

Evaluating the Content and Quality of Next Generation Assessments



Nancy Doorey
Morgan Polikoff

February 11, 2016

Study Overview

This study evaluates the content and quality of assessments for grades 5 and 8 (“capstone” grades for elementary and middle school) for both mathematics and English language arts (ELA/Literacy)

Aims to inform educators, parents, policymakers and other state and local officials of the strengths and weaknesses of several new next-generation assessments on the market (**ACT Aspire**, **PARCC**, **Smarter Balanced**)—as well as how a respected state test (**MCAS**) stacks up

Evaluation criteria drawn from the content-specific portions of the Council of Chief State School Officers’ (CCSSO’s) “Criteria for Procuring and Evaluating High Quality Assessments”

Study Components

Phase 1

- Item Review: Test Forms
- Generalizability (Document) Review: Blueprints, assessment frameworks, etc. (subset of item reviewers)

Phase 2

- Aggregation of Item Review and Generalizability Results and development of consensus statements

Review Panels and Design

- We received **over 200 reviewer recommendations** from various assessment and content experts and organizations, as well as each of the four participating assessment programs.
- In vetting applicants, we prioritized extensive content and/or assessment expertise, deep familiarity with the CCSS, and prior experience with alignment studies. Not eligible: employees of test programs or writers of the standards
- Final review panels were composed of classroom educators, content experts, and experts in assessment. We included at least one reviewer recommended by each participating program on each panel.
- **Seven test forms were reviewed per grade level and content area** (2 forms each for Smarter Balanced, PARCC, and ACT Aspire, and 1 form for MCAS). Reviewers were randomly assigned to forms using a “jigsaw” approach across testing programs to minimize major differences across panels and enhance inter-rater reliability.

Council of Chief State School Officers (CCSSO) Criteria Evaluated

B. Align to Standards – English Language Arts/Literacy

- B.1 Assessing student reading and writing achievement in both ELA and literacy
- B.2 Focusing on complexity of texts
- B.3 Requiring students to read closely and use evidence from texts
- B.4 Requiring a range of cognitive demand
- B.5 Assessing writing
- B.6 Emphasizing vocabulary and language skills
- B.7 Assessing research and inquiry
- B.8 Assessing speaking and listening
(measured but not counted)
- B.9 Ensuring high-quality items and a variety of item types

C. Align to Standards – Mathematics

- C.1 Focusing strongly on the content most needed for success in later mathematics
- C.2 Assessing a balance of concepts, procedures, and applications
- C.3 Connecting practice to content
- C.4 Requiring a range of cognitive demand
- C.5 Ensuring high-quality items and a variety of item types

Content criteria: Orange

Depth criteria: Blue

Key Study Questions

1. Do the assessments place strong emphasis on the most important content for college and career readiness (CCR) as called for by the Common Core State Standards and other CCR standards? **(Content)**
2. Do they require all students to demonstrate the range of thinking skills, including higher-order skills, called for by those standards? **(Depth)**
3. What are the overall strengths and weaknesses of each assessment relative to the examined criteria for ELA/Literacy and mathematics? **(Overall Strengths and Weaknesses)**

Overall Findings





















- Each panel reviewed the ratings from the grade 5 and grade 8 test forms, considered the results of the documentation review, and came to consensus on the criterion's rating--assigning the programs a rating on each of the ELA/Literacy and mathematics criterion:
 - **Excellent Match**
 - **Good Match**
 - **Limited/Uneven Match**
 - **Weak Match.**
- The PARCC and Smarter Balanced assessments earned an Excellent or Good Match to the CCSSO Criteria for both ELA/Literacy and mathematics.
- While ACT Aspire and MCAS did well regarding the quality of items and the depth of knowledge assessed (Depth), the panelists found that these two programs do not adequately assess—or may not assess at all—some of the priority content in both ELA/Literacy and mathematics at one or both grades in the study (Content). This may reflect that ACT Aspire and MCAS were not developed with CCSS explicitly in mind.

