and standard-setting guidance	Define expectations for student performance at a specific grade level and content area	Standard-setting panelists
Reporting	Test-score interpretation	Describe the knowledge, skills, and processes that test takers demonstrate and indicate the knowledge and skills that must be developed to attain the next level of achievement	Stakeholders, such as parents, students, teachers, K–12 leaders, and higher education officials

Note. Adapted from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/11/Smarter-Balanced-ELA-Literacy-ALDs.pdf

PARCC released the initial framework for their policy-level PLDs in October 2012. As can be seen in Table 2, the policy-level PLDs include both policy claims and general content claims,

which are each associated with the five performance levels established by the consortium to indicate level of student command (i.e., minimal, partial, moderate, strong, and distinguished). Grade- and subject-specific PLDs were then developed through an iterative process with state and local leaders, revised based on public comment in April 2013, and approved by the Governing Board in late June. While the PLDs do not explicitly note DOK, PARCC reports that the operational work groups took text complexity, range of accuracy, and quality of evidence into account when setting the levels for ELA and used their cognitive complexity framework (i.e., content, practices, stimulus material, response mode, and processing demand) when setting the levels for mathematics.

Table 2

Types of PARCC Performance Level Descriptors

Type	Use	Purpose	Audience
Policy-level PLDs			
Policy claims	Item development and standard setting	Describe the educational implications for students at a particular performance level	Policymakers
General content claims	Item development and standard setting	Describe, in broad terms, the knowledge, skills, and practices students performing at a given performance level are able to demonstrate. General content claims are applicable to any grade level	Item writers and test developers
Grade- and subject-specific PLDs	Item development and standard setting; inform development of curricular and instructional materials; communicate expectations for college and career readiness	Describe the knowledge, skills, and practices students performing at a given performance level/course and grade level are able to demonstrate (e.g., Grade 4 ELA, Grade 8 Math, and Algebra 1)	Item writers and test developers, educators, parents, and students

In addition to using different broad categories of descriptors, PARCC differs in the number of performance levels being used. The consortium is currently using five levels of performance, with the fourth planned as the minimum to classify students as being college and career ready in a content area. Each of these will be associated with cut-scores in order to quantify performance. While PARCC also uses keywords in their definitions for each performance level (i.e., minimal, partial, moderate, strong, and distinguished), they are not integrated into the general content claims or into the grade- and subject-specific PLDs. In addition, the consortium decided to use a standardized definition of Level 1 for all PLDs rather than providing separate descriptors like they did for Levels 2 through 5 (see Table 3). Next steps for PARCC have not been made publicly available.

Table 3

Sample PARCC Grade-Specific PLDs and Smarter Balanced Range ALDs for Grade 11 Writing

Level	PARCC PLDs	Smarter Balanced ALDs
Focus	Writing Knowledge of Language and Conventions: Students demonstrate knowledge of conventions and other important elements of language.	EDIT/CLARIFY: Apply or edit grade-appropriate grammar usage and mechanics to clarify a message and edit narrative, informational, and argumentative texts.
Level 1	Students performing at this level demonstrate a minimal command of the knowledge, skills, and practices embodied by the Common Core State Standards assessed at their grade level.	Students should be able to provide minimal evidence that they can apply or edit the conventions of grade-appropriate, Standard English grammar usage and mechanics to clarify a message and edit narrative, informational, and persuasive/argument texts.
Level 2	Demonstrates limited command of the conventions of Standard English. There are multiple errors in grammar and usage demonstrating minimal control over language. There are multiple distracting errors in grammar and usage that sometimes impede understanding.	Students should be able to provide partial evidence that they can apply or edit the conventions of grade-appropriate, Standard English grammar usage and mechanics to clarify a message and edit narrative, informational, and persuasive/argument texts.
Level 3	Demonstrates inconsistent command of the conventions of Standard English. There are a few patterns of errors in grammar and usage that may occasionally impede understanding.	Students should be able to provide adequate evidence that they can apply and edit with consistent understanding the conventions of grade-appropriate, Standard English grammar usage and mechanics to clarify a message and edit narrative, informational, and persuasive/argument texts.
Level 4	Demonstrates command of the conventions of Standard English consistent with edited writing. There may be a few distracting errors in grammar and usage, but meaning is clear.	Students should be able to provide thorough evidence that they can apply and edit with advanced understanding the conventions of grade-appropriate, Standard English grammar usage and mechanics to clarify a message and edit narrative, informational, and persuasive/argument texts.
Level 5	Demonstrates command of the conventions of Standard English consistent with effectively edited writing. Though there may be a few minor errors in grammar and usage, meaning is clear throughout the response.	

Table 3 shows examples of PARCC grade-specific PLDs and Smarter Balanced range ALDs for Grade 11 writing. The PLD is aligned to the PARCC subclaim on writing knowledge of language and conventions, while the ALD is aligned to the Smarter Balanced target on editing and clarifying writing. In both cases, the emphasis is on specific content (e.g., knowledge, skills, and practices) that students will need to demonstrate concerning the conventions of Standard English grammar to obtain a specific classification. Despite this, the consortia each take different stances when detailing the content, with PARCC taking a deficit approach by focusing on

students decreasing grammar and usage errors, and Smarter Balanced focusing their descriptors on improvements in grammar and mechanics.

Accessibility and Accommodations Guidelines

Smarter Balanced and PARCC have used similar procedures to develop accessibility and accommodations guidelines for English learners (EL) and for students with disabilities (SWD) taking their summative assessments. For example, both conducted reviews of available research and consulted with practitioners and national experts to assemble initial lists of acceptable practices. Furthermore, capitalizing on recent technological advances, both plan to embed EL and SWD supports and tools in their end-of-year tests; for example, read-aloud and glossary functions may be turned on for students who need such support (see Table 4 and Table 5).

