Assessing Today’s Standards:
Multi-Dimensional Test Items

May 2018

Thoughtful, excellent student assessment instruments aren’t just for statewide testing. Educators and measurement specialists have expanded their expectations of how assessment informs learning, based in part on 21st-century standards, new emphases in ESSA, and the growing awareness of the importance of a balanced assessment system. At the same time, the issue of testing time has become prominent in professional and public awareness.

At Measured Progress, as we began to explore the implications of these issues and the new standards—first, the widely adopted college and career readiness (CCR) standards in English language arts and mathematics, and then the emerging Next Generation Science Standards (NGSS®)—it was clear that the new standards require a better approach to assessment. We need to provide timely, targeted instructional information at the district and school levels, and effective, meaningful data for decision making at all levels, including statewide. We need to use the least amount of time to provide the most information possible—to create test designs and test items that deliver strong results, fast.

To accomplish these goals, we created item specs based on the information we want to provide to states, districts, schools, teachers, students, and parents. This evidence-centered design approach is based on extensive research about student learning progressions and grade-level expectations, as well as on detailed scrutiny of the standards.

To make the results more valuable for educators, our content provides results at greater depth than other instruments. We target our content development based on what we want to report: multiple dimensions for each subject area to indicate students’ strengths and needs at an actionable level.

For each content area, this paper provides information about the approach and presents illustrative sample items.

eMPower Assessments: College and Career Readiness

eMPower Assessments™ give districts and states a coherent program for grades 3–8 in reading, writing and language, and mathematics. With all-new content written to CCR standards, three administrations are available per year to provide reliable data that help districts measure student growth toward grade-level standards. States may implement the spring administration as part of their accountability systems. eMPower also gives early insight into college and career readiness through a direct predictive link with the score scale of the SAT® Suite of Assessments.

We probed the research behind CCR standards to inform the development of our detailed item specifications, as described on the following pages. (See also “Research Basis for eMPower Standards.”)

* NGSS is a registered trademark of Achieve. Neither Achieve nor the lead states and partners that developed the Next Generation Science Standards were involved in the production of this product, and do not endorse it.
Assessing Today’s Standards: Multi-Dimensional Test Items

Reading

eMPower reading items assess essential college and career readiness standards. These standards encompass comprehension and critical thinking skills (analysis and interpretation). Many assessments report an overall score of students’ aptitude in these skills; eMPower goes further by reporting one score for comprehension and one score for analysis and interpretation for every student. To accomplish this, items appearing on each eMPower reading assessment are categorized as relating to one of those areas. Items designated as “comprehension,” such as the one that follows, ask students to effectively demonstrate key content knowledge. Items designated as “analysis and interpretation” evaluate students’ ability to examine, synthesize, and explain key concepts in texts.

eMPower passages and passage pairs are carefully selected to assess a variety of reading skills at a range of cognitive levels. To genuinely assess the standards, eMPower reading assessments are based on:

- Rich and complex authentic texts that are previously published works, similar to what students experience in the classroom
- Literary and informational passages, including scientific or historical texts that are grade-level appropriate
- Paired passages, because students are expected to comprehend and synthesize multiple points of view and sources

The example below is a cognitively rich, evidence-based selected-response item that assesses students’ ability to make key inferences from their reading and to support their inferences with details from the text. The items refer to a passage called “The King’s Journey,” by Betsy Sterman.¹

This question has two parts. Be sure to answer both parts of the question.

In Passage 2, what is the main problem the King tries to solve?

A  It takes too long to travel through his lands.
B  The roads are not fit for traveling.
C  His subjects do not amuse him.
D  He knows very few people.

Which detail from the passage best supports the answer to the question above?

A  “This is the swiftest carriage in all the land,’ Sir Highfeather said proudly.”
B  “Ohhh,’ he groaned. ‘What a great reward I would give to anyone who could shorten my journey.”
C  “The bridges were in ruins, their planks ripped apart and their ropes whipping in the wind.”
D  “A small story it was, but so funny that the king laughed and asked for another.”

STANDARD: RL.04.01
KEYS: A, B

Writing and Language

eMPower writing and language assessments assess students' abilities to recognize good writing and to make good revisions, as well as to demonstrate their command of English language conventions. As in the College Board’s PSAT™, we accomplish this by using embedded-error passages.

The embedded-error approach has several advantages. First, it creates a very natural way for students to experience the passages. Instead of asking rote questions about grammar rules or parts of speech, eMPower writing and language tests present text holistically, asking students to respond to written text rather than to address grammar and mechanics out of context. This approach addresses writing skills and language skills together as students experience grade-level appropriate writing. Second, students have experience with the embedded-error approach before they see it in the PSAT or other tests in the SAT® Suite of Assessments. Also, the embedded-error approach requires minimal testing time and provides results more quickly and at lower cost than other methods that assess these skills.

