How do you know if they’re getting it? Writing assessment items that reveal student understanding

by Melanie Taylor and Sean Smith

After a unit on seasons, a science teacher included the following item on a test:

Which of the following locations would you expect to have the least temperature variation throughout the year?
(a) Toronto, Canada
(b) London, England
(c) Sydney, Australia
(d) Manta, Ecuador

The teacher had done a thorough job teaching the unit, and based on student responses to questions she asked during class, she was confident they would recognize a location on the equator as the correct answer. She was disappointed and confused when three-fourths of the class missed the item. As she returned the tests, she asked several students why they got the question wrong. Almost without exception, each student asked, “Where’s Ecuador?”

Most teachers have probably had a moment of realization similar to this. They write a test item they are sure is crystal clear, only to find later that students lacked some prerequisite knowledge, or that they interpreted the item differently than intended. Through a project funded by the National Science Foundation, Horizon Research has been developing assessment items for students (in the process, compiling item-writing principles from several sources and adding their own). In this article, we share what we have learned about writing items that reveal student understanding, including best practices and concrete examples of applying them. Although most of the examples given are multiple-choice items, the principles are generally applicable to open-ended questions as well.

Principle 1: Define and clarify the content

The first step in writing a good assessment item is deciding what is “fair game.” Drawing clear boundaries around the intended content ensures that items focus on what we care about and prevents us from introducing content that is out of bounds. In the introductory example, the teacher unintentionally introduced specific geography content. We have had the luxury to spend time thinking with scientists about the content for which we write our assessments. One of our tests focuses on the following statement from Benchmarks for Science Literacy (AAAS 1993):

Food provides the molecules that serve as fuel and building material for all organisms. Plants use the energy from light to make sugars from carbon dioxide and water. This food can be used immediately or stored for later use. Organisms that eat plants break down the plant structures to produce the materials and energy they need to survive. Then they are consumed by other organisms.

We broke down this statement into a series of smaller “sub-ideas” shown in Figure 1. Teachers have
lesson plans, textbooks, district curriculum materials, and state standards or frameworks that can help break down “big” ideas or concepts into smaller pieces. For our own development work, we supplemented the breakdown of the content with a literature review on student misconceptions. Knowing this information helps in writing items that uncover common misconceptions students have and suggests distractors for multiple-choice items. Again, teachers rarely have time for this work, but publications such as Benchmarks for Science Literacy (AAAS 1993) and the Atlas of Science Literacy (AAAS 2001; AAAS 2007) distill much of the research on student thinking and link it to specific science ideas.

**Principle 2: Target the item**

When writing assessment items, try to target just one sub-idea per item. This makes each item precise, enabling a specific and clear assessment of students’ understanding of a particular idea. Plus, the cognitive load of the item, or how much information a student needs to consider when answering the item, increases as an item assesses more ideas. There is an important place for items that ask students to make connections across several ideas. However, if we don’t ask the targeted items, we won’t have a good idea of why students struggle with the more complex ones—do students lack understanding of one or more pieces of content, or are they having difficulty relating the various ideas? Consider the question, “A group of friends is discussing what they learned in science class. One friend says she learned that two of Earth’s plates are slowly moving away from each other. Another friend says he learned that those two plates are touching each other, and always will be. Which student is correct and why?” The question includes two ideas—plates can move away from each other and plates abut other plates on all sides—but requires a third idea (new plate material is constantly being formed) to recognize that both students are correct and explain why. It’s a good synthesis question, but if a student answers incorrectly, the teacher may not know which of the three ideas the student is missing unless the test includes other items that are more targeted. For instance, the following item targets one idea—plates abut other plates on all sides.

Which of the following is always found next to the edge of a plate?

(a) Another plate
(b) A rift valley
(c) An oceanic trench
(d) A large pool of magma
Principle 3: Necessity

Students should have to know the target content to answer a question correctly. Put another way, students should not be able to answer the item correctly without knowing the content. Some refer to this as the “necessity criterion” (Stern and Ahlgren 2002). So how could a student answer an item correctly, without guessing, and not know the content? Middle school students have become test savvy. They look for an answer choice that is longer than the rest or shorter than the rest, or they match patterns of words in the item itself with words in the answer choices. Often, these clues lead them to the right answer. A case in point is the following question about the circulatory system in:

In humans, which of the following systems is responsible for moving substances to and from cells?

(a) The urinary system
(b) The circulatory system
(c) The digestive system
(d) The respiratory system

Students who have not studied human body systems often chose the correct answer (b). Students were simply matching the word “moving” in the question to the word “circulatory” in the answer choice.

Principle 4: Sufficiency

Perhaps the most difficult principle to follow is the “sufficiency criterion” (Stern and Ahlgren 2002), which goes hand-in-hand with necessity. Sufficiency here means that the knowledge in the target content is all a student needs to know to answer the item correctly; answering the item correctly requires no knowledge outside the target content. The assessment item at the beginning of this article clearly failed this criterion, requiring students to know geography content. Again, there is a time for asking students to synthesize different kinds of content, but if a teacher’s goal is to assess understanding of a particular idea, the item should meet the sufficiency criterion.