ELA/Literacy Content Ratings Summary

Criteria	ACT Aspire	MCAS	PARCC	Smarter Balanced
I. CONTENT: Assess the content most needed for College and Career Readiness	L	L	E	E
B.3 Reading: * Tests require students to read closely and use specific evidence from texts to obtain and defend correct responses.	L	G	E	E
B.5 Writing: * Tasks require students to engage in close reading and analysis of texts. Across each grade band, tests include a balance of expository, persuasive/argument, and narrative writing.	L	W	E	E
B.6 Vocabulary and language skills: Tests place sufficient emphasis on academic vocabulary and language conventions as used in real-world activities.	G	L	E	G
B.7 Research and Inquiry: Assessments require students to demonstrate the ability to find, process, synthesize, and organize information from multiple sources.	L	W	E	E
B.8 Speaking and listening: Over time, and as assessment advances allow, the assessments measure speaking and listening communication skills.**	W	W	W	L

LEGEND  Excellent Match  Good Match  Limited/Uneven Match  Weak Match

ELA/Literacy Depth Ratings Summary

Criteria	ACT Aspire	MCAS	PARCC	Smarter Balanced
II. DEPTH: Assesses the depth that reflects the demands of College and Career Readiness				
B.1 Text quality and types: Tests include an aligned balance of high-quality literary and informational texts.				
B.2 Complexity of texts: Test passages are at appropriate levels of text complexity, increasing through the grades, and multiple forms of authentic, high-quality texts are used.***				
B.4 Cognitive demand: The distribution of cognitive demand for each grade level is sufficient to assess the depth and complexity of the standards.				
B.9 High-quality items and variety of item types: Items are of high technical and editorial quality and test forms include at least two item types with at least one that requires students to generate a response.				

LEGEND



Excellent Match



Good Match










Limited/Uneven Match



Weak Match

Criterion B.4 Findings: The Distribution of Cognitive Demand in ELA/Literacy




		Grade 5			Grade 8		
		LEVEL 1	LEVEL 2	LEVELS 3 & 4	LEVEL 1	LEVEL 2	LEVELS 3 & 4
CCSS		18%	37%	46%	10%	44%	46%
ACT Aspire		34%	38%	28%	46%	36%	18%
MCAS		10%	63%	27%	5%	59%	37%
PARCC		5%	45%	50%	2%	29%	69%
Smarter Balanced		19%	59%	22%	15%	41%	44%

LEGEND  Excellent Match  Good Match  Limited/Uneven Match  Weak Match

















Note: Percentages in the table represent percentages of score points at each DOK level. Results for a particular grade and program were generated by averaging across all raters and forms for that grade and program (e.g., averaging the four raters of ACT form 1 and the four raters of ACT form 2 at grade 5).




Mathematics Content Ratings Summary

Criteria	ACT Aspire	MCAS	PARCC	Smarter Balanced
I. CONTENT: Assesses the content most needed for College and Career Readiness	L	L	G	G
C.1 Focus: Tests focus strongly on the content most needed in each grade or course for success in later mathematics (i.e., major work).	W	L	G	G
C.2: Concepts, procedures, and applications: Assessments place balanced emphasis on the measurement of conceptual understanding, fluency and procedural skill, and the application of mathematics.****	—	—	—	—

LEGEND  Excellent Match  Good Match  Limited/Uneven Match  Weak Match

Mathematics Depth Ratings Summary

Criteria	ACT Aspire	MCAS	PARCC	Smarter Balanced
II. DEPTH: Assesses the depth that reflects the demands of College and Career Readiness				
C.3 Connecting practice to content: Test questions meaningfully connect mathematical practices and processes with mathematical content.				
C.4 Cognitive demand: The distribution of cognitive demand for each grade level is sufficient to assess the depth and complexity of the standards.				
C.5 High-quality items and variety of item types: Items are of high technical and editorial quality and test forms include at least two item types, at least one that requires students to generate a response.				

LEGEND  Excellent Match  Good Match  Limited/Uneven Match  Weak Match

Criterion C.4 Findings: The Distribution of Cognitive Demand in Mathematics

		Grade 5			Grade 8		
		LEVEL 1	LEVEL 2	LEVELS 3 & 4	LEVEL 1	LEVEL 2	LEVELS 3 & 4
CCSS		43%	50%	7%	51%	40%	9%
ACT Aspire	L	23%	40%	37%	20%	45%	35%
MCAS	E	40%	58%	2%	40%	46%	14%
PARCC	G	34%	55%	11%	13%	62%	24%
Smarter Balanced	G	46%	36%	18%	16%	75%	9%

LEGEND  Excellent Match  Good Match  Limited/Uneven Match  Weak Match

Note: Percentages in the table represent percentages of score points at each DOK level. Results for a particular grade and program were created by averaging across all raters and forms for that grade and program (e.g., averaging the four raters of ACT form 1 and the four raters of ACT form 2 at grade 5).

