Assessing the Skills and Abilities in Math and Science of ELLs with Low English Proficiency: A Promising New Method

Rebecca Kopriva

How do you reliably assess ELLs' real knowledge and skills in math and science? Accommodations, if properly identified and used (which is still quite a challenge), seem to be effective for ELLs with higher proficiency in English. However, traditional methods for testing students do not work well for ELLs with lower levels of English acquisition. For instance, ELL students with lower proficiency levels do not perform well on:

- multiple-choice items, because the required discriminations between choices demand fine-tuned language skills; or
- constructed-response items, unless scoring procedures are in place to allow for code-switching and greater use of visuals.

Typical scoring is a problem because many ELLs lack the necessary productive language skills. Besides the ELLs' lack of language skills, many tests, particularly large-scale assessments, require cultural and background knowledge outside the experience of lower-English-proficient ELL students and these unfamiliar contexts can confuse, rather than assist, their comprehension.

Two *Obtaining Necessary Parity through Academic Rigor* (ONPAR) grants have been funded through the U.S. Department of Education (USDE) for the purpose of building prototype large-scale items in science and mathematics that would be appropriate for ELLs with little proficiency in English. The computer-based items are being built to be interactive. Multi-semiotic representations, such as animation and simulation, greatly minimize the use of text in presenting the item questions. As response opportunities have been a major stumbling block for these students, the ON-PAR items have created novel approaches that allow students to interact with stimuli and demonstrate what they know with almost no language. Native or home language (L1) support and additional visual cues are used to support words or phrases, and to 'act out' action language presented in the remaining text.

The items being developed are particularly impressive because they reflect more cognitively complex maths and science problems. Many recall items can be handled adequately with plain language and static visuals, and do not need many interactive computer capacities. More complex academic items, on the other hand, generally require more complex and abstract language to express the questions suitably and register responses. As such, ONPAR has focused on whether dynamic computer capacities can be used effectively to limit or omit abstract language without changing the complexity of the targeted science or math content.

Research and development questions

To test their viability and effectiveness, ONPAR items are being built from traditional items, with the goal of measuring the same targeted content as the original item. Several steps were involved in developing the ONPAR items. For instance, the construct-relevant and construct-irrelevant components of traditional test items had to be identified (i.e., what portions of an item are necessary to determine a student's skills and knowledge, and what portions of an item are extraneous to the content being tested), so the construct-relevant, or targeted, portions could be translated to the ONPAR versions while the irrelevant components which cause problems could be reduced in the ONPAR items.

Two types of ONPAR items were built in the science study to investigate 'how low could we go' in reducing the language. One version will be used in the mathematics research. The low language (LL) items use simple, sentence-level prompts. If the student so requests, L1 or English audio translations assist student comprehension of the item prompt. The very low language (VL) versions use simple, phrasal-based

language prompts and avoid L1 translations. The language on both the ONPAR versions is supported on the computer through rollovers of concepts and sentences that offer pictures and animations to explain meaning. The LL version used a speaker icon that spoke the concept or verb phrase in L1 or English (as chosen previously by the student). A third support was an animated icon that demonstrated how the student should provide a response (e.g., a graph line that moves, showing that the student should anticipate where the graph line should be). The items were analyzed for their behavior with the various ELL groups and were judged by an expert panel for content coverage and for comparability to their traditional item models. Discourse analysis[1] of the traditional and ONPAR items also was undertaken.

Besides building the items, the ONPAR study asked whether the items could be used effectively for low-English-proficient students instead of the traditional statewide tests, and if they could meet the technical standards of the large-scale tests so the scores could be considered comparable to native English students taking the regular test. To complete this portion of ONPAR, a series of cognitive labs were conducted and randomized experimental large-scale studies were scheduled in science and in math at the elementary and middle-school grade levels. The science study was completed in 2008-09; the math study will be conducted this fall. Both investigations are looking at how ELLs at different levels of language acquisition perform compared to nonELLs on both the traditional and ONPAR items. The math project will study how well some other students with language challenges, such as students with learning disabilities in reading, and deaf and hard-of-hearing students, are performing on both sets of items. Expert judgments and statistical analysis of the types performed on large-scale statewide tests are examining the results.