The necessity and sufficiency criteria are important because they affect the validity of an assessment. We want to know that if students get a question right, it is because they used the content being assessed. Conversely, if students get a question wrong, it should be because they did not know the content. We do not want students to use knowledge outside the target content to answer items through process of elimination, or to answer an item incorrectly because it required knowledge outside the target content. We want items to reveal what students understand about the content we are assessing.

So when is it fitting to ask students to draw on various types of content to answer an item? First, if a teacher is confident that the knowledge is prerequisite (e.g., elementary science content), it is reasonable to expect students to make use of it. Conversely, if students get a question wrong, it should be because they did not know the content. We do not want students to use knowledge outside the target content to answer items through process of elimination, or to answer an item incorrectly because it required knowledge outside the target content. We want items to reveal what students understand about the content we are assessing.

Principle 5: Don’t teach

In writing items, it is important to keep the entire assessment in mind to avoid one item teaching content that may be assessed in another. Providing some context can be important for situating the content of an item, but we have to be careful that the context does not teach other ideas we want to assess. (We don’t believe that students can learn science well from reading...
an assessment item, but the test-wise student will use clues in one item to answer another one.) For example, one item about whether sugars are food for plants stated, “Plants make sugars from carbon dioxide and water. Are sugars food for plants? Explain why or why not.” We quickly realized that this item teaches students that plants make sugars from carbon dioxide and water, one of the other ideas included on the assessment. To fix the problem, simply delete the introductory sentence.

**Principle 6: Mirroring**

“Mirroring” means that all answer choices should reflect the question asked. Mirroring helps students know what to expect from the answer choices and prevents them from ruling out certain choices simply because they do not answer the question asked. For example, here is an item that violates this principle:

How do plants get their food?
(a) Plants make food from minerals, water, and sunlight.
(b) Plants make food from carbon dioxide, water, and sunlight.
(c) Plants get food by absorbing water through their leaves.
(d) Plants get food by absorbing nutrients through their roots.

The item asks how plants “get” their food, but only two answer choices say “get.” Students may rule out choices a and b because they do not mirror the question asked. Because b is the correct answer, a student would miss the item for the wrong reason. Similarly, we do not want students to rule out wrong answers just because the answers do not address the question asked, in essence making the item easier and unable to provide valid information about what the student knows.

**Principle 7: Maximize comprehension**

Assessment items should use language that is accessible to as many students in the class as possible. Items should assess students’ knowledge and understanding of science content, not their reading ability. It is impossible to completely disentangle...
the two, but following certain principles can facilitate comprehension, which is especially important for English language learners. The guidelines discussed here were developed specifically for these students (Kopriva 2000), but have proved helpful for all students:

- First, sentences should be brief and straightforward, with a simple phrase structure and no added clauses. Some items should provide students with engaging contexts, but the content of the item should not become lost in the context.
- Second, items should restrict language to present tense and active voice. Students who are learning English learn these constructions first. One assessment item in force and motion originally read, “A shopping cart is pushed so that it moves faster and faster.” To increase the likelihood of comprehension, we changed it to read, “A man is pushing a shopping cart. The cart is moving faster and faster.”
- Third, any visuals such as diagrams, tables, and pictures should promote understanding of what the item is asking. Do not distract students unnecessarily by including a picture just to include a picture. Unless the point of an item is to have students interpret a visual (e.g., a table of data), no information should appear in the visual that the text of the item does not already include. Also, it is important that any labels in the diagram are consistent with the wording used in the item.

These three guidelines for maximizing comprehension are particularly important, but there are many more. The Council of Chief State School Officers has produced a helpful, comprehensive guide titled Ensuring Accuracy in Testing for English Language Learners (See Resources; Kopriva 2000).

**Principle 8: Avoid test-wiseness**

As we mentioned above, students often try to use test-wiseness rather than science content to answer items, especially when students are unfamiliar with the content. What else can teachers do to ensure that an item is assessing understanding and not test-taking skills? Figure 2 shows some of these and the accompanying item-writing guidelines. These guidelines are specific to multiple-choice questions.

**Now it’s your turn**

Our assessment-development efforts have taught us an important lesson: It is deceptively easy to write a bad item. Writing good items is hard work. Figure 3 is a checklist from the principles discussed in this article; most apply equally well to multiple-choice and open-ended items. If you are wondering how to start applying the principles, you might consider evaluating some of your existing assessments against the checklist. As the principles become second nature, you will gain confidence that your assessment items reveal students’ thinking about the content, and you might just avoid any “Where’s Ecuador?” moments.

**Acknowledgment**

The work described in this article is funded by the National Science Foundation (NSF) under grant number EHR-0335328. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

**References**


**Resources**

The Council of Chief State School Officers, Ensuring Accuracy in Testing for English Language Learners guide—www.ccsso.org

---

Melanie Taylor (mtaylor@horizon-research.com) is a research associate and Sean Smith (ssmith62@horizon-research.com) is a senior research associate for Horizon Research, Inc., located in Chapel Hill, North Carolina.