Table 4

Accessibility Tools and Supports

Type	PARCC	Smarter Balanced
Universal accessibility tools (Embedded supports available to all students)	Audio amplification Blank paper (not embedded) Eliminate answer choices Flag items for review General administration directions read aloud and repeated as needed Highlight tool Magnification/enlargement device Noise buffers Note pad Pop-up glossary Redirect student to test (not embedded) Spell checker Writing tools	Answer eliminator Calculator (for certain items) Context-accurate and grade-appropriate glossaries Expandable passages Highlighting Marking for review Note-taking tool (ELA performance task) Pausing Spell checker Tab-enter navigation Variable font size Writing tools (i.e., bold, italics, underline; indent; cut, copy, paste; undo/redo)
Accessibility features (Embedded supports assigned to students as needed)	Answer masking Background/font color (color contrast) General administration directions/clarified Line reader tool Masking Text-to-speech (Mathematics)	Font background color alternatives Print on request (items, stimuli, passages)

Table 5

Accommodations for Students With Disabilities and/or English Learners

Type	PARCC	Smarter Balanced
Presentation and response	Assistive technology Braille edition/note-taker Closed-captioning of video Descriptive video Familiar test administrator Paper-pencil edition Scribing or speech-to-text (Mathematics; SWDs, ELs) Tactile graphics Video of a human interpreter for the deaf	Online refreshable braille Text-to-speech
Timing and scheduling	Extended time (SWDs, ELs) Frequent breaks (SWDs, ELs) Time of day	Extended time Frequent breaks
Setting	Adaptive or specialized furniture Separate or alternate location Small group Special lighting Specified area or preferential seating	Quiet environment
Special access	Calculation device English/native language word-to-word dictionary (ELs) Read aloud or text-to-speech (ELs) Test directions clarified in native language (ELs) Text-to-speech—read aloud (ELA) Speech-to-text, dictating/transcription (ELA) Video of a human interpreter (ELA) Word prediction (ELA performance task)	Online English glossary

However, the consortia diverge in their approaches to revising their guidelines. Smarter Balanced tested the validity of some of the supports (e.g., extended time, frequent breaks, and print on request) and tools (e.g., highlighting, answer eliminator, and calculator) during their spring 2013 pilot and in the practice tests that have been made publicly available. Once the results are analyzed, Smarter Balanced will finish development of their comprehensive framework, which will be further updated after field testing in 2014. In contrast, PARCC solicited two rounds of public comment to their draft manual during the winter and spring of 2013, approved the first edition of the manual in late June, and will have states and stakeholders further vet the manual during the summer. Once this vetting is complete, PARCC will pilot

supports and tools during item development research and field testing slated for late 2013 and early 2014.

Table 4 and Table 5 display current supports and tools that are planned by both consortia. While generally similar, given their different approaches to guideline development, it is not surprising that PARCC provides more specific guidelines than Smarter Balanced. This is primarily true for the accommodations guidelines. Table 4 provides recent guidelines concerning accessibility supports, while Table 5 provides guidelines for accommodations. Information for these tables was obtained from the first edition of the PARCC Accommodations Manual (2013), the Smarter Balanced Pilot Test Accessibility and Accommodations Guidelines (2013), and the Smarter Balanced Accessibility and Accommodations Policies and Guidelines: Chief Instructional Officers Update (Boshamer & Swaffield, 2013).

Technology Readiness

The Technology Readiness Tool (TRT) was launched during spring 2012 as a joint venture between Smarter Balanced and PARCC. Created by Pearson, the TRT includes an online survey where authorized users can input information about the current technology at their schools or districts (see https://www.techreadiness.net). Once analyzed, stakeholders from the school to state level can use the TRT disaggregation tools to create customized tabular or graphical gap-analysis reports indicating the current state of readiness for their consortium's technology-based summative assessment.

Currently the consortia plan to extract (download) and analyze available TRT survey data twice yearly through 2014. During its first year of availability, though, the consortia completed two additional extractions (September 2012 and February 2013) in order to better support states. Most of the results publicly available for the TRT are based on the baseline survey extraction. During this first round of data collection, 32 of 46 potential states had a school participate, and 19 states had over half of their schools participate. Because data are not available for most schools, and some of the available survey responses were incomplete, these data are difficult to interpret.

Results thus far indicate that the majority of schools:

- Use MS Windows, with most of the remaining using an Apple Mac or iOS operating system;
- Meet recommended specifications of at least 1GB of RAM; and
- Meet requirements concerning display size (≥ 9.5 inches) and resolution (≥ 1024 x 768 pixels).

Table 6

Technology Specifications for Computers

	PARCC		Smarter Balanced	
Type	Minimum	Recommended	Minimum	Recommended
Computer OS				
Windows	XP (Service Pack 3)	7 +	XP (Service Pack 3)	7 +
MAC OS	OS 10.5	OS 10.7 +	OS 10.4.4	OS 10.7 +
Linux	Ubunto 9-10, Fedora 6	Ubunto 11.10, Fedora 16 +	Ubunto 9-10, Fedora 6	Ubunto 11.10, Fedora 16
Chrome OS	OS 19	OS 19 +	OS 19 +	OS 19 +
Tablet OS				
iOS	iPad 2 with 6	iPad 2 + with 6	iPad 2 with 6	iPad 3 + with 6
Android	4.0	4.0 +	4.0 +	4.0 +
Windows	8	8 +	8 + (excluding RT)	8 + (excluding RT)
Memory				
RAM	512 MB	1 GB +	128 MB (Windows) 256 MB (MAC) 64 MB (Linux)	1 GB
Processor (MHz)	--	--	233 MHz (Windows) 300 MHz (MAC) 233 MHz (Linux)	1 GHz
Hard drive (free space)	--	--	52 MB (Windows) 200 MB (MAC) 64 MB (Linux)	80 GB drive or at least 1 GB free space
Connectivity	Wired or wireless networks	Wired or wireless networks	Wired or wireless networks	Wired or wireless networks
Screen				
Size	9.5 in +	9.5 in +	9.5 in or larger	9.5 in or larger
Resolution	1024 x 768 +	1024 x 768 +	1024 x 768 +	1024 x 768 +
Input devices	Keyboard and mouse/touchpad or touchscreen	Keyboard and mouse/touchpad or touchscreen	--	--
Audio	Head/earphones and microphones	Head/earphones and microphones	Head/earphones	Head/earphones

However, sizable proportions of the devices—approximately 15–30%—fail to meet some of the specified requirements (see Table 6), and these shortcomings tend to be more prevalent in PARCC than in Smarter Balanced states. Available data and the large quantity of missing data thus suggest that technology readiness will be a continuing challenge for both consortia. As

shown in Table 6, both have developed guidelines for purchasing new computers and tablets to support assessment implementation, and both have established minimum specifications, which they will support through the 2014–2015 school year in an effort to give schools more time to build their capacity.

In addition to the TRT, both consortia have developed tools to help support states' capacity to assess their technology needs. The Smarter Balanced web-based diagnostic tool detects the operating system, browser, browser security, and bandwidth for the computer being used (see http://www.smarterbalanced.org/smarter-balanced-assessments/technology/). Smarter Balanced also has an online calculator where stakeholders can calculate how many administration days the ELA and mathematics assessments will require based on the number of students, computers, hours per day, and Internet connection speed available (see http://www3.cde.ca.gov/sbactechcalc). PARCC provides states with a spreadsheet that schools or districts can use to assess their capacity. In this case, the tool helps administrators take into account the number of students at different grade levels and the preferred number of administration days at the school while trying to determine their future technology needs (see http://parcconline.org/assessment-administration-guidance). Smarter Balanced, and most likely PARCC, will also be offering paper-and-pencil versions of the tests to states during the 2014–2015 school year to address gaps in technology capacity.

