

# Appendix G

## Testing Program Responses to Study and Descriptions of Test Changes for 2015–2016

### ACT Aspire

#### Response to Report

One primary purpose of this study was to identify areas for improvement in each of the four evaluated testing programs. The study's findings include insights that promise to advance the industry standards for assessment quality. ACT Aspire is taking this opportunity to listen to the findings and implement adjustments to the assessment.

There are also aspects of the assessment that we have identified as areas for change through our own internal analyses, and we have already made design adjustments that will improve ACT Aspire in 2015–16. Both ACT Aspire's response to the study and the 2015–16 design adjustments will be discussed here.

#### Response to Study Findings about ACT Aspire

ACT Aspire is planning changes to two key elements for which the study found limited alignment with the CCSSO Criteria in English Language Arts:

##### 1 Writing

- ◆ Although the ACT Aspire Writing test was intentionally designed to have writing tasks that do not contain the heavy reading load of “writing to sources” tasks, we are currently exploring updated designs that would supplement the current items with tasks that measure these valuable literacy skills. These tasks would also improve coverage of the “Assessing research and inquiry” criterion in the CCSSO framework.

##### 2 Reading

- ◆ In response to the findings about distribution of Depth of Knowledge (DOK), ACT Aspire has already increased the percentage of upper-level DOK items. This effort will build on changes already in effect for the 2016 assessments (DOK 3 items in grades 5 and 8 will increase from 31 percent in 2015 to 38 percent in 2016).
- ◆ ACT Aspire is adding new technology-enhanced item designs that emphasize selecting evidence directly from the passage to support claims and interpretations. While some of these new TE items will be operational in 2016, ACT Aspire is continuing to explore new ways to assess student use of evidence from texts.

ACT Aspire would also like to make a clarification about terminology used to classify texts on the Reading test:

- ◆ The study findings indicate that while ACT Aspire is a Good Match in Depth, the tests should have “additional literary narrative text, as opposed to literary informational texts.” The study’s ELA/Literacy panel has made a different interpretation than ACT Aspire of CCSSO criterion B.1 that this finding refers to. Differences of interpretation around genre definitions are understandable, but it is important to note the effects on the study outcomes. The CCSSO criteria B.1 refers to texts that are “balanced across literary and informational text types and genres” and does not specify a balance of fiction and nonfiction in the literary category. The study’s panel interprets “literary” text types as only including literary narrative *fiction*. ACT Aspire, however, interprets “literary” to include both literary fiction and literary nonfiction passages that have a narrative structure. In accordance with B.1 (“In all grades, informational texts are primarily expository rather than narrative in structure”), ACT Aspire does not include texts that have a primarily narrative structure in the informational category. Aspire’s interpretation of criterion B.1 results in a stronger match to the specified balance of text types.

In math, ACT Aspire is enacting changes in response to three of the study findings:

- 1 Range of item types
  - ◆ The report distinguished levels of use of multiple-choice items of 50 percent and 75 percent and preferred 50 percent or less. With regard to technology-enhanced items, the report also recommended “using them strategically (read sparingly)” in order to use resources wisely. ACT Aspire will apply its research to make high-quality items of all types and look to expand what is possible in directions that involve technology.
- 2 Content focus
  - ◆ The report recommended “an increase in the number of items focused on the major work of the grade.” We will be working to increase this focus and gathering data to understand the balance in terms of promoting college and career readiness.
- 3 Depth of Knowledge
  - ◆ The study’s review panel recommended “the addition of more items at grade 8 that assess standards at DOK 1.” ACT Aspire currently has a plan in place that will increase the number of DOK 1 questions.

## 2015–16 Test Program Changes

In an effort to continuously improve ACT Aspire, we have already made adjustments in the following three categories for 2015–16:

- 1 Timing Adjustments – Based on customer feedback and in order to allow all students a better opportunity to show what they know and can do, we will be adjusting the time per test by five to ten minutes (Writing will not change). (See Tables G-1 and G-2 for more information on timing and point adjustments by grade and category.)
- 2 Adjustments to English Test – Adding six multiple-choice items for grades 3, 4, and 5.
- 3 Adjustments to Math Test – Adding six multiple choice items for grades 3, 4, and 5; removing one constructed-response (CR) item from grades 3, 4, and 5.

