Lessons Learned from Students about Assessment and Instruction

Joey, can you tell me what you were thinking when you were looking at the diagram?

“It looks like the keel of a boat.”

There comes a moment in every teacher’s experience when you ask a student a question and his or her answer is not what you expected—startling, perhaps. Such was our experience as researchers when we asked fourth graders to work on the task in figure 1 and talk to us about it.

Surprisingly, Joey was not seeing two figures at all. Instead, his attention was focused on the space between the figures. This is an example of the lessons we learned from students about this task, one of the released items from the 1996 Fourth-Grade National Assessment of Educational Progress (NAEP). The task is scored with a five-point rubric, based on the number of likenesses and differences between the shapes that students identify in their responses (see fig. 2).

The purpose of our work was to analyze the alignment between the task and certain benchmarks and standards, and whether the task elicits student thinking that matches the quality of the learning goals. While performing our analysis, we learned that good alignment with a standard does not necessarily mean that the task is of high quality. As this article describes, careful analysis of student interviews and papers showed us certain flaws in the task, even though the task was highly aligned with a Standard from NCTM’s Principles and Standards for School Mathematics (2000).

Alignment of Assessment Tasks with Learning Standards

In this era of educational reform and accountability, many states, districts, and schools are attempting to develop new curricula and assessment instruments to determine the extent to which students are meeting rigorous academic standards. The issue of alignment is becoming central to these developments. Policymakers, educators, and journalists have called for better alignment between standards and assessments (NCTM 1995; Quality Counts 2001; Romberg and Wilson 1995; Romberg et al. 1992). Alignment in the context of assessment usually means the degree to which a test (or test item) assesses the same learning goals as a given standard or set of standards (Wilson and Kenney 2001). Because of the high-stakes nature of assessment as a tool for educational reform,
issues of alignment have taken on increased immediacy. Calls for alignment have been addressed in varied ways (see LaMarca et al. 2000; Webb 1999). Researchers have initiated several projects to evaluate the alignment of an entire test with a given set of standards. This study, in contrast, took a more narrow and in-depth approach to the alignment of a single assessment item with one or more learning standards.

The task we chose to analyze is a fourth-grade released item from the 1996 National Assessment of Educational Progress (NAEP). An assessment task in this study is the test item along with accompanying materials, such as a scoring rubric or sample student work. The task selected is an extended constructed-response item. Students are asked to examine two figures—a rectangle and a parallelogram of the same length and width, drawn on a grid—and to describe all the ways the figures are alike and all the ways they are different. We selected this task for analysis primarily because data was available about how students scored on the task across the nation. We also had access to a national random sample of 237 student solutions for the task. To learn how students made sense of the task, in the spring of 2000 we conducted individual interviews with thirteen students from two geographically and demographically diverse areas.

In the interviews, students completed the task and we asked them to verbally explain their solution. We completed each interview by asking the students to tell us how they would describe the two figures on the phone to a friend who could not view the figures.

Before examining the student data, we first applied an alignment procedure developed by Project 2061 of the American Association for the Advancement of Science (see www.project2061.org). After completing our analysis individually, we combined our results to reach consensus. Our analysis clearly demonstrated that the Geometry Standard for Grades 3–5 from Principles and Standards for School Mathematics (NCTM 2000) is the most closely aligned to the task (see Wilson and Kitchen 2002 for a full summary of the completed analysis):

Identify, compare, and analyze attributes of two- and three-dimensional shapes and develop vocabulary to describe the attributes (p. 164)

We investigated the Standard in two parts: (1) whether students identified, compared, and analyzed the attributes of the shapes in the task; and (2) whether students used geometric vocabulary to describe the attributes of the shapes.

Figure 1
Fourth-grade NAEP item

Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all your work.

In what ways are the figures above alike? List as many ways as you can.

In what ways are the figures above different? List as many ways as you can.

In the interviews, students completed the task and we asked them to verbally explain their solution. We completed each interview by asking the students to tell us how they would describe the two figures on the phone to a friend who could not view the figures.