Comparing the Smarter Balanced and PARCC Content Specifications

Expectations for the comparability of Smarter Balanced and PARCC summative assessments vary. However, to achieve this, the two systems must be built on comparable content specifications, which map the claims and targets each assessment will address. Smarter Balanced published their content specifications for ELA and mathematics in February 2012, and PARCC published drafts of their content frameworks in both subjects in November 2011 with revisions made available in November 2012. We used these documents to conduct a general comparison of the proposed content in both consortia's assessments. Our analysis addresses content targets only, as PARCC has yet to specify the DOK or cognitive complexity at which they expect each target to be addressed.

As noted in our previous report (Herman & Linn, 2013), both consortia organize their specifications into the major claims about student performance that their summative assessments are intended to substantiate, and these are generally similar across the two. For example, in their ELA summative assessments both will focus on claims dealing with reading, writing, and research, with Smarter Balanced also focusing on listening from their speaking/listening claim.

In contrast, PARCC plans to develop optional local assessments focused on their fourth claim of speaking and listening.

Smarter Balanced defines relevant content for each claim in terms of assessment targets, which typically combine multiple standards. In contrast, PARCC defines their claims in terms of evidence statements, which correspond to individual standards in ELA, but in mathematics also include integrative evidence statements that combine many standards (e.g., all of the standards related to functions). We thus deconstructed the Smarter Balanced assessment targets into their constituent standards and used the PARCC evidence statements comprising individual standards to conduct a standards-to-standards comparison.

Several caveats are important in considering the comparisons made: First, we excluded PARCC's integrative evidence statements, because it was not clear how so many individual math standards—eight to 10 in a number of cases—could be incorporated into a single item or task. Second, PARCC has not yet articulated evidence statements for their ELA research claim, and Smarter Balanced has not yet specified content targets by grade for either their communication/reasoning or modeling claims in mathematics. As a result, our analyses likely underestimate both consortia's plans for specific standards coverage. Furthermore, since we deconstructed the Smarter Balanced targets, our analyses no longer represent their specific plans.

English language arts. Table 7 summarizes the consortia's plans for ELA standards coverage by grade, excluding those for speaking and listening since assessment plans for this claim appear in flux for both. In general, the results indicate that both plan to address the vast majority of the standards, that Smarter Balanced coverage is slightly more expansive than that of PARCC, and that almost all of both consortia's targets are in common.

Looking more closely at the specific Common Core State Standards (CCSS) for ELA that each consortium plans to address by grade, one can see that standards that are not included in the specifications likely overlap with other ELA standards that are planned for assessment. For example, excluding the standards for speaking and listening, there are 36 ELA standards for Grade 3 and 37 standards for Grades 4 and 5, respectively. As displayed in Appendix Table A1, the Smarter Balanced content specifications address all but five standards for each of these grade levels, and the PARCC specifications—not including the research claim—include all but eight standards for Grade 3 and all but seven standards for Grades 4 and 5, respectively. The table also shows that both consortia are consistent in some of the standards they exclude, for example, foundational skills in reading (ELA Standards 3.3 and 3.4) and an overall standard on being able to read and comprehend text at the high end of grade-level proficiency (Standard 3.10), which could be seen as overlapping with the prior reading standards. Each consortium also appears to

have made some unique decisions about which standards might be overlapping with others, such as PARCC's exclusion of conventions of language (Standards 3.1–3.3).

Table 7

ELA/Literacy CCSS Content Specifications Comparison

		# standards covered		
Grade	# Common Core State Standard	Smarter Balanced	PARCC	Both consortia
3	36	31	28	27
4	37	32	30	29
5	37	32	30	29
6	35	32	30	29
7	35	31	30	29
8	35	30	30	29
9	35	29	30	26
10	35	30	30	27
11	35	30	29	26

Table 8

Mathematics CCSS Content Specifications Comparison

		# standards covered		
Grade	# Common Core State Standard	Smarter Balanced	PARCC	Both consortia
3	25	25	22	22
4	28	28	28	28
5	26	26	26	26
6	29	29	28	28
7	24	24	24	24
8	28	28	26	26

Mathematics. Table 8 summarizes the consortia's plans for mathematics standards coverage, excluding high school.[1] The Smarter Balanced specifications show attention to all

[1] Smarter Balanced specified assessment targets linked to standards only at the high school level. The specifications for Claims 3 and 4 address a large range of content goals but do not directly reference individual standards. Further, at the high school level, PARCC provided separate specifications for integrated math and for the standard Algebra 1, Geometry, and Algebra 2 sequence. Direct comparisons between the two consortia were therefore not possible.

Common Core standards across Grades 3 through 8, as do PARCC's, save for a few scattered omissions in Grades 3, 6, and 8. Thus, the majority of their targets are in common. The specific standards addressed for Grades 3 through 5 in mathematics can be found in Appendix Table A2.

Expectations for Deeper Learning Representation

CRESST is planning to convene an expert panel to establish standards for intellectual rigor for state assessments of the Common Core State Standards. The standards are intended to provide a key indicator in state decisions about test adoption. In both ELA and mathematics, the panels will bring together experts in the CCSS, expert teachers, learning and assessment scholars, and specialists in diverse learners at the elementary, middle, and high school levels. The panels will consider data on the cognitive complexity and DOK expectations of the Common Core tests and that of respected international and U.S. tests of student proficiency and college readiness. Taking into account these data, the panel will reach consensus on the extent to which deeper learning, as defined by higher levels of cognitive complexity and DOK, ought to be represented on state tests. While the official standard setting exercise awaits final analysis and reporting of RAND's analysis of current, highly regarded tests, we offer preliminary thoughts on appropriate standards based on existing data on DOK expectations implicit in the CCSS. Our data come from expert consensus panels established by states and nationally prominent researchers to identify the DOK expectations for the CCSS in ELA and mathematics. These sources include:

- The Florida State Department of Education;
- The Iowa State Department of Education;
- Norman Webb's study of the alignment between the CCSS and six nationally available high school tests (Webb, 2012a, 2012b); and,
- David Conley and colleagues' study of the alignment between the CCSS and prior state standards at the high school level (Conley et al., 2011).

