CRESST REPORT 849

ON THE ROAD TO ASSESSING DEEPER LEARNING: WHAT DIRECTION DO TEST BLUEPRINTS PROVIDE?

SEPTEMBER 2015

Joan L. Herman
Deborah La Torre Matrundola
Jia Wang

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

On the Road to Assessing Deeper Learning: What Direction Do Test Blueprints Provide?

CRESST Report 849

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

September 2015

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 2012-8075 from The William and Flora Hewlett Foundation with funding to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

To cite from this report, please use the following as your APA reference: Herman, J. L., La Torre Matrundola, D., & Wang, J. (2015). *On the road to assessing deeper learning: What direction do test blueprints provide?* (CRESST Report 849). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

ON THE ROAD TO ASSESSING DEEPER LEARNING: WHAT DIRECTION DO TEST BLUEPRINTS PROVIDE?

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

Abstract

This study examines the extent to which deeper learning is expected to be present in the new college and career ready (CCR) standards. This is done by examining the distribution of items and tasks at high levels of cognitive demand (DOK3 and DOK4) in the summative test blueprints developed by the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (Smarter Balanced). The study found that while only 10–20% of the consortia's assessment items and tasks appear to require higher levels of cognitive demand, approximately 30–45% of the total possible raw scores are allocated to deeper learning. Furthermore, the analyses indicated that while the end-of-year (EOY) exams are focused on relatively lower level items, components of the performance tasks primarily concentrate on deeper learning and higher levels of thinking. If the consortia maintain the levels of cognitive demand specified in their blueprints, there is no doubt that this will result in an increase in intellectual demand from prior state tests.

Introduction

States, districts, and schools across the country have been preparing for new tests of college and career ready standards. Last year's field testing, conducted by both PARCC and Smarter Balanced, provided a general sense of the increased demands these new tests will bring. "Wow, this is hard!" seemed a common student refrain based on media reports. For some, the exclamation marked dismay; for others, it marked pleasure in being challenged to think and solve problems in new ways.

Preparing for these new expectations, however, requires more than emotional reactions and more than a general sense that they are "harder." What is needed is a better understanding of the ways in which these tests will be more challenging and, particularly, the extent to which they will assess deeper learning. In this report, we use the Partnership for Assessment of Readiness for College and Careers and the Smarter Balanced Assessment Consortium test blueprints to forecast the deeper learning challenges that the new tests will bring and suggest ways in which this analysis might inform curriculum and instruction.

We start by sharing the concept of deeper learning and the metric we are using to gauge its representation. We then describe how publicly available test blueprints enable us to predict how deeper learning likely will be distributed on PARCC and Smarter Balanced operational tests. Finally, we share the results of our analysis and suggest implications for curriculum and instruction.

Background on Deeper Learning

Deeper learning is the concept we use to capture the major changes in learning expectations that today's new college and career standards for English language arts (ELA), mathematics, and science are intended to embody. These standards reflect a general consensus that, to be prepared for success in college and work, students need to develop deeper content knowledge and be better able both to apply their knowledge to think critically and solve complex problems and to communicate their knowledge and skills with others (William and Flora Hewlett Foundation, n.d.). These capabilities are the essence of deeper learning as defined by the new standards.[1] They also characterize the nature of academic knowledge and skills that our new tests must address to be valid measures of and to reinforce the development of the new standards.

How does one determine how well deeper learning is addressed in the new CCR tests? We are using Norman Webb's Depth-of-Knowledge (DOK) classification scheme (see Webb, Alt, Ely, & Vesperman, 2005; http://wat.wceruw.org) to make a determination, because his scheme has been used in prior studies of state tests and thus enables an easy comparison between the new tests and prior practice. The scheme uses the following four levels to characterize the DOK and thinking required to respond to an item:

- DOK1: Recall of a fact, term, concept, or procedure; basic comprehension
- DOK2: Application of concepts and/or procedures involving some mental processing
- DOK3: Applications requiring abstract thinking, reasoning, and/or more complex inferences
- DOK4: Extended analysis or investigation that requires synthesis and analysis across multiple contexts and non-routine problems and applications

We have argued elsewhere that DOK3 and DOK4 represent important aspects of deeper learning, because to answer items or tasks at these levels students have to apply and synthesize their knowledge and engage in critical thinking and reasoning (Herman & Linn, 2013). Further, both levels have been grossly underrepresented in most prior state tests. For example, RAND's analysis of the DOK assessed in released items and tests from the 17 states reputed to have the most challenging state assessments showed that virtually all of the selected and constructed

[1] As defined by the William and Flora Hewlett Foundation (n.d.), deeper learning also includes constructs focusing on collaboration, self-directed learning, and academic mindset that are not directly represented in the standards, nor are they targets for either the PARCC or Smarter Balanced end-of-year summative assessments.

response items in mathematics were categorized as DOK1 or DOK2, with similar results for the selected response items of reading and writing (Yuan & Le, 2012). The situation was better for constructed response items for those states that included such items with more than half the constructed response reading tasks at or above DOK3, and for the eight states that directly assessed writing, the writing prompts were nearly uniformly classified at DOK3 or DOK4.

A Quick Review of Evidence-Centered Assessment Design

To understand how blueprints currently offer an advance view of what the assessments from PARCC and Smarter Balanced will assess, consider the evidence-centered design (ECD) process that both are using to develop their systems. ECD starts with the basic premise that assessment is a process of reasoning from evidence to evaluate specific claims about student capability. In essence, student responses to assessment items and tasks provide the evidence for the reasoning process, and psychometric and other validity analyses establish the sufficiency of the evidence for substantiating each claim (see Pellegrino, Chudowsky, & Glaser, 2001).

PARCC	Smarter Balanced
1. **Reading:** Students read and comprehend a range of sufficiently complex texts independently.	1. **Reading:** Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts.
2. **Writing:** Students write effectively when using and/or analyzing sources.	2. **Writing:** Students can produce effective and well-grounded writing for a range of purposes and audiences.
3. **Research:** Students build and present knowledge through research and the integration, comparison, and synthesis of ideas.	3. **Speaking and Listening:** Students can employ effective speaking and listening skills for a range of purposes and audiences.
	4. **Research/Inquiry:** Students can engage in research/inquiry to investigate topics, and to analyze, integrate, and present information.

Figure 1. PARCC and Smarter Balanced claims for the ELA summative assessments.

The ECD process starts with a clear delineation of the claims that are to be evaluated and the evidence that can be used to substantiate the claims, which provides a clear and transparent foundation for assessment development. Both PARCC and Smarter Balanced have reorganized the Common Core State Standards into core claims about student competency in ELA and mathematics that their tests are designed to evaluate. Both start with an overall claim about students becoming college and career ready and then subdivide these overall expectations into more specific subclaims for ELA and mathematics. Figure 1 and Figure 2 summarize the PARCC and Smarter Balanced claims for both subject areas. (We return later to an analysis of these claims.)