To provide students the holistic experience with each passage, we write our embedded-error passages to be good works of writing, not to exemplify problem writing. We write the passages first, and determine what to modify to ask meaningful questions, second. We craft the passages according to the grade-range genres emphasized in college and career readiness standards: narrative and expository for grades 3-5, expository and argument in grades 6-8.

The example that follows assesses students’ ability to include effective transitions in their writing. The items refer to a passage called “Great Aunt Lily’s Attic,” by Measured Progress.

This question has two parts. Be sure to answer both parts of the question.

Which sentence should be added after sentence 26 to make the best ending for the passage?

A Chair: It’ll be nice to be part of the family again!
B Chair: I wonder if the living room has changed much.
C Chair: I can’t wait to not be covered in dust anymore!
D Chair: It’ll be hard to not miss the peace and quiet a little.

Which detail from the passage best supports the answer above?

A “That got me rocking, and I woke up.”
B “Do you recall when we used to be in the living room?”
C “Album: I wish we were back in the living room with the family.”
D “Sara: Let’s get some cloths to dust everything off.”

STANDARD: W.04.03.e
KEYS: A, C

“Great Aunt Lily’s Attic” © 2015 by Measured Progress.
In the first part of this EBSR, the student analyzes the passage to determine the best conclusion for the narrative. The second part asks the student to select specific evidence from the passage to support that choice. This provides some of the depth of an open-response question, but gets at it efficiently.

eMPower writing and language assessments provide important data to teachers regarding their students' proficiency with written communication. If a state, district, or school also uses a direct writing prompt, eMPower results supplement and strengthen the evidence from the prompt.

Mathematics

From the outset, we wanted to break new ground in mathematics assessment by reflecting the two dimensions of the college and career readiness standards in mathematics: concepts and procedures, and mathematical practices. To accomplish this, our content developers probed the practices to create detailed “focus points” and item specifications for each practice—beyond those for concepts and procedures. (See “Sample: eMPower Assessments Mathematical Practices Focus Points.”)

Each eMPower mathematics assessment includes items that address a range of focus points within each mathematical practice. The specific mathematical practice focus points, and dual coding to both concepts and procedures and to practices, are unique to Measured Progress and to eMPower.

The item that follows assesses mathematical practice 3 for grades 3–5: “Construct viable arguments and critique the reasoning of others.” To assess this practice, we developed two distinct focus points: construct arguments (Practice 3A), and evaluate arguments (Practice 3B). The item on the next page assesses the second focus point—evaluating arguments. The primary C&P standard assessed is 03.MD.07.d.
Assessing Today’s Standards: Multi-Dimensional Test Items

Students were asked to find the area of this figure. Five children tried to find the area.

![Diagram of a figure with dimensions: 8 feet by 3 feet, 8 feet by 2 feet, and 2 feet by 3 feet.]

- **Kelly**
  - $8 \times 2 = 16$
  - $8 \times 3 = 24$
  - $16 + 24 = 40$
  - 40 square feet
- **Trina**
  - $8 \times 2 = 16$
  - $6 \times 3 = 18$
  - $16 + 18 = 34$
  - 34 square feet
- **Sandy**
  - $8 \times 8 = 64$
  - $2 \times 3 = 6$
  - $64 - 6 = 58$
  - 58 square feet
- **Won**
  - $8 \times 3 = 24$
  - $2 \times 5 = 10$
  - $24 + 10 = 34$
  - 34 square feet
- **Bob**
  - $8 \times 3 = 24$
  - $2 \times 8 = 16$
  - $24 + 16 = 40$
  - 40 square feet

Which **two** children’s work and answer are correct?

- A. Kelly
- B. Trina
- C. Sandy
- D. Won
- E. Bob

**Keys:** B, D

**Distractor Rationales:**
- A. Student multiplies given side measures, and does not decompose rectangles.
- B. Key
- C. Student multiplies given side measures, and does not understand how to decompose rectangles.
- D. Key
- E. Student multiplies given side measures, and does not decompose rectangles.