In addition, we use the Smarter Balanced content specifications, which designate DOK goals for the consortium's assessment targets, which themselves are a reorganization of the CCSS for test development, interpretation, and reporting purposes, and have been approved by all participating states. We also include preliminary data from RAND's analysis of respective national and international tests (Yuan & Le, 2012).

In each case, Norman Webb's framework is used to characterize DOK expectations, where

- DOK1 is characterized by recall of a fact, term, concept, or procedure; basic comprehension.

- DOK2 entails application of concepts and/or procedures involving some mental processing.
- DOK3 requires abstract thinking, reasoning, and/or inferences that are more complex.
- DOK4 involves extended analysis or investigation that requires synthesis and analysis across multiple contexts and nonroutine applications.

CCSS Depth of Knowledge Designations: English Language Arts

Table 9 displays how several states and national studies characterize DOK expectations of the CCSS for ELA. The table shows, by source, the mean percentage of standards and/or assessment targets designated at each DOK level.

Table 9

DOK Levels Reflected in ELA CCSS: Percentage Represented by Source, Subject Area, and School Level

ELA DOK level	Florida	Iowa	Smarter Balanced	Webb (2012a)	Conley et al. (2011)
Elementary school					
1	10.3	25.0	24.5	--	--
2	43.9	30.7	37.7	--	--
3	44.9	29.7	26.4	--	--
4	0.9	11.5	11.3	--	--
Middle school					
1	1.1	20.2	21.2	--	--
2	37.4	25.9	26.9	--	--
3	52.6	33.4	32.7	--	--
4	8.9	19.2	19.2	--	--
High school					
1	0.0	18.8	22.4	1.7	4.5
2	30.7	30.5	28.6	10.9	9.0
3	59.1	33.0	28.6	57.1	52.2
4	10.2	17.8	20.4	30.3	34.3

Note. Iowa and Smarter Balanced designated multiple DOK ratings to many individual standards. For purposes of this analysis, each unique combination of standard and DOK level was treated as a separate target (e.g., if standard one is specified at DOK2 and DOK3, this constitutes two separate targets). Column percentages reflect the percentage of total targets at each level.

The data in Table 9 show a range of DOK expectations for ELA, depending on the source and, to a lesser extent, school level. At the elementary and middle school levels, the chief contrasts are between Florida's expectations and those of Iowa and Smarter Balanced—DOK expectations are generally very similar across the latter two, which is to be expected given that

Iowa is part of the Smarter Balanced Assessment Consortium. Florida is less likely to designate standards at either the lowest or the highest levels of DOK, providing higher concentrations at DOK2 and DOK3. Across all three sources, there is a general increase in expected rigor from elementary to middle school, which is maintained at the high school level. While at the elementary school level, standards appear to be distributed across DOK levels, at the middle and high school levels DOK3 and DOK4 are designated for roughly half the standards. It should be noted that the two high school-only, more research-oriented studies (Conley et al., 2011; Webb, 2012a) designate more than 80% of the standards at DOK3 or DOK4 and seem to be outliers relative to state-designated expectations. Across all five sources, the percentage of standards at DOK3 ranges from 29% to 59% and at DOK4 from 10% to 34%.

Across all grade levels, there is consensus that the majority of standards require thinking at or above DOK2, and that the majority of standards at the middle and high school levels call on DOK3 or DOK4. Mean and median values across the three states provide similar values: The percentage of standards characterized as DOK3 or DOK4 are 42%, 55%, and 56% for elementary, middle, and high school ELA standards respectively.

However, given the variation across the small number of other sources, and because the Smarter Balanced specifications have been vetted across their member states, we recommend relying on their specifications as representing the most widespread consensus on DOK expectations. Norman Webb's framework suggests that consistency between standards and a test is established if at least 50% of the items addressing a standard meet the designated DOK level. Based on this rather conservative criterion, and using the Smarter Balanced specifications, at least approximately 20% of ELA items on a given elementary test should address DOK3 or DOK4, and at least 25% of the items on middle or high school tests should incorporate these levels. Bringing the Webb (2012a) and Conley et al. (2011) studies into the equation would suggest that approximately 35% of high school ELA tests should address higher DOK levels.

CCSS Depth of Knowledge Designations: Mathematics

Table 10 shows the distribution of DOK across the mathematics CCSS. Here, although we find variation across sources and school levels, one trend is very clear: DOK expectations for mathematics appear relatively less rigorous than do those in ELA. The mean across all the sources indicates that, for all grade levels, approximately 20% of the standards call on these higher levels (DOK3 or DOK4). However, rather than lower intellectual demand, the DOK distributions may indicate a mismatch between the nature of the mathematics demands and the DOK framework. That is, the DOK framework does not take into account the complexity of the mathematics that students need to process, but only the extent of processing.

Table 10

DOK Levels Reflected in Math CCSS: Percentage Represented by Source, Subject Area, and School Level

Math DOK level	Florida	Iowa	Smarter Balanced	Webb (2012b)	Conley et al. (2011)
Elementary school					
1	30.4	42.5	24.3	--	--
2	60.2	46.8	39.6	--	--
3	9.4	10.8	25.4	--	--
4	0.0	0.0	10.7	--	--
Middle school					
1	11.8	44.3	21.0	--	--
2	75.6	42.7	41.4	--	--
3	12.6	12.9	26.5	--	--
4	0.0	0.0	11.1	--	--
High school					
1	17.1	44.1	27.3	17.4	21.2
2	67.9	43.8	40.9	61.0	53.9
3	14.5	11.6	22.7	20.7	19.7
4	0.0	0.3	9.1	0.5	4.7

Note. Iowa and Smarter Balanced designated multiple DOK ratings to many individual standards. For purposes of this analysis, each unique combination of standard and DOK level was treated as a separate target (e.g., if standard one is specified at DOK2 and DOK3, this constitutes two separate targets). Column percentages reflect the percentage of total targets at each level.

With that caveat in mind, the majority of mathematics standards across school levels are judged at DOK1 or DOK2, with the latter being the relatively most frequent expectation. DOK4 is relatively absent, except in the Smarter Balanced specifications where 9% of the targets are so designated. This relative absence may be explained by at least two factors: First, DOK4 requires a problem involving complex thinking and reasoning that extends over a significant period of time, which by definition is not possible through typical end-of-year assessment items. Secondly, only the Smarter Balanced specifications specifically incorporate the mathematical practice standards into their assessment targets, which call on extended reasoning, modeling, and communication; in contrast, the other groups rated only the mathematics content standards. As a result, the Smarter Balanced DOK designations appear more rigorous, with roughly a third of the targets identified at DOK3 or DOK4.