TABLE G-1

## Timing Adjustments

ACT Aspire Summative Testing Time Adjustments (in minutes)								
Grade	English (Current)	English (New)	Math (Current)	Math (New)	Reading (Current)	Reading (New)	Science (Current)	Science (New)
<b>3</b>	30	<b>40</b>	55	<b>65</b>	60	<b>65</b>	55	<b>60</b>
<b>4</b>	30	<b>40</b>	55	<b>65</b>	60	<b>65</b>	55	<b>60</b>
<b>5</b>	30	<b>40</b>	55	<b>65</b>	60	<b>65</b>	55	<b>60</b>
<b>6</b>	35	<b>40</b>	60	<b>70</b>	60	<b>65</b>	55	<b>60</b>
<b>7</b>	35	<b>40</b>	60	<b>70</b>	60	<b>65</b>	55	<b>60</b>
<b>8</b>	35	<b>40</b>	65	<b>75</b>	60	<b>65</b>	55	<b>60</b>
<b>*EHS</b>	40	<b>45</b>	65	<b>75</b>	60	<b>65</b>	55	<b>60</b>

\*Early High School

TABLE G-2

## English and Mathematics: Number of Points by Reporting Category

Reporting Category	GRADE									
	3 (new)	3 (old)	4 (new)	4 (old)	5 (new)	5 (old)	6	7	8	EHS
	# of Points									
<b>English</b>										
Production of Writing	12-14	9-11	8-10	6-8	8-10	6-8	11-13	9-11	9-11	12-14
Knowledge of Language			3-5	2-4	3-5	2-4	2-4	4-6	4-6	6-8
Conventions of Standard English	17-19	14-16	17-19	14-16	17-19	14-16	19-21	19-21	19-21	29-31
<b>Total for English</b>	<b>31</b>	<b>25</b>	<b>31</b>	<b>25</b>	<b>31</b>	<b>25</b>	<b>35</b>	<b>35</b>	<b>35</b>	<b>50</b>
<b>Mathematics</b>										
Number & Operations in Base 10		5-7	5-8	3-5	5-8	3-5	1-3	1-3	1-3	0-2
Number & Operations - Fractions	3-5	2-4	6-8	4-6	6-8	4-6	1-3	1-3	1-3	0-2
The Number System							3-5	3-5	2-4	1-3
Number & Quantity										1-3
Operations & Algebraic Thinking	6-8	3-5	4-6	3-5		3-5	1-3	1-3	0-2	0-2
Expressions & Equations							3-5	3-5	5-7	2-4
Ratios & Proportional Reasoning							3-5	3-5	0-2	1-3
Algebra										2-4
Functions									3-5	3-5
Measurement & Data (measurement)							0-2	0-2	1-3	1-3
Geometry		3-5		3-5	4-6	3-5	5-7	4-6	6-8	5-7
Measurement & Data	5-7	3-5		3-5		3-5				
Measurement & Data (data)							0-2	1-3	1-3	1-3
Statistics & Probability							3-5	3-5	4-6	4-7
Justification & Explanation	12	16	12	16	12	16	16	16	20	20
<b>Total for Mathematics</b>	<b>39</b>	<b>37</b>	<b>39</b>	<b>37</b>	<b>39</b>	<b>37</b>	<b>46</b>	<b>46</b>	<b>53</b>	<b>53</b>

## MCAS

### Response to Report

Our goal as a Commonwealth is to ensure that every Massachusetts student is prepared to succeed in postsecondary education and compete in the global economy. We have been administering annual assessments in Massachusetts since 1998 as our way of holding ourselves accountable for our progress toward this goal. The Massachusetts Comprehensive Assessment System (MCAS) tests are generally considered the gold standard of state assessments. They hold students to high expectations—in most cases, equivalent to the proficiency standard on the National Assessment of Educational Progress (NAEP)—and use a variety of question formats to ensure that we assess the full range of student abilities. Over the years we have refined the assessments to adapt to changes in the curriculum frameworks, most notably the incorporation of the Common Core State Standards into our 2010 frameworks, and to improve the quality of the assessment over time.

Our students and educators have accomplished incredible things under this system. Massachusetts' NAEP scores have moved from middle of the pack to leading the nation, and our students have scored well on international assessments. We have also made substantial progress toward closing the proficiency gaps between student subgroups, and we have dramatically reduced our dropout rate and increased our cohort graduation rate. That success would not have been possible without a high-quality assessment providing feedback on student, school, district, and state achievement and progress.

The Massachusetts Comprehensive Assessment Systems was a terrific twentieth-century assessment—but it has reached a point of diminishing returns. In 2015, MCAS was administered for the eighteenth year. We have a better understanding now than we did a decade or two ago about learning progression in mathematics, text complexity and the interplay of reading and writing, and the academic expectations of higher education and employers. And we now know that nearly one-third of our public high school students who go on to enroll in Massachusetts public colleges take at least one remedial course in their first semester, suggesting that the curriculum and assessments they have experienced have not adequately prepared them for the world beyond high school. Indeed, MCAS was never designed to be an indicator of college and career readiness. We joined the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium specifically in order to partner with other states in developing an assessment that is more closely aligned to these expectations.