PARCC	Smarter Balanced
1. **Major Content with Connections to Practices:** The student solves problems involving the Major Content for the grade/course with connections to the Standards for Mathematical Practice.	1. **Concepts and Procedures:** Students can explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency.
2. **Additional and Supporting Content with Connections to Practices:** The student solves problems involving the Additional and Supporting Content for the grade/course with connections to the Standards for Mathematical Practice.	2. **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.
3. **Highlighted Practices MP.3 and 6 with Connections to Content (expressing mathematical reasoning):** The student expresses grade/course level appropriate mathematical reasoning by constructing viable arguments, critiquing the reasoning of others and/or attending to precision when making mathematical statements.	3. **Communicating Reasoning:** Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others.
4. **Highlighted Practice MP.4 with Connections to Content (modeling/application):** The student solves real-world problems with a degree of difficulty appropriate to the grade/course by applying knowledge and skills articulated in the standards for the current grade/course (or for more complex problems, knowledge and skills articulated in the standards for previous grades/courses), engaging particularly in the Modeling practice, and where helpful making sense of problems and persevering to solve them, reasoning abstractly and quantitatively, using appropriate tools strategically, looking for the making use of structure, and/or looking for and expressing regularity in repeated reasoning.	4. **Modeling and Data Analysis:** Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems.
5. **Fluency:** The student demonstrates fluency in areas set forth in the Standards for Content in grades 3–6.	

Figure 2. PARCC and Smarter Balanced claims for the mathematics summative assessments.

Each claim is further defined by specific evidence statements (PARCC) or assessment targets (Smarter Balanced) that the claim encapsulates. These statements or targets essentially represent particular Common Core standards or clusters of standards, and for Smarter Balanced also indicate the DOK level at which each target may be assessed.

The targets and evidence statements become the subjects of item or task specifications. The specifications provide guidance and rules for item writers to follow in developing items that address each assessment target or evidence statement. The ideal specification provides sufficient guidance so that two item writers working independently from the same specification would generate essentially comparable items or tasks for a given assessment target or evidence statement—such that students would be expected to perform similarly on both.

Item writers then use the specifications to generate items and tasks, which in turn are subjected to content and bias reviews as well as pilot testing. Items and tasks that survive this process as substantively and psychometrically sound are then assigned to test forms according to blueprints. These blueprints provide the rules for assembling items that will be administered to students so that the operational test forms will adequately represent the claims and range of evidence required to draw valid inferences about student proficiency relative to the claims. Test forms are then field tested and additional reliability and validity studies conducted.

Test Blueprints: Plans for Assembling Items Into Test Forms

Although the PARCC and Smarter Balanced operational tests were implemented for the first time in spring 2015, the blueprints they used to create the test forms were made available to the public in previous years (see http://www.parcconline.org/assessments/test-design/ela-literacy/test-specifications-documents; http://www.parcconline.org/assessments/test-design/mathematics/math-test-specifications-documents; http://www.smarterbalanced.org/smarter-balanced-assessments).[2] The blueprints lay out the standards content represented in each test form, by grade level, and give some indication of the depth of knowledge at which the assessment targets will addressed.

Blueprint formats and specifications vary for the two consortia, in part because of differences in their on-demand test designs. Although both systems include both end-of-year (EOY) on-demand and performance assessment components, the on-demand tests from Smarter Balanced utilize computer-adaptive testing (CAT), which essentially individualizes the items that are administered to students based on their prior responses. PARCC, in contrast, will use fixed form assessments, which are common across students.

The Smarter Balanced blueprints are organized by claim area and specify the number of CAT and/or performance assessment items that will be included in each content category (e.g., literary versus informational text within reading) and indicate the number of items that will address each assessment target or group of assessment targets. The blueprints indicate the depth of knowledge at which each assessment target can be assessed—more than one level can be specified—and establish a minimum number of items at DOK3 and/or DOK4 for each test, among other details.

PARCC provides similar levels of detail in its English language arts blueprints, termed Common Forms Specifications, but organizes them by task type (e.g., literature analysis,

[2]Original analyses for this report were completed using earlier draft versions of the PARCC (2013) and Smarter Balanced (2014) blueprints. For purposes of this report, analyses were redone using the recently released 2015 versions of the PARCC ELA and Smarter Balanced ELA and mathematics blueprints.

research). Cognitive complexity ratings, however, are not yet provided to the public. In mathematics, PARCC uses its three task types to organize its High Level Blueprints by grade, which in turn are linked to the claims that PARCC's assessment is designed to evaluate. Type I items or tasks are intended to assess basic mathematical concepts skills and procedures; Type II are intended to assess mathematical reasoning and ask for written arguments and justifications; and Type III math tasks are intended to assess modeling and real world problem solving applications. Although the blueprints do not specify DOK levels or other cognitive complexity distribution, it is possible to infer them from the description of task types and sample items PARCC provides.[3] We will discuss this further in our analysis.

Depth-of-Knowledge Expectations Evidence in Blueprints

We used the information in the blueprints to estimate the extent to which deeper learning will be represented in the Smarter Balanced and PARCC summative assessments, using two different but related metrics. The first is the proportion of items or tasks that are likely to be at DOK3 or DOK4. The second is the proportion of the total raw score value that will be accounted for by items and/or tasks at DOK3 or DOK4. The latter value takes into account the higher score values that are often associated with tasks at higher levels of complexity, for example those that call for an explanation and/or more extended performance tasks. Data are reported by claim, where possible, as this is the level at which individual scores will be reported, based on current plans.

Smarter Balanced ELA Summative Assessments

Table 1 shows the distribution of DOK3 and/or DOK4 items for Smarter Balanced ELA CAT and performance task assessments for reading, writing, speaking and listening, and research claims. Because the distributions are identical across all grades for the performance task components and nearly so for the CAT component, this table shows the mean distribution across elementary, middle, and high school (see Appendix for results for each grade span). All items and tasks include given stimuli (e.g., literary texts, informational texts) and may involve the analysis of multiple texts. Although we focus in this report on the DOK findings, the data provide important content information. Forty percent of the CAT items will address the reading claim and its constituent assessment targets, while writing and research will be the focus of the performance tasks. As would be expected, the data also indicate that a relatively small proportion of the CAT items are at or above DOK3, because the CAT format is not conducive to the kinds

[3] See http://www.parcconline.org/ela-plds and http://www.parcconline.org/math-plds for overviews of the PARCC Cognitive Complexity Frameworks for ELA/Literacy and Mathematics.

of extended response tasks that typically address higher DOK levels. In contrast, the performance task component is centered on these higher level applications.

Table 1

Smarter Balanced ELA Summative Assessment: Means Across Grades, Number, and Percentage of Items by Claim, Total, and Minimum at DOK3/4

Claim	Mean # items	% Total items	Min # DOK3/4 items	Min % DOK3/4 items
CAT component				
Claim 1: Reading	16.3	40%	2.3	14%
Claim 2: Writing	10.0	24%	1.0	10%
Claim 3: Speaking and listening	9.0	22%	0.0	0%
Claim 4: Research	6.0	15%	0.0	0%
Subtotal	41.3	100%	3.3	8%
Performance task component				
Claim 1: Reading	0.0	0%	n/a	n/a
Claim 2: Writing	3.0	50%	3.0	100%
Claim 3: Speaking and listening	0.0	0%	n/a	n/a
Claim 4: Research	3.0	50%	2.0	100%
Subtotal	6.0	100%	6.0	100%

Note. Our analyses do not differentiate between DOK3 and DOK4 because specifications did not consistently differentiate the two in setting minimums.

The representation of DOK in English language arts, however, changes substantially when one considers the proportion of total raw score points that will be based on these higher DOK levels. The difference, of course, occurs because items that are more complex typically are worth more points—for example, CAT items are often scored on a rubric from 0–2 points (partially to fully correct), while essay rubrics have more extended 1–4 point scales. Based on the blueprints, coupled with our analysis of the points associated with items and tasks in sample items and practice tests, we assume the following: (a) The CAT items at DOK3 are typically worth 2 points each. (b) The writing task in the performance assessment will be worth 10 points (i.e., 4 points for organization/purpose, 4 points for evidence/elaboration, and 2 points for conventions). (c) The performance assessment items addressing research will be worth 2 points. Based on these assumptions, we estimate, for every student who takes the assessment, the following minimal proportions of their total possible raw score will be based on DOK3 and/or DOK4:

- Elementary school (Grades 3–5): Minimally 36%
- Middle school (Grades 6–8): Minimally 39%

- High school (Grades 9–12): Minimally 39%

As noted, these are based on the minimums that Smarter Balanced has established in its blueprint specifications. However, for some students, the percentages may be much higher, if based on their prior item performance, the CAT algorithm assigns high-level items beyond the stated minimum.