Recommendations for State Policymakers

1. Make quality non-negotiable.
2. When developing or revising assessments, carefully prioritize the set of skills and knowledge at each grade that should serve as the focus of instruction, building public understanding and support as you do so.
3. Ensure quality is maintained while addressing concerns about testing time and costs.
4. Work with other state leaders to press the assessment industry and researchers for improvements in test item types and scoring engines to better measure key constructs in a cost-effective way.

Recommendations for Test Developers

1. Ensure that every item meets the highest standards for editorial accuracy and technical quality.
2. Use technology-enhanced items (TEI) strategically to improve test quality and enhance student effort.
3. Focus research and development on areas of targeted importance relative to measuring student performance on CCR standards.

Thank you for your time.
Questions?

polikoff@usc.edu



ELA/Literacy Content Results by Criterion

Criterion B.3 -

“Do the tests require students to read closely and use evidence from texts to obtain and defend responses?”



ACT Aspire

LIMITED/UNEVEN MATCH

Although most reading items require close reading of some kind, too many can be answered without analysis of what was read. Items that purport to require specific evidence from text often require only recall of information from text. To meet this criterion, the test items should require students to cite specific text information in support of some conclusion, generalization, or inference drawn from the text.



2014 MCAS

GOOD MATCH

Most reading items require close reading and focus on central ideas and important particulars. Some questions, however, do not require the students to provide direct textual evidence to support their responses. In addition, too many items do not align closely to the specifics of the standards.



PARCC

EXCELLENT MATCH

Nearly all reading items require close reading, the understanding of central ideas and the use of direct textual evidence.



Smarter Balanced

EXCELLENT MATCH

Nearly all reading items align to the reading standards requiring close reading, the understanding of central ideas, and use of direct textual evidence in support of a conclusion, generalization, or inference.

The following were required to fully meet this criterion:

1. Nearly all reading items require close reading and analysis of text, rather than skimming, recall, or simple recognition of paraphrased text.
2. Nearly all reading items focus on central ideas and important particulars.
3. Nearly all items are aligned to the specifics of the standards.
4. More than half of the reading score points are based on items that require direct use of textual evidence.

ELA/Literacy Content Results

Criterion B.5 -

”Do the tests require students to write narrative, expository, and persuasive/argumentation essays (across each grade band, if not in each grade) in which they use evidence from sources to support their claims?”



ACT Aspire

LIMITED/UNEVEN MATCH

Although the program documentation shows that a balance of all three writing types is required across each grade band, the writing prompts do not require writing to sources. As a result, the program insufficiently assesses the types of writing required by college and career readiness standards.



2014 MCAS

WEAK MATCH

Writing is assessed at only one grade level per band, and there is insufficient opportunity to assess writing of multiple types. In addition, the writing assessments do not require students to use sources. As a result, the program inadequately assesses the types of writing required by college and career readiness standards.



PARCC

EXCELLENT MATCH

The assessment meets the writing criterion, which requires writing to sources. Program documentation shows that a balance of all three writing types is required across each grade band.



Smarter Balanced

EXCELLENT MATCH

The writing items are of high quality, and the writing prompts all require the use of textual evidence. Program documentation shows that a balance of all three writing types is required across each grade band.

The following were required to fully meet this criterion:

1. All three writing types are approximately equally represented across all forms in the grade band (K–5; 6–12), allowing blended types (i.e., writing types that blend two or more of narrative, expository, and persuasive/argumentation) to contribute to the distribution.
2. All writing prompts require writing to sources (meaning they are text-based).

ELA/Literacy Content Results

Criterion B.6 -

“Do the tests require students to demonstrate proficiency in the use of language, including academic vocabulary and language conventions, through tasks that mirror real-world activities?”