Thus, we were not surprised by this report's conclusion that the MCAS does not always measure well what's most important today. This report also confirms that in many ways, PARCC sets a higher bar than MCAS for student performance. This is particularly true as students move up the grades into middle and high school. This higher bar is not simply about being harder: PARCC provides more opportunities for critical thinking, applying knowledge, research, and making connections between reading and writing. More and more schools have upgraded curriculum and instruction to align with our 2010 frameworks. While we adjusted MCAS to test those frameworks, PARCC was built around them. Classroom instruction is now increasingly focused on the knowledge and skills in the frameworks, rather than how to pass a test.

We are proud of what we have accomplished in Massachusetts in the nearly two decades that we have been administering the MCAS. Now that we have the benefit of that experience and have revised our curriculum frameworks to reflect our upgraded learning expectations, it is time to upgrade our assessments too. Our state Board of Elementary and Secondary Education voted in November 2015 to do exactly that.

### 2015–16 Test Program Changes

Over the next few years, we will transition to a new statewide assessment system that will take much of what this report identifies as the strengths of PARCC—high-quality content aligned strongly to college and career ready standards—and combine it with elements of MCAS in the context of a Massachusetts-specific governance system that will allow us to set our own policies on test content, administration, and reporting. With this approach, we will continue to benefit from a high-quality, next-generation assessment while ensuring that the test will reflect the Commonwealth's unique needs and concerns. Most importantly, our students will be better prepared for success after high school—our ultimate goal.

## PARCC

### Response to Report

PARCC would like to thank the study authors and review panelists for a comprehensive, strong study. There are two areas where we would like to present a few additional comments.

### ELA/Literacy Content Rating

#### Panel Recommendation:

*The tests could be strengthened by the addition of research tasks that require students to use two or more sources and, as technologies allow, a listening and speaking component.*

#### PARCC Response:

Every PARCC assessment in grades 3–8 requires students to complete a research simulation task where the student reads two or three sources and must integrate or synthesize the ideas in a written essay. All students also read two literary texts and write a literary analysis (literary research) essay.

The PARCC assessment measures many aspects that are key to the Speaking and Listening standards. PARCC uses multimedia texts to measure comprehension for all students taking its tests online (providing students with opportunities to demonstrate strengths and needs in comprehending audio and audiovisual texts). The CCSS build coherence across the ELA strands and identify similar skills built into both the reading comprehension standards (standards RI.7 and RL.7) and the listening standards. PARCC chose to report students' speaking and listening performance in relation to the reading standards.

The PARCC assessment system includes a robust set of Speaking and Listening tools. All schools administering PARCC in 2015–2016 have access to a comprehensive set of formative assessments and instructional tools to support educators, parents, and students in better understanding students' strengths and needs in speaking and listening. Further information about the PARCC Speaking and Listening tools can be found on PARCC's Partnership Resource Center: <https://prc.parcconline.org/library/speaking-and-listening-overview>.

### Cognitive Demand

#### Panel Recommendation:

*The program could better meet the Depth criteria by ensuring that the distribution of cognitive demand on the assessments provides sufficient information across the range.*

#### PARCC Response:

It is important to note that students who meet Level 1/Level 2 Depth of Knowledge (DOK) for items situated at higher DOK levels are given partial credit points for demonstrating skills that require lower cognitive complexity. Reviewers did not consider the possibility that scoring, rather than adding more Level 1 or Level 2 items, could allow for the balance of item complexities. For more information on the PARCC scoring rubrics and to view released items, visit: <https://prc.parcconline.org/assessments/parcc-released-items>.

The PARCC assessment uses a cognitive complexity framework that was developed by the PARCC consortium to more accurately reflect the demands of the CCSS. This framework received recognition from AERA (2014 Outstanding Contribution to Practice in Cognition and Assessment award). An article detailing the innovations of this framework and potential next steps in research around cognitive complexity has been published in a new book titled *The Next Generation of Testing: Common Core Standards, Smarter-Balanced*.<sup>106</sup>

<sup>106</sup> H. Jiao and R. Lissitz, eds., *The Next Generation of Testing: Common Core Standards, Smarter-Balanced, PARCC, and the Nationwide Testing Movement* (Charlotte, NC: Information Age Publishing, Inc., 2016).