As with ELA, the Smarter Balanced mathematics specifications are the most widely vetted of those available and thus seem the most credible. They also are the most intellectually rigorous of the set. Here, expectations for higher levels of DOK are relatively consistent from elementary

through middle and high school, averaging approximately 35% at DOK3 and DOK4. Using Norman Webb's conservative criterion, a reasonable target for higher levels of DOK would be only half that.

Depth of Knowledge Representation in Respected National and International Assessments

RAND's preliminary analysis of the DOK levels evident in respected national and international tests (Yuan & Le, 2012)—including the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Study (PIRLS), Advanced Placement (AP), and International Baccalaureate (IB) programs—shows similar trends to those in the CCSS DOK specifications. That is, the ELA DOK expectations are higher than are those for mathematics, and there is considerable variation in expectations depending on source, in this case the various tests. For example, RAND's preliminary analysis of the ELA and mathematics tests shows a range of 25% to 100% of open-ended items at DOK3 and DOK4 and for multiple choice items, from 7% to 56% at those levels. In mathematics, 4% to 46% of items were identified at DOK3 and DOK4, and for multiple choice items, 0% to 15% at those levels. In general, RAND's analysis of NAEP, TIMMS, and PIRLS showed relatively little representation of the two highest levels of DOK. Even PISA results showed a preponderance of items at the two lowest levels of DOK, with the highest relative frequency at DOK2. In general, the Smarter Balanced DOK specifications appear to surpass those for all tests analyzed except for the AP and IB tests.

In the following section, we concentrate on the analysis of AP and IB tests, which involved full test forms rather than sample items, so that we could calculate how well each DOK level was represented in student scores. We felt that relative score value was a better metric for evaluating the emphasis given to deeper levels of learning than was the proportion of items. For example, if an essay response at DOK4 is worth a possible ten points and a multiple choice item at DOK1 or DOK2 is worth one point, then an assessment with one essay and ten multiple choice items gives equal attention to DOK4 as to DOK1 and DOK2, even though 91% of the total number of items are at the two lowest levels. We thus converted RAND's preliminary findings on the percentage of open-ended and multiple choice items at each level into a composite percentage of total score points at each level. This was a relatively easy conversion for the IB tests, since they were composed solely of open-ended items, and there was no need to aggregate RAND's findings across the two item types. In the absence of available information, however, our analysis assumed that each item was given an equal score weight, which likely underestimates the weight given to higher levels, which are likely to involve more score points. Further, although upper division IB tests reflected somewhat higher levels of DOK than the lower division one, our results below combine the two.

Our AP analysis combines specific results by course into aggregates for ELA and mathematics. We started with RAND's findings for open-ended and multiple choice items and then used what we knew about the relative score values for these two types of items to compute a weighted average. In ELA, approximately half of a student's total AP score (55%) is based on their responses to the open-ended items, while in mathematics, an average open-ended item seems to be worth about two and a half times the score value of a multiple choice item.

Depth of knowledge levels in English language arts tests. The ELA results for IB and AP are displayed in Table 11. Based on the RAND analysis, all of the IB items address DOK4 and all are essay items (Yuan & Le, 2012).

Table 11

Depth of Knowledge in IB and AP: ELA (RAND Study Data: Embargoed)

		DOK levels				
		1	2	3	4	Total
IB: Writing, lower and upper division						
Open-ended items	*n*	0	0	0	6	6
	%	0	0	0	100	100
Composite total test score	%	0	0	0	100	100
AP: All ELA subjects						
Open-ended items	*n*	0	0	0	6	6
	%	0	0	0	100	100
Multiple choice items	*n*	12	35	60	0	107
	%	11	33	56	0	100
Composite total test score (estimated)[a]	%	5	15	25	55	100

[a]Weighted based on estimated relative score values of open-ended and multiple choice items. Total score estimated as composed of 55% open-ended items and 45% multiple choice items.

We see this same trend in AP results with all of the essay items addressing DOK4. However, the AP exams also include multiple choice items addressing reading, which range from DOK1 to DOK3. In addition, DOK3 comprises the clear majority of multiple choice items (60%). As previously noted, available information suggests that 55% of the total score is attributable to the open-ended, essay items, and the remaining 45% is attributable to multiple choice items. Based on these allocations we find that approximately 55% of the total score possible is at DOK4 and 25% at DOK3, for a total of 80% at these levels. Given the populations for which these assessments are intended, it is not surprising that their DOK expectations seem to exceed, by far, those for the other tests examined.

Depth of knowledge levels in mathematics tests. Table 12, which shows RAND's mathematics results for IB and AP tests, again shows the correspondence between item type and DOK level (Yuan & Le, 2012). Because DOK4 requires extended analysis, this level cannot be addressed through a multiple choice item. While it is not the case, as seen in this table, that all open-ended items draw on deeper levels of knowledge, it is the case that only open-ended items can elicit DOK4.

Table 12

Depth of Knowledge in IB and AP: Mathematics (RAND Study Data: Embargoed)

		DOK levels				
		1	2	3	4	Total
IB: Math, lower and upper division						
Open-ended items	*n*	33	79	43	2	157
	%	21	50	27	1	100
Composite total test score	%	21	50	27	1	100
AP: All math subjects						
Open-ended items	*n*	15	19	27	2	63
	%	24	30	43	3	100
Multiple choice items	*n*	44	86	0	0	130
	%	34	66	0	0	100
Composite total test score (estimated)[a]	%	17	37	42	4	100

[a]The total score composite assumes that each open-ended item score value is 2.5 times that of a multiple choice item.

As with ELA, in the absence of available information, the IB mathematics composite score results assume that score values are the same for all items. For AP, the calculation required making assumptions that are debatable, based on the varying number of score points associated with different open-ended items, which ranged principally from one to four. Absent complete information on the score points associated with each item reviewed by RAND, we assumed that score points varied with DOK levels, and that the higher the DOK level, the more score points associated with the item. Open-ended items at DOK1 were assumed to be worth one point, those at DOK2 were assumed to be two-point items, and so on for DOK3 and DOK4. The composite total test score data in Table 12 reveal that the percentage of items at DOK3 and DOK4 are estimated at 28% and 33% for IB and AP respectively.

As with the analysis of DOK expectations for the CCSS, these results show a significant reduction in DOK demand in mathematics relative to ELA. The distribution of DOK levels here

seems particularly surprising given that these tests are intended for the relatively highest achieving, college bound students and, for AP, are intended to represent college-level work—and again, may indicate the limitations of the DOK metric for capturing intellectual demand in mathematics. Even so, Smarter Balanced expectations for the Common Core seem to surpass those evident in both the IB exam and that of the AP exams.