ACT Aspire

GOOD MATCH

Language items meet the criterion for being tested within writing activities, though more items are needed that are embedded in real world tasks such as editing. The vocabulary items do not meet the criterion because there are too few of them and not enough assess Tier 2 words.



2014 MCAS

LIMITED/UNEVEN MATCH

Vocabulary items are sufficient and generally aligned to the criterion; however, the grade 5 items need more words at the Tier 2 level. Furthermore, a lack of program documentation means that the quality of vocabulary assessments cannot be substantiated across forms. MCAS does not meet the criterion for assessing language skills, which call for them to be assessed within writing assessments that mirror real-world activities including editing and revision.



PARCC

EXCELLENT MATCH

The test contains an adequate number of high-quality items for both language use and Tier 2 vocabulary and awards sufficient score points, according to the program's documentation, to both of these areas.



Smarter Balanced

GOOD MATCH

Language skill items are contained in a sub-score and meet the criterion for being assessed within writing and mirroring real-world activities such as editing and revision. The number of items that test vocabulary is a bit low; further, items coded as vocabulary too often did not test Tier 2 vocabulary words.

The following were required to fully meet this criterion:

1. The large majority of vocabulary items (i.e., three-quarters or more) focus on Tier 2 words and require the use of context, and more than half assess words important to central ideas.
2. A large majority (i.e., three-quarters or more) of the items in the language skills component and/or scored with a writing rubric (i.e., points in writing tasks that are allocated toward a language sub-score), mirror real-world activities, focus on common errors, and emphasize the conventions most important for readiness.
3. Vocabulary is reported as a sub-score or at least 13 percent of score points are devoted to assessing vocabulary/language.
4. Language skills are reported as a sub-score or at least 13 percent of score points are devoted to assessing language skills (language skills items plus score points).

ELA/Literacy Content Results

Criterion B.7 -

“Do the tests require students to demonstrate research skills, including the ability to analyze, synthesize organize, and use information from sources?”



ACT Aspire

LIMITED/UNEVEN MATCH

Although the one item at each grade level involving research and inquiry did indeed require analysis and organization of information, this single item is insufficient to provide a quality measure of research and inquiry.



2014 MCAS

WEAK MATCH

The assessment has no test questions devoted to research.



PARCC

EXCELLENT MATCH

The research items require analysis, synthesis and/or organization and the use of multiple sources, therefore meeting the criterion for Excellent.



Smarter Balanced

EXCELLENT MATCH

The research items require analysis, synthesis and/or organization, and the use of multiple sources.

The following were required to fully meet this criterion:

Three-quarters or more of the research items on each test form require analysis, synthesis, and/or organization of information.

ELA/Literacy Depth Results

Criterion B.1 -

“Do the tests require a balance of high-quality literary and informational texts?”

G **ACT Aspire**

GOOD MATCH

The texts are of high quality, and the proportion of informational texts meets the criterion.

The assessment would better align to the criterion, however, with additional literary narrative text, as opposed to literary informational text.

G **2014 MCAS**

GOOD MATCH

The quality of the texts is very high.

Regarding the balance of text types, some forms had too few informational texts.

G **PARCC**

GOOD MATCH

Although the passages are consistently of high quality, the tests would have better reflected the criterion with additional literary nonfiction passages.

E **Smarter Balanced**

EXCELLENT MATCH

Overall text quality is high and among informational texts there is a high proportion of expository text types.

The following were required to fully meet this criterion:

1. Approximately half of the texts at grades 3–8 and two-thirds at high school are informational, and the remainder literary.
2. Nearly all passages are high quality (previously published or of publishable quality).
3. Nearly all informational passages are expository in structure.
4. For grades 6–12, the informational texts are split nearly evenly for literary nonfiction, history/social science, and science/technical.

ELA/Literacy Depth Results

Criterion B.2 -

“Do the tests require appropriate levels of text complexity, increasing the level each year so that students are ready for the demands of college and career by the end of high school?”



ACT Aspire

GOOD MATCH

It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.



2014 MCAS

GOOD MATCH

It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.



PARCC

GOOD MATCH

It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.



Smarter Balanced

GOOD MATCH

It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.

The following were required to fully meet this criterion:

1. Documentation clearly explains how quantitative data are used to determine grade band placement.
2. Texts are placed at the grade level recommended by the qualitative review.