Before examining the student data, we first applied an alignment procedure developed by Project 2061 of the American Association for the Advancement of Science (see www.project2061.org). After completing our analysis individually, we combined our results to reach consensus. Our analysis clearly demonstrated that the Geometry Standard for Grades 3–5 from Principles and Standards for School Mathematics (NCTM 2000) is the most closely aligned to the task (see Wilson and Kitchen 2002 for a full summary of the completed analysis):

Identify, compare, and analyze attributes of two- and three-dimensional shapes and develop vocabulary to describe the attributes (p. 164)

We investigated the Standard in two parts: (1) whether students identified, compared, and analyzed the attributes of the shapes in the task; and (2) whether students used geometric vocabulary to describe the attributes of the shapes.
What We Learned from the Students

To complete the task, we asked students to use paper and pencil to write about how the two figures are similar and different. In the interviews, we asked the students when they knew to stop working on the task. Nearly all of them responded that they stopped when they could no longer think of anything to write. One boy seemed intent on writing the same number of likenesses and differences. One of us observed that the students also seemed to be influenced by the amount of space provided on the page for the answer; they decided to stop writing when the space for each question was filled. One of the flaws of this task, when it is used as part of a written assessment, is that students have no way of knowing when they have satisfactorily finished the task, so they must rely on other clues.

The data from the 237 student papers and thirteen student interviews demonstrated that the majority of students (64 percent of the student papers) showed evidence of thinking about attributes of the figures, but most (70 percent of the student papers) did not use geometric vocabulary correctly in their responses. We ascribe students’ poor use of geometric vocabulary to the task’s failure to require the use of geometric vocabulary and to students’ poor understanding of measurement concepts, particularly perimeter and congruence.

Because the expectations of the task were ambiguous, students often provided solutions that were nonmathematical. For example, a few students referred to what the shapes reminded them of as a way of thinking about the figures.

Figure 2

Fourth-grade NAEP item scoring guide

Solution
The figures are alike because:

a. They both have 4 sides (or 4 corners or 4 angles).
b. They both have parallel sides.
c. They both have two sets of sides that are the same length.
d. They have the same area.
e. They have the same length (base).
f. They have the same height.
g. They both have little squares.

The figures are different because (do not accept “They’re not both the same shape”):

h. One has 4 equal angles and the other does not.
i. One has right angles or perpendicular lines and the other does not. (Students do not need to make the comparison; that is, they can just say, “One has 4 right angles.”)
j. One is “slantier” than the other.
k. They have different perimeters.

Scoring Guide
1. Incorrect response
2. Student gives a nonspecific response such as “The one on the right is skinnier” OR only one correct reason (alike or different)
3. Student gives one correct reason alike and one correct reason different OR two reasons alike OR two reasons different
4. Student gives two reasons the figures are alike and one reason they are different OR one reason they are alike and two reasons they are different
5. Student gives at least two reasons they are alike and at least two reasons they are different (two alike reasons are not both “a”)
of, such as “the back of a truck” (the rectangle) and “carrots that my cousin chopped diagonally” (the parallelogram). This demonstrated that what sort of solutions would count was unclear to students. They had to infer from the context (“This task is on a mathematics test”) to know that they probably should provide a mathematical answer.

When students did incorporate correct geometric vocabulary in their responses, often the only geometric word they used was squares. For example, during an interview, one student wrote that “they [the figures] are both five squares across and three squares down.” Another wrote, “The two shapes both have 15 complete squares [sic].” and a third wrote, “They both have 3 squares [sic] in each column.” Two students misunderstood both shapes to be squares. One of these students discussed how the shapes looked like squares, and the other wrote that “the squares are the same but 1 is tilted a little.”

While reviewing the student papers, we gave many students a score of four out of five, although their responses lacked the use of correct geometric vocabulary. Students’ only clue that the task entails mathematical language is that it is part of a mathematics test. Nothing about the task design encourages students to use geometric language, nor does the scoring rubric reward students who do so. We do not mean to suggest that students should be punished for not using correct geometric vocabulary; our point is only that this task did not instruct students to use precise mathematical language.

We also found that students struggled to use geometric vocabulary to discuss the perpendicular sides in the rectangle and the nonperpendicular sides in the parallelogram. Students used words such as straight or up and down as substitutes for perpendicular, and angled or slanted to talk about the angles of the parallelogram. None of the students we interviewed knew the word parallelogram. Many students began their work by numbering the two shapes so that they could refer to them as “Figure 1” and “Figure 2.” A few correctly identified the first figure as a rectangle.

When we gave the students an opportunity to talk about the task by describing the shapes to us in a phone conversation, some students identified additional attributes of the figures. They began to notice the angle measures that were different in the two figures, and one student began to see the parallelogram as a rectangle plus a right triangle. These findings show that given the opportunity to respond in ways other than simply with pencil and paper, students may be capable of communicating more profound mathematical understandings than they would in writing.