Based on the blueprints and the possible DOK levels at which each target *could* be assessed, we also estimated the maximum percentage of total score points that could be attributed to DOK3 and DOK4 items for a student. Here we assumed that any target that might be assessed at Level 3 or 4 would be—for example, a target that was specified at DOK2 and DOK3 was assumed to be assessed at DOK3, and a target specified at DOK3 or DOK4 was assumed to be assessed at DOK4. Based on these assumptions, more than two thirds of the total possible score could be based on DOK3 and/or DOK4. It is unlikely that any student will be assessed at the maximum, but it is likely that students will be administered more than the minimum and that somewhere in the middle—approximately 50% of the total score—might be a reasonable estimate.

Smarter Balanced Math Summative Assessments

Table 2 shows the distribution of DOK3 and DOK4 items for the Smarter Balanced math CAT and performance task components for the concepts and procedures, problem solving and modeling, and communicating reasoning claims.[4] As with ELA, because the distributions for the performance task components are identical across grades and those for the CAT component are very similar, the table summarizes mean distributions across elementary, middle, and high school (see Appendix for results for each grade span).

Here, as in ELA, we see less attention to DOK3 and DOK4 in the CAT component. More specifically, DOK3 and DOK4 are not represented for Claim 1, which constitutes the majority of the CAT items, but are represented in sizable proportions for the relatively few items addressing Claims 2/4 and 3. These latter claims are the focus of the performance assessment component, where half the items are at least at DOK3.

Again, the picture changes when we examine the proportion of total possible score points associated with items at the highest DOK levels, because DOK3 and DOK4 items and tasks are worth more points than those at lower DOK levels. Based on the blueprints, and coupled with our analysis of the points associated with items and tasks in sample items and practice tests, we assume that CAT items at DOK3 will be worth 2 or 3 points each, with a mean of 2.5. We also

[4]Smarter Balanced has combined Claims 2 and 4 for the purpose of score reporting.

assume that higher level items on the performance task component will be evenly split between DOK3 and DOK4, with an average value of 3 points each. With these assumptions in mind, we estimate that, for all students, the following minimal proportions of their Smarter Balanced math scores will be based on DOK3 and/or DOK4 items:

- Elementary school (Grades 3–5): Minimally 38%
- Middle school (Grades 6–8): Minimally 39%
- High school (Grades 9–12): Minimally 37%

Table 2

Smarter Balanced Math Summative Assessment: Means Across Grades, Number, and Percentage of Items by Claim, Total, and Minimum at DOK3/4

	CAT component			
Claim	Mean # items	% Total	Min # DOK3/4	Min % DOK3/4
CAT component				
Claim 1: Concepts and procedures	20.1	59%	0.0	0%
Claim 2/4: Problem solving & modeling and data analysis	6.0	18%	2.0	33%
Claim 3: Communicating reasoning	8.0	23%	2.0	25%
Subtotal	34.1	100%	4.0	12%
Performance task component				
Claim 1: Concepts and procedures	0.0	0%	n/a	n/a
Claim 2/4: Problem solving & modeling and data analysis	4.0	67%	2.0	50%
Claim 3: Communicating reasoning	2.0	33%	0.0	0%
Subtotal	6.0	100%	2.0	33%

Note. Our analyses do not differentiate between DOK3 and DOK4 because specifications did not consistently differentiate the two in setting minimums.

These proportions are very similar to those for ELA and again represent the minimum of what every student will be administered. The maximum that a student *could* see is considerably higher—approximately 70% of all assessment targets designated as possibly DOK3 or DOK4 were assessed at these highest levels. However, as with the ELA, the likely maximum is lower.

PARCC ELA Summative Assessments

Table 3 and Table 4 reorganize the data in the PARCC blueprints to display the number and percentage of items, score points, and total possible raw score points representing each of PARCC's major claims for its ELA summative assessment. Since the PARCC blueprints do not include explicit DOK or cognitive complexity specifications, we use score points accorded to various item types as a rough indicator.

As shown in Table 3, the PARCC EOY test concentrates on their reading claim and, across grades, each item is worth 2 points. In Grades 3–5, the assessment is to be composed of 10 items for a total of 20 points, and for Grades 6–12 there will be 16 items for a total of 32 points. With all items, students are required to read given texts.

Table 3

PARCC ELA Summative Assessment: Distribution of EOY Items by Claim

	EOY component		
Claim by grade	# Items	% EOY total items	# Score points
Elementary school: Grades 3–5			
Claim: Reading	10	100%	20
Claims: Writing & research	0	0%	0
Subtotal	10	100%	20
Middle and high school: Grades 6–12			
Claim: Reading	16	100%	32
Claims: Writing & research	0	0%	0
Subtotal	16	100%	32

Table 4 shows the distribution of PARCC's performance assessment component by grade and claim. At each grade level, the assessment is composed of three major tasks—one involving literary analysis of two given texts, one involving a research simulation based on three given texts, and a narrative task based on one short text. Within each task type, students respond to reading questions about each text for 2 points, and then respond to a prose constructed response (PCR) task. The resulting essay is scored for written expression, knowledge of language and conventions, for literary analysis and research tasks, and for reading/use of evidence. Depending on the grade level, total possible scores on PCR tasks for reading range from 3 to 4 points (Grades 3–5 and Grades 6–12, respectively); for written expression from 9 to 12 points (Grades 3–5 and Grades 6–12, respectively); and knowledge of language and conventions have a possible score of 3 points. As with the Smarter Balanced performance tasks, the PARCC performance component addresses research and writing, but also incorporates significant attention to its reading claim.

Table 4

PARCC ELA Summative Assessment: Distribution of Performance Assessment Items by Claim

Claim by grade	Total # items	# Short items (2 points)	# Higher level PCR items	% Higher level PCR items
Elementary school: Grade 3				
Claim: Reading	18	16	2	11%
Claim: Writing	3	0	3	100%
Claim: Research[a]	(7)	(6)	(1)	(100%)
Subtotal	21	16	5	24%
Elementary school: Grades 4-5				
Claim: Reading	20	18	2	10%
Claim: Writing	3	0	3	100%
Claim: Research[a]	(9)	(8)	(1)	(100%)
Subtotal	23	18	5	22%
Middle and high school: Grades 6-12				
Claim 1: Reading	18	16	2	11%
Claim 2: Writing	3	0	3	100%
Claim 3: Research[a]	(9)	(8)	(1)	(100%)
Subtotal	21	16	5	24%

[a]For purposes of assessment, the research claim is subsumed under those for reading and writing; every student takes one research simulation task. Because of this, the counts and percentages for the research claim are presented in parentheses and are not included in the subtotals.

To get a rough estimate of the percentage of items and total score value that is based on high complexity level items, and again based on our analysis of sample items and practice tests, we conservatively estimate that 2-point items will likely reflect DOK1 and/or DOK2, and that the PCR tasks will be at DOK3 or DOK4.[5] Based on these assumptions, Table 5 shows the score values associated with these item types by claim and grade level. These data indicate that approximately 45% of students' total possible ELA scores on the summative assessment will be based on items and tasks that tap high-level cognitive demands. This estimate also takes into account that approximately one third of a student's total score on the tests should represent each level on the PARCC cognitive complexity framework for ELA/Literacy (i.e., low, medium, high; PARCC, personal communication, October 3, 2014).