ELA/Literacy Depth Results

Criterion B.4 -

”Are all students required to demonstrate a range of high order, analytical thinking skills in reading and writing based on the depth and complexity of the standards?”



ACT Aspire

WEAK MATCH

To better reflect the depth and complexity of the standards, both grade-level tests should require more items with higher cognitive demands, although this problem is greater at grade 8.



2014 MCAS

LIMITED/UNEVEN MATCH

More items that measure the higher levels of cognitive demand are needed to sufficiently assess the depth and complexity of the standards.



PARCC

EXCELLENT MATCH

The test is challenging overall; indeed the cognitive demand of the grade 8 test exceeds that of the CCSS.



Smarter Balanced

GOOD MATCH

The cognitive demand of items cover a sufficient range and, in grade 8, the percentage of more demanding items (DOK 3 and 4) correspond well to the demand of the standards. However, the grade 5 test needs more items at higher levels of cognitive demand to reflect fully the depth and complexity of the standards.

The following were required to fully meet this criterion:

To receive the highest rating on this criterion, the distribution of cognitive demand on test forms had to match the distribution of cognitive demand of the standards as a whole and match the higher cognitive demand (DOK 3+) of the standards. Note that criterion B.4 is not a rating of test difficulty. Assessments that do not match the distribution of complexity of the standards, including if they have too many high Depth of Knowledge items, may receive a rating of less than Excellent Match.

ELA/Literacy Depth Results

Criterion B.9 -

”Are a variety of item types used, including at least one that requires students to generate, rather than select, a response, and are the test items of high quality?”

E

ACT Aspire

EXCELLENT MATCH

The test includes items that exhibit high technical quality and editorial accuracy. Multiple item formats are used, including student-constructed responses.

E

2014 MCAS

EXCELLENT MATCH

Multiple item formats are used, including student-generated response items. The items exhibit high technical quality and editorial accuracy. The paper-and-pencil format precludes the use of technology-enhanced items, but the criterion for multiple item types is met.

E

PARCC

EXCELLENT MATCH

The tests use multiple item formats, including student-constructed responses.

G

Smarter Balanced

GOOD MATCH

The tests use multiple formats and technology enhanced items including constructed responses. However, editorial or technical issues, including readability, were noted in a number of items.

The following were required to fully meet this criterion:

1. At least two item formats are used, including one that requires students to generate, rather than select, a response.
2. All or nearly all operational items reviewed reflect both high technical quality and high editorial accuracy.

Now for Mathematics...

Content Results by Criterion

Criterion C.1 -

"Do the tests focus strongly on the content most needed for success in later mathematics?"



ACT Aspire

WEAK MATCH

ACT forms do not consistently place sufficient emphasis on the major work of the given grade, due in part to intentional test design, which requires inclusion of selected content from earlier and later grades. Still, many of the items coded to standards from lower grades do not address the major work of the relevant grade.



2014 MCAS

LIMITED/UNEVEN MATCH

The grade 8 assessment is focused on the major work of the grade. The grade 5 assessment is significantly less focused on the major work of the grade than called for by the criterion, as it samples content across the full set of standards for the grade.



PARCC

GOOD MATCH

While the grade 8 tests focus strongly on the major work of the grade, the grade 5 tests fall short of the threshold for required for the top rating.



Smarter Balanced

GOOD MATCH

While the grade 8 tests focus strongly on the major work of the grade, the grade 5 tests fall short of the threshold for required for the top rating.

The following were required to fully meet this criterion:

The vast majority (i.e., at least three-quarters at elementary grades, at least two-thirds in middle school grades, and at least half in high school) of score points in each assessment focuses on the content that is most important for students to master in that grade in order to reach college and career readiness (also called the major work of the grade), and at least 90 percent of the major work clusters must be assessed by at least one item.

Mathematics Content Results

Criterion C.2-

“Do the tests assess a balance of concepts, skills, and applications?”

Qualitative statements rather than the ratings awarded to this criterion (not used in the determination of the overall Content rating)

In general, the test forms from all four programs showed attention to conceptual understanding, procedural skill, and application.