In summary, we found that even though the task was highly aligned with the Geometry Standard, it was not effective at prompting students to use correct geometric vocabulary. This finding suggests that alignment of a test or test item with a set of learning goals is necessary but not sufficient to ensure that the test or test item is of high quality. For a test or test item to be of high quality, it not only must be aligned with a set of high-quality learning goals but also must elicit student thinking that is aligned with the learning goals. That the designers of NAEP were interested only in whether students could identify, compare, and analyze the shapes is quite possible. For this study, the purpose of the task on the NAEP is unimportant. Our original intent was to judge the effectiveness of the Project 2061 procedure by analyzing the alignment of the task with select standards. After applying the alignment procedure, we discovered that student responses did not match the second part of the learning goal that aligned best with the task. Although this finding may not be of significance to the NAEP designers, it illustrated for us that alignment of a task with standards does not ensure high-quality student responses.

Alignment of a task with standards does not ensure high-quality student responses

How Our Analysis Informs Instruction

Because of the many issues we have discussed and the task’s potential to generate interesting and rich mathematical discourse, we believe that this task may be best suited for classroom use. Working in small groups, students initially could use rulers and protractors to analyze the shapes before engaging in a whole-class discussion of their findings. In the classroom, the teacher could both guide and prompt students to find similarities and differences between the figures while supporting their use of correct geometric vocabulary.
With the teacher’s guidance, students also could learn how to distinguish between responses that are appearance-based (that is, have informal, visual, or descriptive characteristics), attribute-based, and based on seeing the figures as members of a class (Lehrer, Jenkins, and Osana 1998). In terms of the van Hiele levels of geometric thinking (van Hiele 1986), this entails helping students move from level 1, the visual stage at which students characterize shapes by their appearances, to level 2, the descriptive/analytic stage at which students recognize and can characterize shapes by their properties. The majority of the students’ papers and responses in our study were appearance-based and attribute-based. Teachers could use this task as an activity or as part of a series of lessons in which students focus on the properties of figures to learn how to identify them as members of a class.

Our analysis of student work samples also indicated that students may not be accustomed to analyzing figures to discover their differences. For students to provide more ways in which the figures are similar than dissimilar was common. This finding indicates that students probably are more accustomed to analyzing figures for their similarities than for their differences. Although this finding may be specific only to this task, we believe that students generally need more experience with investigating how figures are different.

In addition, even students who completed the task and were identified by their teachers as high-achieving were incapable of writing about how the figures are similar and different because they lacked the vocabulary to do so. For us, the significance of this finding is that geometric vocabulary must be explicitly taught to students. Moreover, several of the teachers of the students who participated in the interviews stated that their students had not yet studied geometry prior to the interviews. Most likely, students’ lack of preparation in geometry contributed to their incorrect and inadequate use of geometric vocabulary and the difficulties that they had in describing how the figures are different. Principles and Standards for School Mathematics (NCTM 2000) recommends that students study geometry throughout the early elementary grades.

Concluding Remarks

Our study highlights the value of analyzing student work samples to understand whether a task elicits the level of student thinking stated in the aligned learning goal. In this case, we believe that the task we analyzed has certain strengths and flaws that make it better suited for classroom use than as part of an assessment instrument. Although the task actually was well aligned with a learning standard that we identified when carrying out the alignment procedure, the student data demonstrated that the task did a poor job of prompting students to use correct geometric vocabulary. When we told students to explain their thinking while completing the task, they were more likely to think of the shapes as members of a class of shapes and to use correct geometric vocabulary.

Alignment of a task with a learning goal or set of learning goals is necessary but not sufficient to ensure that the task is of high quality. This study demonstrates the value of analyzing multiple pieces of student work, as well as the importance of listening to students explain their thinking, in order to reveal information about the qualities of a mathematics task that an alignment procedure by itself may not reveal.

References


NCTM's 2004 Each One, Reach One Membership Drive

Join your colleagues as they help NCTM launch the 2004 Each One, Reach One membership campaign. If you—and each one of your NCTM colleagues encouraged just one mathematics educator to join, NCTM would double its strength!

Watch your mailbox for additional information about the campaign, which includes exciting prizes for you and a world of opportunity for professional development for your recruited member. Participating is fun, easy, and rewarding.