[5]Based on a review of PARCC sample items, it seems likely that some of the 2-point items that address relatively high levels of text complexity and/or which draw on multiple texts reflect DOK3.

Table 5

PARCC ELA Summative Assessment: Percentage of Total Possible Score Points From Higher Level Items (PCR Tasks) for Both the EOY and Performance Assessment

	Total possible score			
Claim by grade	2 point items	3–12 point items	All items	% High-level PCR items
Elementary school: Grade 3				
Claim: Reading	52	6	58	10%
Claim: Writing	0	36	36	100%
Claim: Research[a]	(0)	(15)	(15)	(100%)
Subtotal	52	42	94	45%
Elementary school: Grades 4-5				
Claim: Reading	56	8	64	13%
Claim: Writing	0	42	42	100%
Claim: Research[a]	(0)	(19)	(19)	(100%)
Subtotal	56	50	106	45%
Middle and high school: Grades 6-12				
Claim: Reading	68	8	76	11%
Claim: Writing	0	45	45	100%
Claim: Research[a]	(0)	(19)	(19)	(100%)
Subtotal	68	53	121	44%

[a]For purposes of assessment, the research claim is subsumed under those for reading and writing; every student completes one research simulation task, which totals 12 points. Because of this, the counts and percentages for the research claim are presented in parentheses and not included in the subtotals.

PARCC Mathematics Summative Assessments

Table 6 and Table 7 reorganize the data in the PARCC math blueprints by claim for the EOY and performance assessment components. The tables display by claim and grade level the number of items planned at each score value, the total number of items and possible raw score points, and the percentage of items expected at high levels of cognitive complexity. Again, in the absence of direct specification of cognitive complexity, we use score value as a rough indicator. Similar to PARCC's ELA assessment, we consider score values of three and above as addressing higher DOK levels and deeper learning.

Table 6

PARCC Math Summative Assessment: Distribution of EOY Items by Claim

	Mean #					
Claim by grade	1 point	2 points	4 points	Total # items	Total score points	% High-level items
Elementary school: Grade 3	34.00	5.00	0.00	39.00	44	0%
Claim A: Major content with connections to practice						
Claim B: Additional and supporting content with connections to practice						
Claim E: Demonstrate fluency						
Elementary school: Grades 4-5	28.00	8.00	0.00	36.00	44	0%
Claim A: Major content with connections to practice						
Claim B: Additional and supporting content with connections to practice						
Claim E: Demonstrate fluency						
Middle school: Grades 6-8	25.33	6.67	1.33	33.33	44	3%
Claim A: Major content with connections to practice						
Claim B: Additional and supporting content with connections to practice						
Claim E: Demonstrate fluency						
High school[a]	19.33	12.17	2.83	34.33	55	5%
Claim A: Major content with connections to practice						
Claim B: Additional and supporting content with connections to practice						
Claim E: Demonstrate fluency						

[a]High school summarizes course specifications for Algebra 1, Math I, Geometry, Math II, Algebra 2, and Math III.

As Table 6 shows, the EOY test concentrates on PARCC's three mathematics claims:

- Claim A: Major content with connections to practices
- Claim B: Additional and supporting content with connections to practices
- Claim E: Fluency

As noted earlier, PARCC organized their math blueprint by task type and do not differentiate the number of items addressing each claim. Instead, the blueprint specifies that

Claims A, B, and E—the more basic knowledge-oriented standards—will all be addressed by Type I items. They may also involve mathematical practices.

Given the emphasis on claim and task type, it is not surprising to see that the EOY exam generally concentrates on basic knowledge and application levels. However, the exam does appear to call on progressively more and some deeper applications as the student's grade level advances.

Table 7 reveals that PARCC's performance assessment component addresses the following claims, as well as the claims addressed in the EOY assessment:

- Claim C: Highlighted practices MP.3 and 6 with connections to content (expressing mathematical reasoning)
- Claim D: Highlighted practice MP.4 with connections to content (modeling/ application)

The performance assessments are composed of previously defined Type I, II and III items, with the latter two devoted to Claims C and D. As a reminder, Type II and III items likely align with DOK3, as at least half their point values are awarded based on the quality of student reasoning and/or modeling.

Table 7 shows that, based on the PARCC blueprint, all items addressing Claims C and D will reflect deeper levels of learning, while those assessing Claims A, B, and E, as expected, remain focused on basic knowledge and application. In terms of the proportion of items, those addressing basic knowledge and applications constitute more than half the performance assessment component, while those communicating reasoning (Claim C) and modeling (Claim D) each draw less than a quarter of the items. Allocations of higher level items are roughly similar for elementary and middle school grades, but increase slightly at the high school levels because of the increased demands specified for Algebra 2 and Math III courses.

Table 7

PARCC Math Summative Assessment: Distribution of Performance Assessment Items by Claim

	Mean #								
Claim by grade	1 point	2 points	3 points	4 points	6 points	Total # items	Claim % of total items	Claim % of high-level items	Total score points
Elementary school: Grades 3–5									
Claims A, B, E (Type I)	7.33	2.33	0.00	0.00	0.00	9.67	58%	0%	12.00
Claim C: Expressing mathematical reasoning	0.00	0.00	2.00	2.00	0.00	4.00	24%	100%	14.00
Claim D: Modeling/application	0.00	0.00	2.00	0.00	1.00	3.00	18%	100%	12.00
Subtotal	7.33	2.33	4.00	2.00	1.00	16.67	100%	41%	38.00
Middle school: Grades 6–8									
Claims A, B, E (Type I)	8.67	1.67	0.00	0.00	0.00	10.33	60%	0%	12.00
Claim C: Expressing mathematical reasoning	0.00	0.00	2.00	2.00	0.00	4.00	23%	100%	14.00
Claim D: Modeling/application	0.00	0.00	2.00	0.00	1.00	3.00	17%	100%	12.00
Subtotal	8.67	1.67	4.00	2.00	1.00	17.33	100%	41%	38.00
High school[a]									
Claims A, B, E (Type I)	10.00	0.00	0.00	0.00	0.00	10.00	54%	0%	10.00
Claim C: Expressing mathematical reasoning	0.00	0.00	2.00	2.29	0.00	4.29	23%	100%	15.14
Claim D: Modeling/application	0.00	0.00	2.00	0.00	2.33	4.33	23%	100%	20.00
Subtotal	10.00	0.00	4.00	2.29	2.33	18.62	100%	46%	45.14

[a]High school summarizes course specifications for Algebra 1, Math I, Geometry, Math II, Algebra 2, and Math III. Relative to the other courses, specifications for Algebra 2 and Math II include an additional high-level item for each of Claims C and D.

Again, however, a truer picture of how much deeper learning counts is found in the weight given to higher level items in computing students' total score. Based on our assumptions about the relationship between score points accorded to an item and/or task type, Table 8 shows the distribution of item score values by grade level and claim and the proportion of score values that can be attributed to higher level items. Here we see a progression of increasing weight being given to higher level items, from approximately one third of the total score value in elementary school to approaching half at the high school level. Again, this increase at the high school level is due to the additional higher level items specified for Algebra 2 and Math III. The weight through Algebra I, Geometry, Math I, and Math II is similar to the middle school allocations.