However, each program fell short of the goal of balance (which was operationalized as an even distribution) in one way or another. For ACT at both grades, reviewers noted that items directly assessing procedural skill were underrepresented. For MCAS at grade 5, reviewers found few items assessing conceptual understanding and an overabundance of application items. The grade 5 PARCC exam similarly had an overabundance of application items, some of which reviewers noted had shallow contexts. Finally, the Smarter Balanced exams at both grade levels had a slight wealth of application items, and reviewers also noticed that some forms were more heavily focused on applications than others.

The following were required to fully meet this criterion:

On each test form, at least 25 percent and no more than 50 percent of score points are allocated to each of the three categories: mathematical concepts, procedures/fluency, and applications.

Mathematics Depth Results

Criterion C.3 -

*”Do the tests connect
mathematical practices to
content?”*



ACT Aspire

EXCELLENT MATCH

All items that are coded to mathematics practices are also coded to one or more content standard.



2014 MCAS

EXCELLENT MATCH

Although no items are coded to mathematical practices, the practices were assessed within items that also assessed content.



PARCC

EXCELLENT MATCH

All items that are coded to mathematics practices are also coded to one or more content standard.



Smarter Balanced

EXCELLENT MATCH

All items that are coded to mathematics practices are also coded to one or more content standard.

The following were required to fully meet this criterion:

1. All or nearly all items that assess mathematical practices also align to one or more content standards.

Mathematics Depth Results

Criterion C.4 -

”Are all students required to demonstrate a range of high-order, analytical thinking skills in mathematics based on the depth and complexity of the standards?”



ACT Aspire

LIMITED/UNEVEN MATCH

At both grades 5 and 8, the test forms include significantly more items of high cognitive demand (DOK 3) than reflected in the standards, and proportionately fewer at the lowest level (DOK 1). While these items increase the challenge of the tests, standards that call for the lowest level of cognitive demand (DOK 1) may be under-assessed.



2014 MCAS

EXCELLENT MATCH

At each grade level, the distribution of cognitive demand closely reflects that of the standards.



PARCC

GOOD MATCH

The distribution of cognitive demand of items reflects that of the standards very well at grade 5, while the grade 8 test includes proportionately more items at the higher levels of cognitive demand (DOK 2 and 3). As a result, grade 8 standards that call for the lowest level of cognitive demand may be under-assessed.



Smarter Balanced

GOOD MATCH

The distribution of cognitive demand of items reflects that of the standards very well at grade 5. At grade 8, the test includes proportionately fewer items at the lowest levels of cognitive demand (DOK 1) than in the standards, and proportionately more items at the mid-level of cognitive demand (DOK 2). As a result, grade 8 standards that call for the lowest level of cognitive demand may be under-assessed.

The following were required to fully meet this criterion:

To receive the highest rating on this criterion, the distribution of cognitive demand on test forms had to match the distribution of cognitive demand of the standards as a whole and match the higher cognitive demand (DOK 3+) of the standards. (As was the case in the ELA review of cognitive demand, this is not a rating of test difficulty. Assessments that do not match the distribution of complexity of the standards, including if they have too many high Depth of Knowledge items, may receive a rating of less than Excellent Match.

Mathematics Depth Results

Criterion C.5 -

”Are a variety of item types used, including at least one that requires students to generate, rather than select, a response? Are the test items of high quality?”



ACT Aspire

EXCELLENT MATCH

The program uses multiple item types, including constructed response on the Extended Task items. These items, although they carry high point values, are limited in number; the rest of the items are predominantly multiple-choice.

The large majority of items are of high technical and editorial quality, with only very minor issues of editing, language or accuracy. At the grade 8 level, some items appear to be susceptible to simplification by use of calculators, which are allowed on all items at grade 8, in contrast to the other programs that allow them on a restricted set of items.



2014 MCAS

EXCELLENT MATCH

Both grade 5 and grade 8 forms include multiple item types, including constructed-response. The items are of high technical and editorial quality, with very minor issues of editing, language, and accuracy at grade 8.



PARCC

GOOD MATCH

The program includes a wide variety of item types, including several that require student-constructed responses. However, there are a number of items with quality issues, mostly minor editorial but sometimes mathematical.



Smarter Balanced

LIMITED/UNEVEN MATCH

The program includes a wide variety of item types, many of which make effective use of technology.