ONPAR item measuring buoyancy

This item determines the relative position of objects in water and the resulting water displacement based on the object's density and volume (go to www.onpar.us/buoyancy.html for more information on this item). Students view an animation showing three balls placed on a platform suspended over beakers; the platform is removed. First (Figure 1), students roll over the balls to determine that the metal balls are solid and the wood ball is hollow. They move the balls up and down to show their relative position in the water when the board is taken away and the balls drop into the water.

Next (Figure 2), students drag the water level to a position reflecting

Figure 1. What will happen to balls in water?
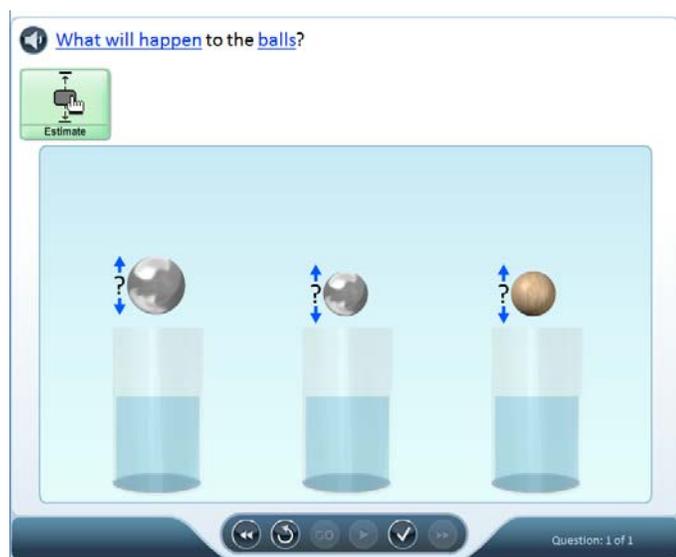
Figure 2. What will happen to the water?
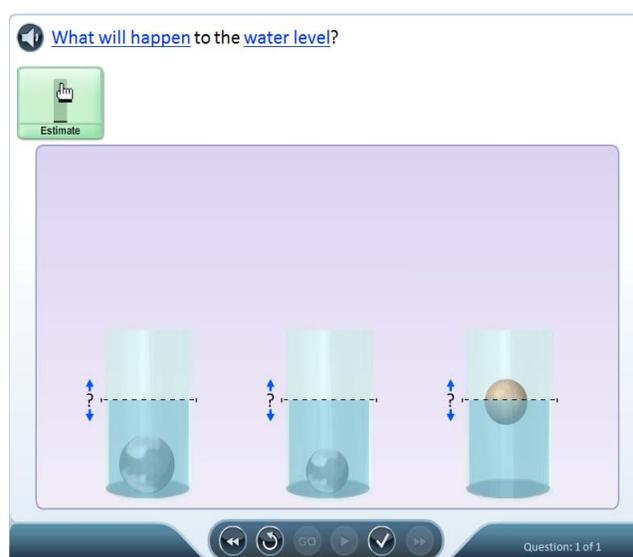
the position of the balls as they placed them in Figure 1.

ONPAR approach: In the first scene, students must compare the properties of each of the balls, determining that density, not size, determines where the balls will go in the water. Students compare wood and metal of the same size and metal of different sizes. In the second scene, students demonstrate knowledge about water displacement. Students' answers from the first scene carry to the second scene to compare relative water displacement.

Traditional item approach: In this item, students are asked to compare two steel balls of different sizes, indicate which water level will be the highest, provide an explanation, then compare a wood ball and steel ball of the same size, indicating which water level would be the highest, and provide an explanation. This item requires extensive language to explain the problem, and it requires students to produce language to respond.

Comparison: The ONPAR item asks students to interact with the screen elements and engage in the experiment, as compared to their more indirect relationship with the content in the traditional item. In ONPAR, the students are *demonstrating* their conceptual mastery, maintaining a depth of knowledge for the subtle comparisons based on several factors and demonstrating knowledge of cause-and-effect relationships. The traditional item asks students to explain but, depending on their

meta-cognitive abilities and their proficiency with language, their responses may or may not represent the true sophistication of their knowledge.

Analysis of ONPAR science items
Research on prototype items focused on discourse analysis, cognitive lab results, and the comparability of the computer interactive assessment to a traditional paper-and-pencil test as well as the comparability of specific items on the computer-based assessment to "matching" items on the paper-and-pencil test. The controlled experimental study provides a final look at the "goodness" of the prototype items in measuring the skills and knowledge of 4th and 8th grade students—it provides a first measure of the success of the ONPAR-Science project.