Table 8

PARCC Math Summative Assessment: Percentage of Total Possible Score Points from Higher Level Items (PCR Tasks) for Both the EOY and Performance Assessment

	Total possible score			
Claim by grade	1–2 point items	3–6 point items	All items	% High-level PCR items
Elementary school: Grades 3–5				
Claims A, B, E (Type I)	56.00	0.00	56.00	0%
Claim C: Expressing mathematical reasoning	0.00	14.00	14.00	100%
Claim D: Modeling/application	0.00	12.00	12.00	100%
Subtotal	56.00	26.00	82.00	32%
Middle school: Grades 6–8				
Claims A, B, E (Type I)	50.67	5.32	55.99	10%
Claim C: Expressing mathematical reasoning	0.00	14.00	14.00	100%
Claim D: Modeling/application	0.00	12.00	12.00	100%
Subtotal	0.00	31.32	81.99	38%
High school[a]				
Claims A, B, E (Type I)	53.66	11.33	65.00	17%
Claim C: Expressing mathematical reasoning	0.00	15.14	15.14	100%
Claim D: Modeling/application	0.00	20.00	20.00	100%
Subtotal	0.00	46.48	100.14	46%

[a]High school summarizes course specifications for Algebra 1, Math I, Geometry, Math II, Algebra 2, and Math III. Relative to the other courses, specifications for Algebra 2 and Math II include an additional high-level item for each of Claims C and D.

Concluding Thoughts

This report shares an analysis of the PARCC and Smarter Balanced blueprints to project the extent to which deeper learning will be reflected in the consortia's summative assessment systems. We grounded our analysis in the evidence-centered design (ECD) process utilized by both consortia, and used the distribution of items and tasks at high levels of cognitive demand (DOK3 and DOK4) as indicators of deeper learning. We believe that several implications and calls for action are evident in our findings.

Study Implications

Methodology. A very different picture of representation of deeper learning emerges when one considers the percentage of the total raw score that is attributable to higher level items rather than metrics based on the number or proportion of items. In examining the former, only 10–20% of the consortia's assessment items and tasks appear to require higher levels of thinking, but based on the analysis of the raw score values associated with these items, approximately 30–45% of the total possible raw score is allocated to deeper learning. Historically, counts and proportions of items have been used in considering the alignment between standards and assessment (see, for example, Webb et al., 2005; Yuan & Le, 2012). With new tests, which include technology-enhanced and other new formats that vary item score values, the field needs to move to new metrics for conducting alignment studies. At this point in the consortia's development process, we believe that raw score value provides the better indicator. Down the line, based on operational tests, however, it will also be important to examine the weight given deeper learning when raw scores are scaled and converted to the scale scores that are used for reporting and comparison.

Performance assessments. Study analyses also make clear the strong relationship between the performance assessment components and opportunities to assess deeper learning. For both consortia's assessments and across both English language arts and mathematics, our analysis indicates that the bulk of the EOY exams are focused on relatively lower level items, while the performance task components concentrate on tasks that draw on deeper learning and higher levels of thinking. While this relationship is not a surprise, it is worth underscoring as next generation tests, in addition to PARCC and Smarter Balanced, are produced. Without a performance assessment or extended response component, any test will have difficulty incorporating deeper learning goals—or at least the depth of knowledge of tests without a performance assessment component will require serious scrutiny.

Increased intellectual demand. Perhaps the most telling implication of the study involves the dramatic increase in intellectual demand that the PARCC and Smarter Balanced tests will bring, and what this increase portends for student performance. If these consortia hold to their plans, there is no doubt that their tests will be much more demanding than those most students have previously faced. Recall that prior studies have shown that most current state tests focus nearly exclusively on lower levels, and only those states that include extended constructed responses (only 8 of 50 in one study) consistently reached higher levels (Webb, 2002a, 2002b; Yuan & Le, 2012). The low levels of state tests, in fact, were a motivating factor in the federal government's substantial investment in PARCC and Smarter Balanced. Faced with increased intellectual challenge, many students likely will have difficulty performing well and in fact, Kentucky and New York, the two states that already have transitioned to CCR-aligned tests, have seen test scores plummet. A drop in scores is to be expected, and the public needs to be prepared, as many others have noted.

Students are unlikely to perform well because they previously have not had the opportunity to learn and attain the new, more demanding college and career ready standards that the new tests address. Importantly, schools and teachers have not had a full opportunity to learn how to teach the new standards nor the resources to do so, according to surveys of states, districts, and teachers (see Rentner, 2013; Rentner & Kober, 2014; Scholastic & the Bill and Melinda Gates Foundation, 2014). These studies show that standards implementation is well underway in most places, and teachers are positive, but teachers also indicate that they and their students need support to attain success.

Opportunities for Action

We believe our findings also have implications for action moving forward. Both new college and career ready standards and new tests of them require deeper learning. Adapting to these new standards is not simply a matter of alignment to the surface content details of the new standards, but must enable students to apply, communicate, and extend their knowledge and skill to solve complex problems and meet new situations. This is the essence of the higher levels of DOK: At these levels, students must be able to go beyond the basic concepts and procedures they have learned to use and integrate this knowledge to think critically, reason with evidence, and explain their thinking. Teachers need to be able to incorporate these higher level demands into their teaching and to engage students in instructional activities and assessments that ask students to extend their learning. One step in making this transition involves educating teachers about the types of item prompts that focus on higher levels of DOK. For example, teachers who want to use DOK3 and DOK4 items should consider using prompts such as those shown in Figure 3. (See Hess, 2013, for additional guidance and examples.)

Level	Language arts	Mathematics
DOK3	For tasks requiring complex text-based inferences: • Explain the most likely reason why . . . . Use evidence from text to support your reasoning. • Select the sentences/paragraphs that best show the author's point of view. • Summarize the central idea of a complex passage . . . . • What conclusions can you draw from . . . . Support your answer with evidence from the text. • Why did the author use . . . ? • Write a continuation of a story	For tasks requiring the integration of multiple concepts and/or procedures: • What decision do the data support? Explain. • Explain how/whether/why or why not . . . . • What relationships do you find in the data? Use them to predict what will happen. • Show the reasoning you used . . . . • Using evidence, explain how you know that . . . . • Which graph/equation represents . . . ?
DOK4	• Use the evidence from two sources to develop and support a claim. • Using more than one source, develop a thesis to explain about . . . . • Write an essay about the similarities and differences between two stories . . . . • Use multiple sources to formulate and answer a research question.	In response to a novel or ill-defined problem: • Create a model to describe the situation. • Explain why your design is better. • Conduct a survey to answer . . . . • How could you minimize . . . . Explain your thinking. • What is the most cost-effective plan? Explain.

Figure 3. Sample item prompts to incorporate higher level demands.

Language arts examples. The following items from PARCC's Grade 8 Practice Test (2014) are used to illustrate the changes in demands as an assessment item moves from DOK2 to DOK3 and DOK4. In each of these items, the student is asked to read and analyze a 34-paragraph passage from the novel *Confetti Girl*, which focuses on the differences in perspective of a daughter and father. In Figure 4, the item asks the student to answer two questions concerning the meaning of the word sarcasm in the passage. This requires the student to use contextual cues within the specified passage to first define and then provide an example of sarcasm, both of which require some level of mental processing beyond recall or reproduction. In contrast, Figure 5 includes an adaptation of the same item that has less scaffolding, involves some higher level processing, and requires the student to provide a short explanation. More specifically, students are now provided with a quote from the passage involving sarcasm and have to discern what it indicates about the girl's relationship with her father. Finally, Figure 6 presents an extended version of the same basic activity where the student must write a multi-paragraph essay where they have to synthesize the tension in *Confetti Girl* with the tension presented in a passage from a second novel.