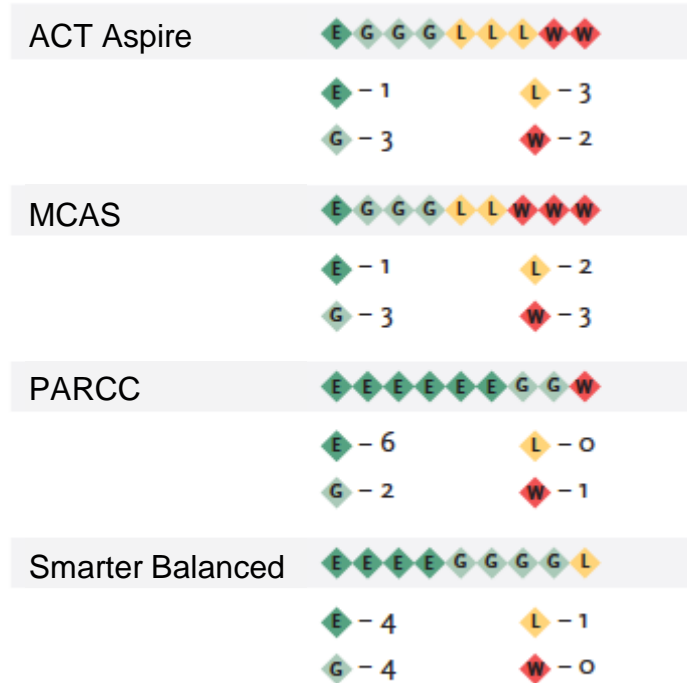
The program could be improved by ensuring that virtually identical items are not presented to individual students. Further, a good deal of variability across forms and grades is observed, with some forms fully meeting the item quality criterion and others only partially meeting it. Issues exist with the editorial quality and mathematical accuracy of individual items, most of which are minor but some of which could impact assessment of the targeted skill, resulting in a rating of Limited/Uneven.

The following were required to fully meet this criterion:

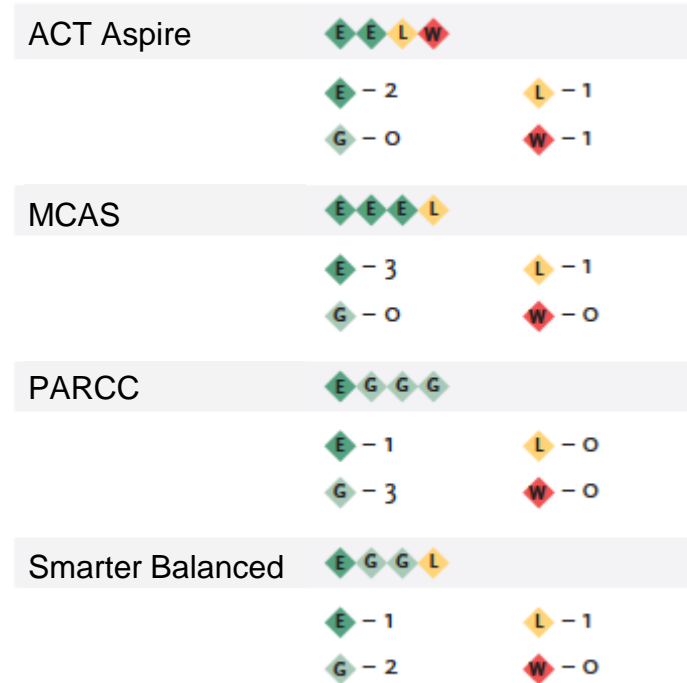
1. At least two item formats are used, including one that requires students to generate, rather than select, a response.
2. All or nearly all operational items reviewed reflect both high technical quality and high editorial accuracy.

Ratings Tally by Program [to be re-formatted without the counts below]

ELA/Literacy Ratings Tally by Program



Mathematics Ratings Tally by Program



LEGEND  Excellent Match  Good Match  Limited/Uneven Match  Weak Match

Program Strengths and Areas for Improvement

ACT Aspire

In **ELA/Literacy**, ACT Aspire receives a Limited/Uneven to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The combined set of ELA/Literacy tests (reading, writing, and English) requires close reading and adequately evaluates language skills. More emphasis on assessment of writing to sources, vocabulary, and research and inquiry, as well as increasing the cognitive demands of test items, will move the assessment closer to fully meeting the criteria. Over time, the program would also benefit by developing the capacity to assess speaking and listening skills.