**Depth of Knowledge:** 2
**Focus/Bullet:** 3B/Bullet 3
**Primary Content Domain/Cluster:** 03.MD.03.07.d
**Secondary Content Domain/Cluster:** 03.MD.03.07.b
**Additional Mathematical Practices:** Problem Solving, Focus 1C/Bullet 1

By assessing specific practice focus points as well as concepts and procedures, eMPower mathematics assessments provide a second key dimension of insight to support instructional decisions.
Science Secure Item Bank: 3-Dimensional Standards

The new three-dimensional science standards call for a shift in how we teach and assess science. Integrating the Science and Engineering Practices with the Disciplinary Core Ideas and the Crosscutting Concepts requires new thinking about how to write items and how to design tests—to get at the complexity of the standards without excessive test length.

Based on these needs, Measured Progress is creating a secure item bank for science, based on A Framework for K–12 Science Education and the Next Generation Science Standards (NGSS®). The item bank will provide clients with an ever-increasing selection of high-quality assessment items that they can use to create tests.

We agree with the recommendations of the NRC BOTA report and the CCSSO SAIC assessment framework that because the NGSS performance expectations (PEs) are complex, with significant depth, breadth, and rigor, we need richer tasks with which to assess students. We can't assess these complex, multi-dimensional science performance expectations with only simple, discrete, primarily multiple-choice items as we have in the past.

Our item cluster model addresses this challenge. An item cluster is a set of items all associated with a common stimulus that presents a science phenomenon or engineering design problem. From an instructional perspective, the best units for three-dimensional learning follow a storyline, in which students begin with an overarching question about a phenomenon or problem and then figure out the progression of research, investigation, and analysis needed to make sense of the situation. When we are assessing students, we want them to be following a similar process to explain a phenomenon or solve a real problem, so we create clusters with items that mirror the kinds of thinking and sequence of actions that students are engaging in in the classroom. We plot out a storyline that moves from gathering information, to reasoning, to making sense of the original phenomenon.

Most of our item clusters are aligned to a PE bundle, meaning a set of PEs that together provide the construct knowledge to explain the phenomenon or solve the problem presented in the stimulus. Our PE bundles typically incorporate two or three PEs. One advantage of clusters is that they thoroughly assess the targeted PEs. However, with constraints on testing time, relying on clusters may not provide adequate standards coverage in each test administration. To offer flexibility to increase the number of PEs covered on a test, we will also include some stand-alone items in the bank.

All the items in our science item bank align to more than one dimension. Both stand-alone items and individual items in clusters are at least two-dimensional, if not three-dimensional. Also, each cluster collectively assesses all three dimensions of the target PEs. This multi-dimensional alignment, combined with the predominant use of item clusters, ensures coverage of the breadth, depth, and core intent of the

---

3 NGSS is a registered trademark of Achieve. Neither Achieve nor the lead states and partners that developed the Next Generation Science Standards were involved in the production of this product, and do not endorse it.


Assessing Today’s Standards: Multi-Dimensional Test Items

PEs. Further, when we evaluate items for alignment, we indicate dimension alignment only for items that truly require the use of a dimension, rather than those that merely use the context of a dimension. In this way we assure strong fidelity to and measurement of the constructs in the standards.

The following pages display a grade 5 cluster that assesses PE 5-ESS1-2, which requires students to “represent data in graphical displays to reveal patterns of daily changes in length and direction of shadows, day and night, and the seasonal appearance of some stars in the night sky.”

This cluster specifically looks at shadows. The “hook” in the stimulus on the following page is a student seeing her shadow changing—something that many elementary students notice when playing outside. She wants to know if there is a pattern for the change. The stimulus goes on to provide details that will support students in making sense of the shadow changes, with an investigation setup and related data to analyze and interpret.

### Investigating Shadows

A student sees that the length of her shadow changes during the day. She wants to find out whether there is a pattern for how her shadow changes. To test this, she pierces a pin through the top of a box and measures the length of the pin’s shadow several times on a mostly sunny winter day. This diagram shows the pin and the box at 9:00 a.m. The data table shows the lengths of the shadow at different times during the day. At 10:00 a.m., the sky was cloudy and there was no shadow.

![Investigating Shadows Diagram]

<table>
<thead>
<tr>
<th>Time</th>
<th>Length (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 a.m.</td>
<td>45</td>
</tr>
<tr>
<td>10:00 a.m.</td>
<td>no shadow</td>
</tr>
<tr>
<td>11:00 a.m.</td>
<td>29</td>
</tr>
<tr>
<td>12:00 p.m.</td>
<td>27</td>
</tr>
<tr>
<td>1:00 p.m.</td>
<td>32</td>
</tr>
<tr>
<td>2:00 p.m.</td>
<td>39</td>
</tr>
</tbody>
</table>

The items in the cluster then assess whether students can indeed represent data in graphical displays to reveal patterns of daily changes in the length of shadows, as the PE specifies.
Item 1 is a TEI focusing on transforming the data from the table into a bar graph. This item is two-dimensional; it aligns to the Science and Engineering Practice (SEP) of analyzing data, and to the Crosscutting Concept (CCC) dimension of patterns, asking students to include an estimate for the shadow length at the missing time. Thus students have to apply both their basic graphing skills and their knowledge of patterns.