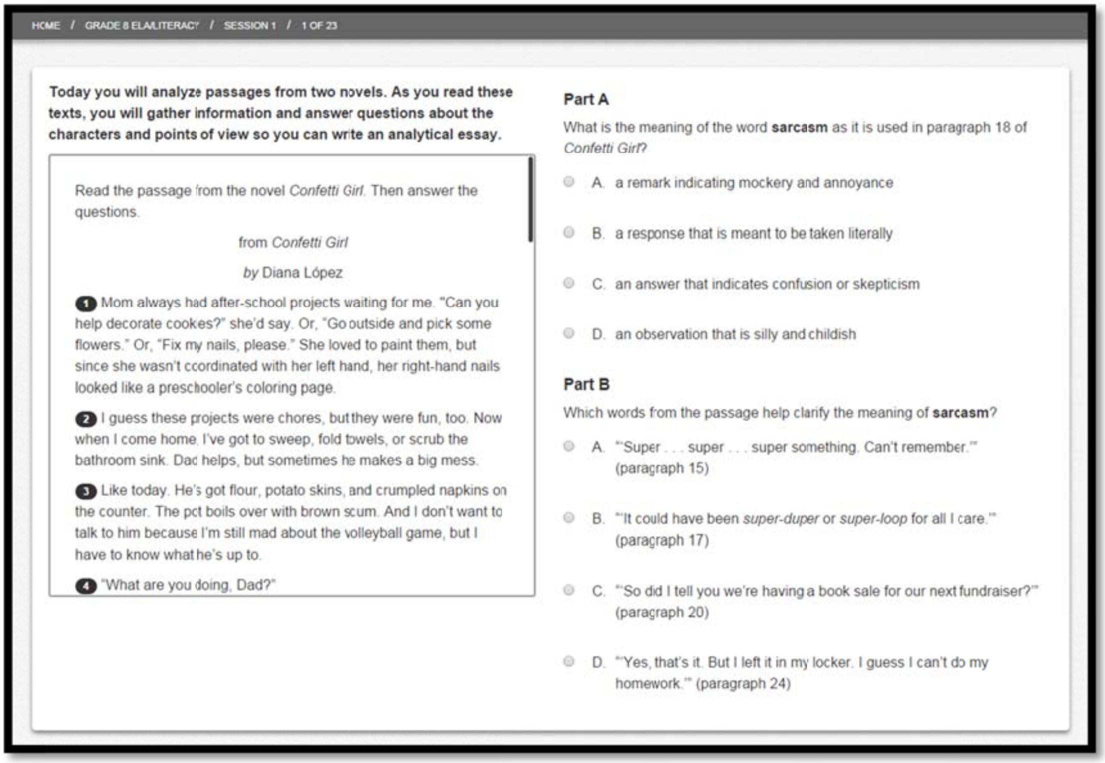


Figure 4. DOK2 ELA item on reading. Item 1 from Grade 8 ELA Practice Test, by PARCC (2014). Retrieved from http://parcc.pearson.com/practice-tests/english/

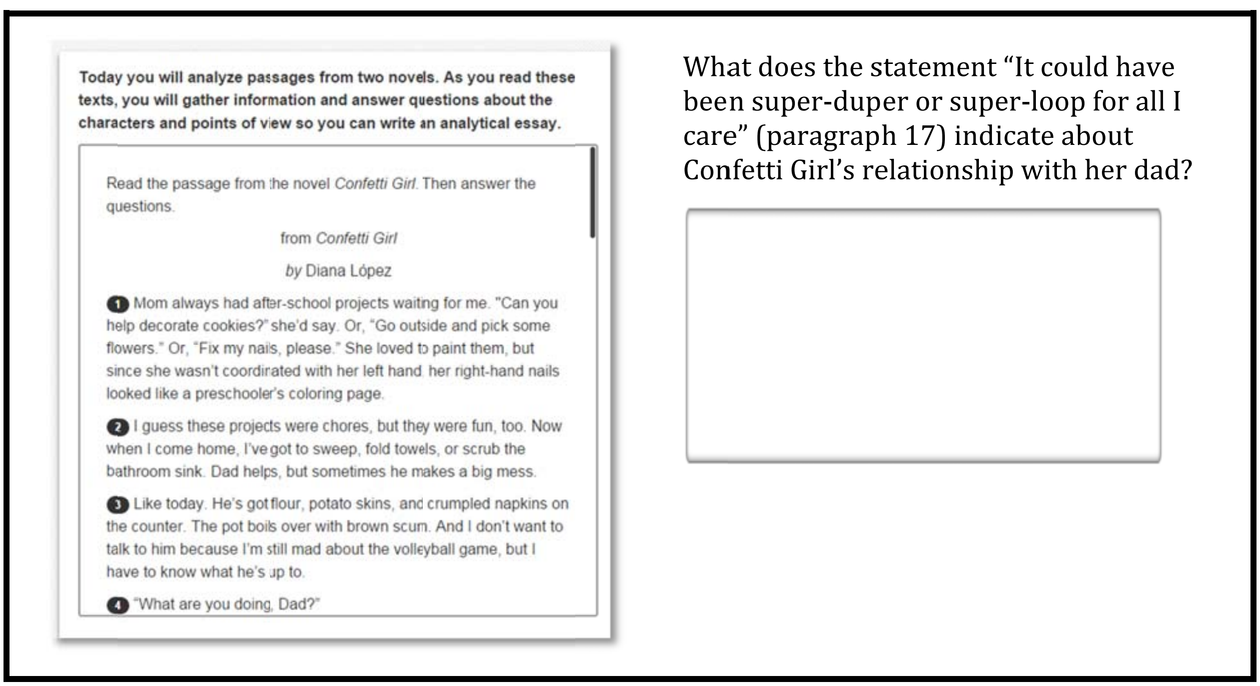


Figure 5. DOK3 ELA item on reading. Adapted from Item 1 from Grade 8 ELA Practice Test, by PARCC (2014). Retrieved from http://parcc.pearson.com/practice-tests/english/

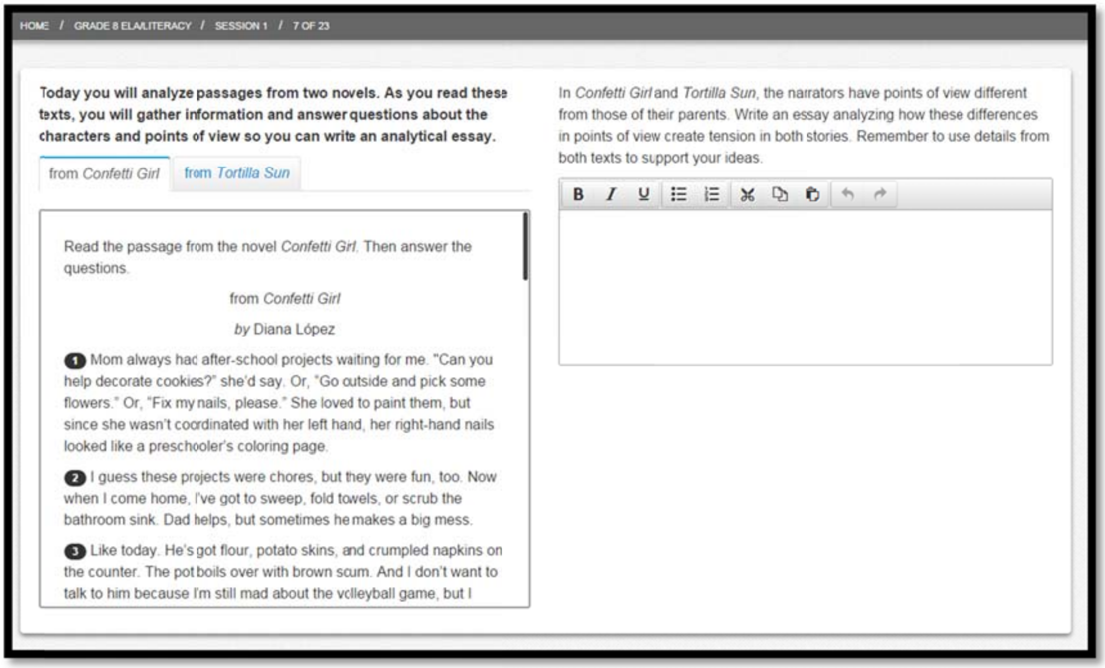


Figure 6. DOK4 ELA item on reading. Item 7 from Grade 8 ELA Practice Test, by PARCC (2014). Retrieved from http://parcc.pearson.com/practice-tests/english/