In **mathematics**, ACT Aspire receives a Limited/Uneven to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. Some of the mismatch with the criteria is likely due to intentional program design, which requires that items be included from previous and later grades.

The items are generally high quality and test forms at grades 5 and 8 have a range of cognitive demand, but in each case the distribution contains significantly greater emphasis at DOK 3 than reflected in the standards. Thus, students who score well on the assessments will have demonstrated a strong understanding of the standards' more complex skills. However, the grade 8 test may not fully assess standards at the lowest level of cognitive demand. The tests would better meet the CCSSO Criteria with an increase in the number of items focused on the major work of the grade and the addition of more items at grade 8 that assess standards at DOK 1.

Program Strengths and Areas for Improvement

MCAS

In **ELA/Literacy**, MCAS receives a Limited to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The test requires students to closely read high-quality texts and a variety of high-quality item types. However, MCAS does not adequately assess several critical skills, including reading informational texts, writing to sources, language skills, and research and inquiry; further, too few items assess higher-order skills. Addressing these limitations would enhance the ability of the test to signal whether students are demonstrating the skills called for in the standards. Over time, the program would also benefit by developing the capacity to assess speaking and listening skills.

In **mathematics**, MCAS receives a Limited/Uneven Match to the CCSSO Criteria for Content and an Excellent Match for Depth relative to assessing whether students are on track to meet college and career readiness standards. The MCAS mathematics test items are of high technical and editorial quality. Additionally, the content is distributed well across the breadth of the grade level standards, and test forms closely reflect the range of cognitive demand of the standards. Yet the grade 5 tests have an insufficient degree of focus on the major work of the grade.

While mathematical practices are required to solve items, MCAS does not specify the assessed practices(s) within each item or their connections to content standards. The tests would better meet the criteria through increased focus on major work at grade 5 and identification of the mathematical practices that are assessed—and their connections to content.

Program Strengths and Areas for Improvement

PARCC

In **ELA/Literacy**, PARCC receives an Excellent Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The tests include suitably complex texts, require a range of cognitive demand, and demonstrate variety in item types. The assessments require close reading, assess writing to sources, research, and inquiry, and emphasize vocabulary and language skills. The program would benefit from the use of more research tasks requiring students to use multiple sources and, over time, developing the capacity to assess speaking and listening skills.

In **mathematics**, PARCC receives a Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The assessment is reasonably well aligned to the major work of each grade. At grade 5, the test includes a distribution of cognitive demand that is similar to that of the standards. At grade 8, the test has greater percentages of higher-demand items (DOK 3 and 4) than reflected by the standards, such that a student who scores well on the grade 8 PARCC assessment will have demonstrated strong understanding of the standards' more complex skills. However, the grade 8 test may not fully assess standards at the lowest level (DOK 1) of cognitive demand.

The test would better meet the CCSSO Criteria through additional focus on the major work of the grade, the addition of more items at grade 8 that assess standards at DOK 1, and increased attention to accuracy of the items—primarily editorial, but in some instances mathematical.

Program Strengths and Areas for Improvement

Smarter Balanced

In **ELA/Literacy**, Smarter Balanced receives a Good to Excellent Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The tests assess the most important ELA/Literacy skills of the CCSS, using technology in ways that both mirror real-world uses and provide quality measurement of targeted skills. The program is most successful in its assessment of writing and research and inquiry. It also assesses listening with high quality items that require active listening, which is unique among the four programs. The program would benefit by improving its vocabulary items, increasing the cognitive demand in grade 5 items, and, over time, developing the capacity to assess speaking skills.

In **mathematics**, Smarter Balanced has a Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The test provides adequate focus on the major work of the grade, although it could be strengthened at grade 5.

The tests would better meet the CCSSO Criteria through increased focus on the major work at grade 5 and an increase in the number of items on the grade 8 tests that assess standards at the lowest level of cognitive demand. In addition, removal of serious mathematical and/or editorial flaws, found in approximately one item per form, should be a priority.