Summary

In summary, we see a range of depth of knowledge expectations within and across those for the Common Core State Standards and those evident in respected national and international tests. Given this variation, it is difficult to establish a firm standard for DOK representation.

However, because the specifications developed by the Smarter Balanced consortium have been the most widely vetted and the other sources involve only a single state or a single study, we find the Smarter Balanced formulation currently the most credible. Not only were the specifications established by an expert panel, similar to those assembled by other sources, but they were also reviewed and approved by all member states. The Smarter Balanced specifications further distinguished themselves by being generally higher in DOK expectations than the vast majority of tests reviewed by RAND. The only exceptions were the AP and IB exams for English language arts, which are generally intended for academically elite students. Even for these latter tests, Smarter Balanced expectations for higher levels of thinking were 50% to 75% of those for IB and AP respectively. Given these general patterns, we believe that Smarter Balanced specifications provide the best source of guidance currently for DOK specifications. Moreover, if it is important to include DOK4, and we believe it is, then RAND's test results make clear that assessments must include extended open-ended tasks. The more extended open-ended items, the greater the likelihood that DOK4 will be represented.

What would this mean? Smarter Balanced's DOK expectations in English language arts start at about 38% at DOK3 and DOK4 at the elementary school level and move to roughly 50% at DOK3 and DOK4 at the middle and high school levels. In mathematics, proportions at these two highest levels range from 32% to 38%, with no clear pattern by school level. Further, across both grades, the clear majority of targets (more than 75%) are characterized as DOK2 or above. This alone is a substantial change from most current state assessments.

Conclusion and Recommendations

Both PARCC and Smarter Balanced have made substantial progress over the course of the year and plans—based on content specifications, blueprints and released sample items—bode well for the representation of deeper learning in both consortia's assessments. Both have increased oversight on item and task development in an effort to assure that items and tasks will

be well crafted and rigorous. Largely because of these concerns, PARCC has delayed their pilot testing plans. In contrast, Smarter Balanced held to their schedule, completing pilot testing plans and developing and sharing practice tests for use with all of their states.

The analysis of Smarter Balanced and PARCC content specifications reveals close similarity in plans for content coverage, as might be expected since both are supposed to be aligned with the CCSS. However, plans for relative rigor in the two tests are not yet known since PARCC has yet to specify publicly the DOK or cognitive complexity for their individual evidence statements. Regardless of how similar the rigor of the items and tasks, however, the consortia achievement or performance level descriptors will limit the comparability of scores from one consortium's tests to the others. For example, PARCC is planning to classify student performance relative to five proficiency levels, while Smarter Balanced is planning to use four levels.

Finally, our preliminary analysis of DOK requirements in the Common Core State Standards and RAND's analysis of the DOK expectations evidenced in respected national tests lead us to conclude that Smarter Balanced specifications represent a reasonable and rigorous target for deeper learning.

Based on our analyses, we recommend the following:

1. To provide some flexibility, particularly in the transition from less to far more rigorous tests, we suggest the following interim benchmarks for DOK representation in annual, state accountability tests:
    - ELA elementary: 33% of total possible score points at DOK3 and DOK4, and at least 75% at or above DOK2, with at least one extended performance task.
    - ELA middle and high school: 50% of total possible score points at DOK3 and DOK4, and at least 75% at or above DOK2, with at least one extended performance task.
    - Mathematics, all school levels: 33% of total possible score points at DOK3 and DOK4, and at least 75% at or above DOK2, with at least one extended performance task.
2. We suggest further consideration of the sensitivity of the DOK metrics for adequately capturing deeper learning in mathematics. We look forward to RAND's findings in this area.
3. While both consortia indicate that they considered complexity when developing their ALDs or PLDs, neither was explicit in how these were being applied to the descriptors showing expected learning progressions for specific content areas and/or grade levels. As such, we recommend the consortia make this information more explicit to aid item writing as well as interpretation by stakeholders.

4. Preliminary data reveal that 15% to 30% of technology (i.e., computers and tablets) in the states may fail to meet some of the specified requirements for the summative assessments. This is a concern particularly for the implementation of higher complexity items, which may require enhanced technology. Therefore, we recommend that both consortia increase their efforts to make district and school administrators aware of the technological capacity required for the implementation of the tests in 2014–2015 and beyond.

References

Boshamer, C., & Swaffield, S. (2013, March 12). *Smarter Balanced accessibility and accommodations policies and guidelines: Chief instructional officers update* [PowerPoint slides]. Columbia, SC: South Carolina Department of Education.

Conley, D. T., Drummond, K. V., de Gonzalez, A., Seburn, M., Stout, O., & Rooseboom, J. (2011). *Lining up: The relationship between the Common Core State Standards and five sets of comparison standards.* Eugene, OR: Educational Policy Improvement Center.

Herman, J., & Linn, R. (2013). *On the road to assessing deeper learning: The status of Smarter Balanced and PARCC assessment consortia* (CRESST Report 823). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, & Student Testing.

Partnership for Assessment of Readiness for College and Careers. (2013). *PARCC accommodations manual: First edition*. Washington, DC: PARCC Assessment Consortia.

Smarter Balanced Assessment Consortium. (2013). *Smarter Balanced pilot test accessibility and accommodations guidelines*. Retrieved from http://sbac.portal.airast.org/Pilot_Test/resources/PilotTestAAGuidelines_022613.pdf

Webb, N. L. (2012a). *Alignment analysis of the English language arts Common Core State Standards for Grades 9-12 and six assessments—ACT, SAT, PSAT, PLAN, EXPLORE, and ReadiStep.* Indianapolis, IN: Indiana Department of Education.

Webb, N. L. (2012b). *Alignment analysis of the mathematics Common Core State Standards for Grades 9-12 and six assessments—ACT, SAT, PSAT, PLAN, EXPLORE, and ReadiStep.* Indianapolis, IN: Indiana Department of Education.

Yuan, K., & Le, V. (2012). *Estimating the percentage of students who were tested on cognitively demanding items through the state achievement tests* (WR-967-WFHF). Santa Monica, CA: RAND.