Mathematics examples. The following items from the Smarter Balanced Grade 8 Practice Test (2014) and Sorting Equations and Identities from the Mathematics Assessment Resource Service (Mathematics Assessment Resource Service [MARS], Shell Center, University of Nottingham, 2012) are used to illustrate the changes in demands as an assessment item moves from DOK2 to DOK4.

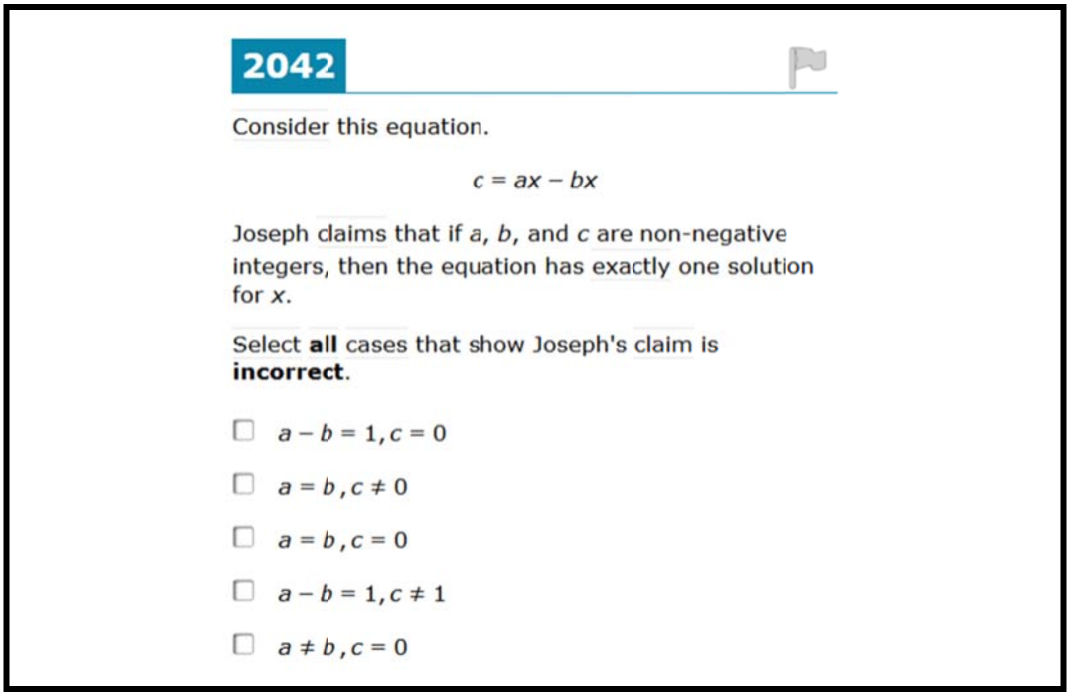
Write down an example of an equation that has:

a. One solution. ____________________

b. An infinite number of solutions. ____________________

c. No solutions. ____________________

Figure 7. DOK2 mathematics item on linear equations. Item adapted from Sorting Equations and Identities (FAL13) Pre-Post Test; by MARS, Shell Center, University of Nottingham (2012). Retrieved from http://map.mathshell.org/materials/download.php?fileid=688


2042

Consider this equation.

$$c = ax - bx$$

Joseph claims that if a, b, and c are non-negative integers, then the equation has exactly one solution for x.

Select **all** cases that show Joseph's claim is **incorrect**.

- ☐ $a - b = 1, c = 0$
- ☐ $a = b, c \neq 0$
- ☐ $a = b, c = 0$
- ☐ $a - b = 1, c \neq 1$
- ☐ $a \neq b, c = 0$

Figure 8. DOK3 mathematics item on linear equations. Item 2042 from Grade 8 Mathematics Practice Test, by Smarter Balanced Assessment Consortium (2014). Retrieved from http://sbac.portal.airast.org/wp-content/uploads/2013/08/G8_Practice-Test-Scoring-Guide-8-30-14-Final.pdf

Consider this equation.

$$c = ax - bx$$

Joseph claims that if a, b, and c are non-negative integers, then the equation has exactly one solution for x.

Kim disagreed with Joseph, claiming that if a, b, and c are non-negative integers, then the equation has no solutions for x.

Do you think either Joseph or Kim are correct?

If you think Joseph is wrong, explain the mistake and how many solutions you think x has in this equation.

If you think Kim is wrong, explain the mistake and how many solutions you think x has in this equation.

Figure 9. DOK4 mathematics item on linear equations. Item adapted from Grade 8 Mathematics Practice Test, by Smarter Balanced Assessment Consortium (2014). Retrieved from http://sbac.portal.airast.org/wp-content/uploads/2013/08/G8_Practice-Test-Scoring-Guide-8-30-14-Final.pdf

In each of these items, students are asked to work with examples of linear equations with one solution, infinitely many solutions, or no solutions (CCSS.Math.Content.8.EE .C.7a). In the DOK2 version of this task, shown in Figure 7, students have to apply their knowledge of linear equations to write examples of each type of equation. In contrast, the task shown in Figure 8 explores this standard at the next DOK level by providing a false claim, and then asking the student to identify which of the five given cases disprove the claim. Not only does this problem have multiple answers (Options 2 and 3), but it requires the student to think abstractly as they evaluate the claim. Finally, Figure 9 extends this problem by asking the student to evaluate a second claim (x has one solution; x has no solutions), and then provide written arguments as to why the two claims are both incorrect.

ECD tools could help. Teachers should consider how they can routinely incorporate such deeper learning questions within their ongoing curriculum and instruction. The evidence-centered design process, in fact, may offer some tools to help teachers do so. Consider the products of the test development processes both consortia have used: They have established claims about student performance that their ELA and mathematics tests are intended to evaluate. These represent major competencies that students are expected to develop, and the big ideas of curriculum goals: for example, "Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts; students solve real world problems, engaging particularly in the modeling practice" (see Figure 1 and Figure 2). Just as classroom curriculum lays out a progression of learning targets that students will need to accomplish to reach these broader goals, the consortia have defined the specific assessment targets that constitute their claims and have created items and/or task specifications to measure each one. Furthermore, in the case of Smarter Balanced, they also specified the DOK levels of the specified targets. Intended for item writers, these publicly available models and templates could also be used by teachers to create classroom assessments, particularly to integrate higher levels of complexity into their ongoing instruction and assessment.

Granted, these specifications are currently very complex and not particularly user friendly, from a teacher perspective. Nonetheless, if districts, schools, and/or teachers take the time to digest them, the specifications for DOK3 and DOK4 targets can provide some guidance.