## 2015–16 Test Program Changes

In May 2015, the chief state school officers from the PARCC states unanimously voted to streamline the assessment. They accomplished this goal while retaining all the key elements of the test—a strong commitment to quality and reliability, measurement of the full range of the standards, and the ability to get results back to teachers and parents quickly, so that they can help meet the needs of students for the coming school year. The following changes to the test design will be instituted in the 2015–16 school year:

- ◆ The two testing windows (the performance-based and end-of-year components) in mathematics and English language arts/literacy (which includes reading and writing) will be consolidated into one. The single testing window will simplify administration of the test for states and schools. Schools will have up to thirty school days to administer the test, and the testing window will extend from roughly the 75 percent mark to the 90 percent mark of instructional time. Most schools will complete testing in one to two weeks during that window.

The testing time for students will be reduced by about ninety minutes overall (sixty minutes in mathematics; thirty minutes in English language arts/literacy). The result will be that the total testing time for ELA/Literacy and mathematics will be approximately 8.5 hours at grades 3–5, 9.2 hours at grades 6–8, and 9.7 hours at grade 11. There will also be more uniformity of test unit times, allowing for easier scheduling in schools.

- ◆ Each PARCC assessment is administered in multiple sections, called units. The number of test units was reduced for all students, and includes three units in English language arts/literacy and three or four units in mathematics.

The testing time was shortened by reducing the number of score points and items in both subject areas. The tables below show a comparison of score points between the previous test design and the redesign.

**TABLE G-3**

### Comparison of Score Points in the Previous ELA/Literacy Design and the Redesign

		Previous Two Administrations	Adopted Single Administration
<b>Grade 3</b>	Reading Points	64	58
	Writing Points	36	36
	Total Points	100	94
	Units	4	3
	Total Testing Time	4.75 hours	4.25 hours*
<b>Grades 4–5</b>	Reading Points	70	62
	Writing Points	36	36
	Total Points	106	98
	Units	4	3
	Total Testing Time	5.0 hours	4.5 hours**
<b>Grades 6–11</b>	Reading Points	94	76
	Writing Points	45	45
	Total Points	139	121
	Units	5	3
	Total Testing Time	5.75 hours	5.2 hours***

\* Add 1.5 hours for field test unit

\*\* Add 1.5 hours for field test unit

\*\*\*Add 1.8 hours for field test unit

TABLE G-4

## Comparison of Score Points in the Previous Mathematics Design and the Redesign

		Previous Two Windows	Adopted Single Administration
<b>Grade 3-8</b>	Short Items	56 pts	40 pts
	Reasoning Items	14 pts	14 pts
	Modeling Items	12 pts	12 pts
	Total Points	82 pts	66 pts
	Units	4 @ varies	4 @ 60 min.
	Total Time on Task	5 hours	4 hours
<b>Algebra I, Geometry, Algebra II, and Integrated Math I, II, III</b>	Short Items	65 pts	49 pts
	Reasoning Items	14 pts	14 pts
	Modeling Items	18 pts	18 pts
	Total Points	97 pts	81 pts
	Units	4 @ varies	3 @ 90 min.
	Total Time on Task	5.3 - 5.5 hours	4.5 hours

- ◆ Standalone field testing will be eliminated. As with all similar testing, field test items—items that could be used in future years—are embedded in each student’s test. Because the performance tasks in English language arts/literacy are longer, a sampling of students had to take a standalone field test unit for these tasks in spring 2015. To further streamline the testing process for all schools, the PARCC field test will now be wrapped into the testing window. Each year, a small percentage of students will take an additional ELA/Literacy unit. Schools and classrooms selected in one year—per the process determined in their state—will in almost all cases not have to field test again for several years.
- ◆ The test design changes do not result in the loss of any performance tasks in English language arts/literacy (there are still three performance tasks). Additionally, there are now two or three text sets included in the units, depending on the grade level (one text set was removed for grades 6–11).

The test design changes do not result in the loss of any reasoning and modeling mathematics items, with the exception of Algebra II and Integrated Math III at the high school level. Short answer items were removed.

For more information, visit <http://parconline.org/assessments/test-design/design-changes>.

## Smarter Balanced

### Response to Report

Tony Alpert, Executive Director  
Luci Willits, Deputy Executive Director

December 9, 2015

Thank you to the Thomas B. Fordham Institute and HumRRO for its diligent work to evaluate the quality of the Smarter Balanced summative assessment and its alignment to the Common Core State Standards.