Appendix:
Smarter Balanced and PARCC Content Specifications

Table A1

CCSS Content Coverage for Grades 3–5 ELA/Literacy for Smarter Balanced and PARCC

Grade 3 ELA			Grade 4 ELA			Grade 5 ELA		
Standards	SBAC	PARCC	Standards	SBAC	PARCC	Standards	SBAC	PARCC
Reading Literature			**Reading Literature**			**Reading Literature**		
ELA-Literacy.RL.3.1	✓	✓	ELA-Literacy.RL.4.1	✓	✓	ELA-Literacy.RL.5.1	✓	✓
ELA-Literacy.RL.3.2	✓	✓	ELA-Literacy.RL.4.2	✓	✓	ELA-Literacy.RL.5.2	✓	✓
ELA-Literacy.RL.3.3	✓	✓	ELA-Literacy.RL.4.3	✓	✓	ELA-Literacy.RL.5.3	✓	✓
ELA-Literacy.RL.3.4	✓	✓	ELA-Literacy.RL.4.4	✓	✓	ELA-Literacy.RL.5.4	✓	✓
ELA-Literacy.RL.3.5	✓	✓	ELA-Literacy.RL.4.5	✓	✓	ELA-Literacy.RL.5.5	✓	✓
ELA-Literacy.RL.3.6	✓	✓	ELA-Literacy.RL.4.6	✓	✓	ELA-Literacy.RL.5.6	✓	✓
ELA-Literacy.RL.3.7	✓	✓	ELA-Literacy.RL.4.7	✓	✓	ELA-Literacy.RL.5.7	✓	✓
ELA-Literacy.RL.3.9	✓	✓	ELA-Literacy.RL.4.9	✓	✓	ELA-Literacy.RL.5.9	✓	✓
ELA-Literacy.RL.3.10	✗	✗	ELA-Literacy.RL.4.10	✗	✗	ELA-Literacy.RL.5.10	✗	✗
Reading: Informational Text			**Reading: Informational Text**			**Reading: Informational Text**		
ELA-Literacy.RI.3.1	✓	✓	ELA-Literacy.RI.4.1	✓	✓	ELA-Literacy.RI.5.1	✓	✓
ELA-Literacy.RI.3.2	✓	✓	ELA-Literacy.RI.4.2	✓	✓	ELA-Literacy.RI.5.2	✓	✓
ELA-Literacy.RI.3.3	✓	✓	ELA-Literacy.RI.4.3	✓	✓	ELA-Literacy.RI.5.3	✓	✓
ELA-Literacy.RI.3.4	✓	✓	ELA-Literacy.RI.4.4	✓	✓	ELA-Literacy.RI.5.4	✓	✓
ELA-Literacy.RI.3.5	✓	✓	ELA-Literacy.RI.4.5	✓	✓	ELA-Literacy.RI.5.5	✓	✓

Grade 3 ELA			Grade 4 ELA			Grade 5 ELA		
Standards	SBAC	PARCC	Standards	SBAC	PARCC	Standards	SBAC	PARCC
ELA-Literacy.RI.3.6	✓	✗	ELA-Literacy.RI.4.6	✓	✓	ELA-Literacy.RI.5.6	✓	✓
ELA-Literacy.RI.3.7	✓	✓	ELA-Literacy.RI.4.7	✓	✓	ELA-Literacy.RI.5.7	✓	✓
ELA-Literacy.RI.3.8	✓	✓	ELA-Literacy.RI.4.8	✓	✓	ELA-Literacy.RI.5.8	✓	✓
ELA-Literacy.RI.3.9	✓	✓	ELA-Literacy.RI.4.9	✓	✓	ELA-Literacy.RI.5.9	✓	✓
ELA-Literacy.RI.3.10	✗	✗	ELA-Literacy.RI.4.10	✗	✗	ELA-Literacy.RI.5.10	✗	✗
Reading: Foundational Skills			**Reading: Foundational Skills**			**Reading: Foundational Skills**		
ELA-Literacy.RF.3.3	✗	✗	ELA-Literacy.RF.4.3	✗	✗	ELA-Literacy.RF.5.3	✗	✗
ELA-Literacy.RF.3.4	✗	✗	ELA-Literacy.RF.4.4	✗	✗	ELA-Literacy.RF.5.4	✗	✗
Writing			**Writing**			**Writing**		
ELA-Literacy.W.3.1	✓	✓	ELA-Literacy.W.4.1	✓	✓	ELA-Literacy.W.5.1	✓	✓
ELA-Literacy.W.3.2	✓	✓	ELA-Literacy.W.4.2	✓	✓	ELA-Literacy.W.5.2	✓	✓
ELA-Literacy.W.3.3	✓	✓	ELA-Literacy.W.4.3	✓	✓	ELA-Literacy.W.5.3	✓	✓
ELA-Literacy.W.3.4	✓	✓	ELA-Literacy.W.4.4	✓	✓	ELA-Literacy.W.5.4	✓	✓
ELA-Literacy.W.3.5	✓	✓	ELA-Literacy.W.4.5	✓	✓	ELA-Literacy.W.5.5	✓	✓
ELA-Literacy.W.3.6	✓	✓	ELA-Literacy.W.4.6	✓	✓	ELA-Literacy.W.5.6	✓	✓
ELA-Literacy.W.3.7	✓	✓	ELA-Literacy.W.4.7	✓	✓	ELA-Literacy.W.5.7	✓	✓
ELA-Literacy.W.3.8	✓	✓	ELA-Literacy.W.4.8	✓	✓	ELA-Literacy.W.5.8	✓	✓
ELA-Literacy.W.3.10	✗	✓	ELA-Literacy.W.4.9	✓	✓	ELA-Literacy.W.5.9	✓	✓
			ELA-Literacy.W.4.10	✗	✓	ELA-Literacy.W.5.10	✗	✓

Grade 3 ELA			Grade 4 ELA			Grade 5 ELA		
Standards	SBAC	PARCC	Standards	SBAC	PARCC	Standards	SBAC	PARCC
Speaking & Listening			**Speaking & Listening**			**Speaking & Listening**		
ELA-Literacy.SL.3.1	✓	✕	ELA-Literacy.SL.4.1	✓	✕	ELA-Literacy.SL.5.1	✓	✕
ELA-Literacy.SL.3.2	✓	✕	ELA-Literacy.SL.4.2	✓	✕	ELA-Literacy.SL.5.2	✓	✕
ELA-Literacy.SL.3.3	✓	✕	ELA-Literacy.SL.4.3	✓	✕	ELA-Literacy.SL.5.3	✓	✕
ELA-Literacy.SL.3.4	✓	✕	ELA-Literacy.SL.4.4	✓	✕	ELA-Literacy.SL.5.4	✓	✕
ELA-Literacy.SL.3.5	✓	✕	ELA-Literacy.SL.4.5	✓	✕	ELA-Literacy.SL.5.5	✓	✕
ELA-Literacy.SL.3.6	✓	✕	ELA-Literacy.SL.4.6	✓	✕	ELA-Literacy.SL.5.6	✓	✕
Language			**Language**			**Language**		
ELA-Literacy.L.3.1	✓	✕	ELA-Literacy.L.4.1	✓	✕	ELA-Literacy.L.5.1	✓	✕
ELA-Literacy.L.3.2	✓	✕	ELA-Literacy.L.4.2	✓	✕	ELA-Literacy.L.5.2	✓	✕
ELA-Literacy.L.3.3	✓	✕	ELA-Literacy.L.4.3	✓	✕	ELA-Literacy.L.5.3	✓	✕
ELA-Literacy.L.3.4	✓	✓	ELA-Literacy.L.4.4	✓	✓	ELA-Literacy.L.5.4	✓	✓
ELA-Literacy.L.3.5	✓	✓	ELA-Literacy.L.4.5	✓	✓	ELA-Literacy.L.5.5	✓	✓
ELA-Literacy.L.3.6	✓	✓	ELA-Literacy.L.4.6	✓	✓	ELA-Literacy.L.5.6	✓	✓