Further, the performance task specifications for both PARCC and Smarter Balanced provide general templates for the design of tasks that address deeper learning levels, particularly for ELA. For example, the task specifications for Smarter Balanced indicate that students will be exposed to at least two stimuli (in each task), consisting of one or more passages from a novel, informational articles, videos, etc. PARCC ELA performance tasks emphasize analysis and synthesis of two texts.

We do not mean to imply that classroom teaching and learning should be reduced to test preparation, but rather that teachers might consider using selected consortia prompts and models in their ongoing classroom assessment and integrate consortia-type assessment in support of their teaching and learning goals. For example, in probing students' reading and use of evidence they should consider including, as part of the classroom repertoire, the types of questions and prompts that are similar to those that will be used in the summative system. Certainly, classroom performance assessment can and should go beyond the bounds of what can be accomplished in a one- or two-day performance task. The general template of the ELA tasks involves having students read multiple sources closely, analyze, and then synthesize and/or compare them in a culminating performance. Similarly, the math tasks have students examine different data representations (e.g., tables, equations, graphs, etc.), analyze an existing model, and then extend on or create a new model or investigation. Both of these provide recipes that can be applied within a unit, as a culminating performance, or an extended research paper. Students could even potentially be given the opportunity to select their own topics and sources. A subsequent report will further explore these ideas.

In conclusion, the study reported here indicates that both the PARCC and Smarter Balanced summative assessment systems, based on their current blueprints, mark a significant step forward in their demands for deeper learning. Students and teachers alike are likely to find the assessments very challenging. However, study results, and the ECD products on which they are based, provide a general roadmap for orienting classroom curriculum, teaching, learning, and assessment toward success.

References

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: University of California, National Center for Research on Evaluation, Standards, & Student Testing.

Hess, K. (2013). *A guide for using Webb's Depth of Knowledge with Common Core State Standards.* Retrieved from http://cliu21cng.wikispaces.com/file/view/WebsDepthofKnowledgeFlipChart.pdf/457670878/WebsDepthofKnowledgeFlipChart.pdf

Mathematics Assessment Resource Service (MARS), Shell Center, University of Nottingham. (2012). *Sorting equations and identities.* Nottingham, UK: University of Nottingham and UC Berkeley. Retrieved from http://map.mathshell.org/materials/download.php?fileid=688

Partnership for Assessment of Readiness for College and Careers (PARCC). (2014). *Grade 8 English Language Arts/Literacy Performance Based Assessment Practice Test*. Retrieved from parcc.pearson.com/practice-tests/english/

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Rentner, D. S. (2013). *Year 3 of implementing the Common Core State Standards: States prepare for Common Core assessments*. Washington, DC: Center on Education Policy. Retrieved from http://www.cep-dc.org/displayDocument.cfm?DocumentID=423

Rentner, D. S., & Kober, N. (2014). *Common Core State Standards in 2014: Districts' perceptions, progress, and challenges*. Washington, DC: Center on Education Policy. Retrieved from http://www.cep-dc.org/displayDocument.cfm?DocumentID=440

Scholastic & the Bill and Melinda Gates Foundation. (2014). *Primary sources: Update: Teacher's views on Common Core State Standards.* New York: Author. Retrieved from http://www.scholastic.com/primarysources/PrimarySources-2014update.pdf

Smarter Balanced Assessment Consortium. (2014). *Practice test scoring guide Grade 8 mathematics. Monterey, CA: CTB McGraw-Hill*. Retrieved from http://sbac.portal.airast.org/wp-content/uploads/2013/08/G8_Practice-Test-Scoring-Guide-8-30-14-Final.pdf

Webb, N. L. (2002a). *An Analysis of the alignment between mathematics standards and assessments for three states*. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, Louisiana.

Webb, N. L. (2002b). *Depth-of-knowledge levels for four content areas*. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin-Madison.

Webb, N. L., Alt, M., Ely, R., & Vesperman, B. (2005). *Web alignment tool (WAT): Training manual 1.1*. Retrieved from http://www.wcer.wisc.edu/WAT/Training%20Manual%202.1%20Draft%20091205.doc

The William and Flora Hewlett Foundation. (n.d.). *What is deeper learning?* Retrieved from http://www.hewlett.org/programs/education/deeper-learning/what-deeper-learning

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 Summative Assessment Grade-Level Results

Table A1

Smarter Balanced ELA Summative Assessment: Means Across Grade Spans, Number, and Percentage of Items by Claim, Total and Minimum at DOK 3/4

	CAT component				Performance task component			
Claim by grade	Mean # items	% Total	Min # DOK3/4	Min % DOK3/4	Mean # items	% Total	Min # DOK3/4	Min % DOK3/4
Elementary: Grades 3-5								
Claim 1: Reading	16	39%	2	13%	0	0%	n/a	n/a
Claim 2: Writing	10	24%	1	10%	3	50%	3	100%
Claim 3: Speaking/listening	9	22%	0	0%	0	0%	n/a	n/a
Claim 4: Research	6	15%	0	0%	3	50%	2	67%
Subtotal	41	100%	3	7%	6	100%	5	83%
Middle: Grades 6-8								
Claim 1: Reading	17	40%	2	12%	0	0%	n/a	n/a
Claim 2: Writing	10	24%	1	10%	3	50%	3	100%
Claim 3: Speaking/listening	9	21%	0	0%	0	0%	n/a	n/a
Claim 4: Research	6	14%	0	0%	3	50%	2	67%
Subtotal	42	100%	3	7%	6	100%	5	83%
High school: Grades 9-12								
Claim 1: Reading	16	39%	3	19%	0	0%	n/a	n/a
Claim 2: Writing	10	24%	1	10%	3	50%	3	100%
Claim 3: Speaking/listening	9	22%	0	0%	0	0%	n/a	n/a
Claim 4: Research	6	15%	0	0%	3	50%	2	67%
Subtotal	41	100%	4	10%	6	100%	5	83%

Table A2

Smarter Balanced Math Summative Assessment: Means Across Grade Spans, Number, and Percentage of Items by Claim, Total, and Minimum at DOK3/4

	CAT component				Performance task component			
Claim by grade	Mean # items	% Total	# DOK3/4	% DOK3/4	Mean # items	% Total	# DOK3/4	% DOK3/4
Elementary: Grades 3-5								
Claim 1: Concepts and procedures	20.0	59%	0	0%	0	0%	n/a	n/a
Claim 2/4: Problem solving & modeling and data analysis	6.0	18%	2	33%	4	67%	2	50%
Claim 3: Communicating reasoning	8.0	24%	2	25%	2	33%	0	0%
Subtotal	34.0	100%	4	12%	6	100%	2	33%
Middle: Grades 6-8								
Claim 1: Concepts and procedures	19.7	58%	0	0%	0	0	n/a	n/a
Claim 2/4: Problem solving & modeling and data analysis	6.0	18%	2	33%	4	67%	2	50%
Claim 3: Communicating reasoning	8.0	24%	2	25%	2	33%	0	0%
Subtotal	33.7	100%	4	12%	6	100%	2	33%
High school: Grades 9-12								
Claim 1: Concepts and procedures	22.0	61%	0	0%	0	0	n/a	n/a
Claim 2/4: Problem solving & modeling and data analysis	6.0	17%	2	33%	4	67%	2	50%
Claim 3: Communicating reasoning	8.0	22%	2	25%	2	33%	0	0%
Subtotal	36.0	100%	4	11%	6	100%	2	33%

Note. Specifications are rounded across grades within each grade span.