While this report focused on the end-of-year test, Smarter Balanced is more than a summative assessment: it's a system to improve teaching and learning. Our system includes optional and flexible interim assessments available throughout the year to help teachers monitor student progress, as well as a Digital Library with thousands of educator-approved classroom resources. Nearly 5,000 educators from across the country helped build the Smarter Balanced system. Smarter Balanced assessments are designed to be administered online and are customized for every child using built-in accessibility resources.

This report recognizes many of these strengths and gives Smarter Balanced an Excellent or Good Match in all but one subcategory. In addition, the report recognizes that Smarter Balanced is the only assessment that measures students' listening skills. We are proud of these ratings. We also recognize that there is always room for improvement. However, one of the greatest strengths of Smarter Balanced, the computer-adaptive feature of the summative assessment, is not addressed in this report. Because it is an adaptive test that is customized for each student, it is difficult to compare the Smarter Balanced summative assessment on an item-per-item basis to a fixed form test that is static. In addition, this study did not consider some other important features of the Smarter Balanced assessment, including the ability of states to work with the service provider of their choice. Finally, Smarter Balanced is arguably the most accessible large-scale assessment system that includes supports for over 90 percent of the consortium's English language learners' primary languages. Individually, these elements are historic; collectively, they are unprecedented.

It is important to note that due to the timing of this study, reviewers were not able to access all of the interactive features that are available to all students during a live test. For instance, reviewers did not interact with features such as highlighting text in passages and test questions, zooming in and out of test pages, making notes about a test question in the notepad, and using strikethrough for answer options. In addition, the study's version of the system did not provide some of the helpful built-in student tools, such as error messages when students use incorrect keys, the ability to mark items, or move forward in the test without answering all the questions on a page.



## 2015–16 Test Program Changes

For the 2015–2016 summative tests, Smarter Balanced members have the flexibility to determine whether classroom activities will be given prior to the administration of performance tasks.

The following table describes the differences between the system used in the study versus the one actually used for students:

**TABLE G-5**

### Differences in Student and Study Interface

<b>Students' Actual Version:</b>	<b>Study Version:</b>
Calculators available in grades 6 through high school for items when not measuring computation.	No calculators.
Tutorials to show students how to use the available tools and to interact with all of the different types of items they might see on a test.	No tutorials.
Grade appropriate and item-specific English glossaries are available for mathematics and English language arts.	No glossaries.
For mathematics, grade appropriate item-specific translated glossaries are available in ten different languages, plus dialects.	No glossaries.
Error messages given to students when they try to enter characters that aren't allowed.	No error messages when equation editor and fill-in blank items are incorrectly populated.
Verdana Font (14 pt)	Times New Roman font (12 pt)
Formatted for best results according to student cognitive labs and field testing.	Format not consistent.
All the research-based tools available as appropriate to the content area and item (as shown in the practice tests).	Limited availability of tools. For example, notepad, underline, highlight, etc. were not available.

As part of the development process, Smarter Balanced collaborated with national experts and local teachers to determine how to best measure critical thinking and problem solving skills as part of college and career ready standards. For example, at times, it is most appropriate to ask students to solve engaging items within a real-world scenario; while at other times, presenting students with an equation to solve is a better way to measure student knowledge. This is reflected in the test blueprint.

It is important to note that with the adaptive test, Smarter Balanced can measure more complex skills for low- and high-performing students alike. In mathematics, Smarter Balanced chose to emphasize the more complex skill sets with the understanding that students must have the procedural knowledge to do well on the test. With English language arts, we will discuss the report's findings with our membership and consider changes.

Smarter Balanced is committed to including only high-quality questions on our tests. We were disappointed to see that reviewers found a handful of questions that needed improvement and received a rating of Limited/Uneven Match. Smarter Balanced has an extensive process for question development to ensure each test item

is extensively reviewed prior to being included on a student's test. Educators including national mathematical, English language arts, and accessibility experts write questions and review them for content accuracy as well as for any potential bias or lack of sensitivity. Questions that do not meet a very high standard are revised or are removed. However, we will use this study to improve our item development review processes. Immediately, Smarter Balanced will initiate a detailed review of the existing test questions based on the feedback from this report.

In addition to this positive review of Smarter Balanced, we were pleased to note that the National Network of State Teachers of the Year echoed complimentary feedback in their report as well. That review looked at many of the same questions as this review. The nation's best teachers said Smarter Balanced provides a better picture of student performance, is grade-level appropriate, and supports great teaching and learning throughout the year.

Thank you again for your review and for the opportunity to provide more context into the reviewers' findings.

Sincerely,

A handwritten signature in black ink that reads "Anthony Alpert". The signature is written in a cursive style with a large initial "A".

Anthony Alpert  
Executive Director  
Smarter Balanced Assessment Consortium