For the science study, three forms of the assessment (traditional, LL ONPAR, and VL ONPAR) were randomized over students. The traditional paper-and-pencil multiple choice and constructed response items were generally from the New England Common Assessment Program (NECAP), the National Assessment of Educational Progress (NAEP), and the Trends in International Math and Science Study (TIMSS). The study was guided by three research questions:
1. When controlled for ability, how does the performance of each group on the LL and VL level test forms compare to their performance on the traditional test form?
2. How does the focal group, the ELLs with low English profi-

ciency, perform relative to nonELLs?
3. What ONPAR item characteristic(s) appear to be effective or not effective?

Approximately 1,000 students from eight districts in three states, grades 4 and 8, participated in the study. ELLs at English proficiency levels 1-3, based on the ACCESS for ELLs™ English Proficiency Test, were the focal group, ELLs with proficiency levels 4 and above were an exploratory group, and nonELLs were the control group.

Most of the ONPAR science items were measuring the same content as the traditional test items from NECAP, NAEP, and TIMSS and the overall cognitive complexity was the same on both the ONPAR prototype items and the traditional assessment items. The students were tested in groups of about 15, with a team of two "testers," and items on a laptop.

For study purposes, the ability of the students was controlled statistically, based on a survey that teachers completed about the science skills of each student. The survey listed each of the concepts measured by the items on the "test" and asked teachers to provide the extent to which the students had demonstrated they understood the concept. Teachers' responses were recorded on a 4-point Likert-type scale.

Results are very promising. While there were significant differences between how LL ELLs and non-ELLs performed on the traditional test, there was *no* significant

difference between the LL ELLs and the nonELLs on the ONPAR test. Furthermore, the nonELLs did not score significantly differently on the two forms of the test, whereas the ELLs did. All of this suggests that the test scores appear to be comparable for the two groups of students, and that tests which use items such as those used in ONPAR can be used reliably and validly by low ELLs on a statewide test while other groups take the regular test.

Additional findings include (1) the scores of the higher level ELLs tended to be somewhere between both groups on both kinds of tests; (2) on the whole, the VL ONPAR test form did more poorly than the LL test form; and (3) an analysis of item characteristics showed that ONPAR items seem to pick up the science ability of students better than the traditional items do.

Conclusions

The collaboration with and cooperation of both state education agencies and local education agencies has been essential. They have assisted in assuring that the ONPAR items are aligned with state content and language proficiency standards, allowed access to items from the state content assessments, and identified schools and students to participate in the research.

There are both advantages and disadvantages to created assessments such as those being developed by ONPAR. The ONPAR approach has advantages over the traditional approach because it:

- Allows creating more equitable items for ELLs (and perhaps other populations);
- Aligns better with inquiry-based and interactive instructional approaches;
- Is more motivating for test-takers; and
- Offers the possibility of embedding and integrating accommodations for test-takers.

The disadvantages of an ONPAR-type approach include:

- Higher cost of item construction than for traditional items; and
- A technical infrastructure for testing which is still not universally available.

Research on ONPAR prototype items has shown that the ONPAR approach allows ELLs a better opportunity to demonstrate their science ability and thus is a more linguistically appropriate assessment tool for this population of learners. ONPAR-Science is an ongoing project; as more prototype items are developed and field tests of the science items continue, updates will be available at www.onpar.us.

Notes

1. For this study, discourse analysis refers to analyses of test items in order to understand item difficulty and accessibility better—especially for differentiated groups of test takers. The model is designed to predict the ovrall success of individual test items based on the presence or absence of specific discourse components.

Rebecca Kopriva, Ph.D., is a senior scientist at the Wisconsin Center for Educational Research (WCER), University of Wisconsin-Madison. E-mail: rkopriva@wisc.edu

---

ask*NCELA*'s Inbox

in which we highlight the answers to commonly asked questions that appear in our email inbox.

Q: Where can I find information on ELL students and services in my state?

A: NCELA's newly launched Title III State Information System has a wealth of up-to-date information on ELLs and Title III, by state. You can find information on English language proficiency standards and assessments, standards and assessments in the content areas, demographic information on ELL students, and a list of technical assistance centers tailored to each state. Visit the Title III State Information System often as NCELA adds to the information available: http://www.ncela.gwu.edu/t3sis

askNCELA@gwu.edu is NCELA's email helpline. We are happy to answer questions and to provide technical assistance information upon request.