Table A2

CCSS Content Coverage for Grades 3–5 Mathematics for Smarter Balanced and PARCC

Grade 3 Mathematics			Grade 4 Mathematics			Grade 5 Mathematics		
Standards	SBAC	PARCC	Standards	SBAC	PARCC	Standards	SBAC	PARCC
Operations and Algebraic Thinking			**Operations and Algebraic Thinking**			**Operations and Algebraic Thinking**		
Math.Content.3.OA.A.1	✓	✓	Math.Content.4.OA.A.1	✓	✓	Math.Content.5.OA.A.1	✓	✓
Math.Content.3.OA.A.2	✓	✓	Math.Content.4.OA.A.2	✓	✓	Math.Content.5.OA.A.2	✓	✓
Math.Content.3.OA.A.3	✓	✓	Math.Content.4.OA.A.3	✓	✓	Math.Content.5.OA.B.3	✓	✓
Math.Content.3.OA.A.4	✓	✓	Math.Content.4.OA.B.4	✓	✓			
Math.Content.3.OA.B.5	✓	✗	Math.Content.4.OA.C.5	✓	✓			
Math.Content.3.OA.B.6	✓	✗						
Math.Content.3.OA.C.7	✓	✓						
Math.Content.3.OA.D.8	✓	✓						
Math.Content.3.OA.D.9	✓	✗						
Number and Operations in Base Ten			**Number and Operations in Base Ten**			**Number and Operations in Base Ten**		
Math.Content.3.NBT.A.1	✓	✓	Math.Content.4.NBT.A.1	✓	✓	Math.Content.5.NBT.A.1	✓	✓
Math.Content.3.NBT.A.2	✓	✓	Math.Content.4.NBT.A.2	✓	✓	Math.Content.5.NBT.A.2	✓	✓
Math.Content.3.NBT.A.3	✓	✓	Math.Content.4.NBT.A.3	✓	✓	Math.Content.5.NBT.A.3	✓	✓
			Math.Content.4.NBT.B.4	✓	✓	Math.Content.5.NBT.A.4	✓	✓
			Math.Content.4.NBT.B.5	✓	✓	Math.Content.5.NBT.B.5	✓	✓
			Math.Content.4.NBT.B.6	✓	✓	Math.Content.5.NBT.B.6	✓	✓
						Math.Content.5.NBT.B.7	✓	✓

Grade 3 Mathematics			Grade 4 Mathematics			Grade 5 Mathematics		
Standards	SBAC	PARCC	Standards	SBAC	PARCC	Standards	SBAC	PARCC
Numbers and Operations - Fractions			**Numbers and Operations - Fractions**			**Numbers and Operations - Fractions**		
Math.Content.3.NF.A.1	✓	✓	Math.Content.4.NF.A.1	✓	✓	Math.Content.5.NF.A.1	✓	✓
Math.Content.3.NF.A.2	✓	✓	Math.Content.4.NF.A.2	✓	✓	Math.Content.5.NF.A.2	✓	✓
Math.Content.3.NF.A.3	✓	✓	Math.Content.4.NF.B.3	✓	✓	Math.Content.5.NF.B.3	✓	✓
			Math.Content.4.NF.B.4	✓	✓	Math.Content.5.NF.B.4	✓	✓
			Math.Content.4.NF.C.5	✓	✓	Math.Content.5.NF.B.5	✓	✓
			Math.Content.4.NF.C.6	✓	✓	Math.Content.5.NF.B.6	✓	✓
			Math.Content.4.NF.C.7	✓	✓	Math.Content.5.NF.B.7	✓	✓
Measurement And Data			**Measurement And Data**			**Measurement And Data**		
Math.Content.3.MD.A.1	✓	✓	Math.Content.4.MD.A.1	✓	✓	Math.Content.5.MD.A.1	✓	✓
Math.Content.3.MD.A.2	✓	✓	Math.Content.4.MD.A.2	✓	✓	Math.Content.5.MD.B.2	✓	✓
Math.Content.3.MD.B.3	✓	✓	Math.Content.4.MD.A.3	✓	✓	Math.Content.5.MD.C.3	✓	✓
Math.Content.3.MD.B.4	✓	✓	Math.Content.4.MD.B.4	✓	✓	Math.Content.5.MD.C.4	✓	✓
Math.Content.3.MD.C.5	✓	✓	Math.Content.4.MD.C.5	✓	✓	Math.Content.5.MD.C.5	✓	✓
Math.Content.3.MD.C.6	✓	✓	Math.Content.4.MD.C.6	✓	✓	Math.Content.5.MD.A.1	✓	✓
Math.Content.3.MD.C.7	✓	✓	Math.Content.4.MD.C.7	✓	✓			
Math.Content.3.MD.D.8	✓	✓						

Grade 3 Mathematics			Grade 4 Mathematics			Grade 5 Mathematics		
Standards	SBAC	PARCC	Standards	SBAC	PARCC	Standards	SBAC	PARCC
Geometry			**Geometry**			**Geometry**		
CCSS.Math.Content.3.G.A.1	✓	✓	CCSS.Math.Content.4.G.A.1	✓	✓	CCSS.Math.Content.5.G.A.1	✓	✓
CCSS.Math.Content.3.G.A.2	✓	✓	CCSS.Math.Content.4.G.A.2	✓	✓	CCSS.Math.Content.5.G.A.2	✓	✓
			CCSS.Math.Content.4.G.A.3	✓	✓	CCSS.Math.Content.5.G.B.3	✓	✓
						CCSS.Math.Content.5.G.B.4	✓	✓

Note. Smarter Balanced Claims 2, 3, and 4 were not prepared in compatible forms.