1. Complete the bar graph to show the length of the pin's shadow each hour from 9:00 a.m. to 2:00 p.m. Include an estimate for 10:00 a.m. Drag each bar to the correct value to change the heights of the bars.

![Bar Graph](image)

Item 2 then asks students to look at the patterns and the cause of the patterns in the data, bringing them further into the PE’s expectation to reveal and understand patterns related to the Disciplinary Core Idea (DCI) concept of the motion of celestial objects—in this case, Earth’s rotation. This item is also two-dimensional, addressing the CCC and DCI dimensions of the PE. The multi-part structure of the item assures readability and clarity, and the multi-select format for Part a encourages students to dig into the patterns for shadow length.
2. **Part a**
   Which statements describe the pattern of change in the length of the pin's shadow? Select all that apply.
   ① The length of the shadow increased and decreased in the morning.
   ② The length of the shadow decreased in the morning and increased in the afternoon.
   ③ The length of the shadow changed less during the middle of the day than during the early morning or late afternoon.
   ④ The length of the shadow changed by the same amount each hour.

**Part b**
What causes the pattern of change in the length of the pin's shadow during a winter day?
① Earth orbits the Sun.
② The Sun orbits Earth.
③ Earth rotates on its axis.
④ The Sun rotates on its axis.

Item 3 also focuses on the CCC and DCI dimensions, extending the students' demonstration of their understanding of the patterns related to shadow formation. Here the students show their understanding that Earth's rotation and the apparent motion of the Sun across the sky will create a pattern in the length of the shadow, and also in its direction.

3. The student observed that the Sun's position in the sky appeared to move over the time she recorded her data. What other pattern could she observe to provide evidence of this?
   ① The position of the box shifted as time passed.
   ② The height of the pin decreased as time passed.
   ③ The direction of the pin's shadow changed for each measurement.
   ④ The thickness of the pin's shadow increased for each measurement.
Item 4 serves as the culminating task in the cluster, requiring synthesis to make sense of this shadow phenomenon. Students demonstrate their understanding by extending their explanation to shadow lengths in summer, based on their understanding of shadow lengths in winter. This item is three-dimensional, requiring students to use their DCI knowledge of celestial motions and effects, their CCC knowledge of patterns, and their SEP knowledge and skill in representing data in displays.

In Part a, students create a table of the expected measurements for the shadow lengths in the summer. The bracketed numbers in the Length column show the options presented to students when they click on a dropdown icon in each cell. Students choose their answer from those options.

4. Six months later, the student repeats her measurements of the length of the pin’s shadow on a sunny summer day. She compares the shadow lengths from the winter day with those from the summer day.

**Part a**
Identify what the measurements of the pin’s shadow from the summer day will most likely be.

<table>
<thead>
<tr>
<th>Time</th>
<th>Length (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 a.m.</td>
<td>[38, 45, 55]</td>
</tr>
<tr>
<td>10:00 a.m.</td>
<td>[26, 43, 60]</td>
</tr>
<tr>
<td>11:00 a.m.</td>
<td>[21, 29, 64]</td>
</tr>
<tr>
<td>12:00 p.m.</td>
<td>[19, 27, 66]</td>
</tr>
<tr>
<td>1:00 p.m.</td>
<td>[24, 32, 62]</td>
</tr>
<tr>
<td>2:00 p.m.</td>
<td>[31, 39, 53]</td>
</tr>
</tbody>
</table>
In Part b, students identify the reasoning for their response to Part a. This underscores the importance and emphasis in the NGSS on evidence, reasoning, and explanations.

**Part b**

Which statement explains the pattern of change in the length of the pin's shadow during a summer day compared to a winter day?

- A The Sun appears higher in the sky during summer because Earth is tilted on its axis and orbits the Sun.
- B The light from the Sun is brighter during summer because Earth is tilted on its axis and closer to the Sun.
- C More sunlight reaches Earth during summer because Earth rotates toward the east at this time of year.
- D The Sun shines on Earth for a longer time each day during summer because Earth rotates more slowly at this time of year.

For more information on eMPower Assessments and the Science Secure Item Bank—or to learn about other Measured Progress products and services—email info@measuredprogress.org.