Performance Based Assessment vs. Traditional Assessments in the Science Classroom

Jim Mauch

Sed690

December 19, 2005

We have a wide variety of purposes for assessment in our classrooms. Policy makers may want to monitor the quality of education. Administrators could monitor program strengths and weaknesses, plan a program, or evaluate alternatives to existing curricula. Teachers use them to determine students' grades, monitor student progress, or motivate students (Dietel, Herman, & Knuth, 1991). The same type of assessment is not always the best answer, but my question I have chosen to debate is whether performance based assessments or more traditional assessments have greater value in the science classroom.

In the process of evaluating assessments used, it is important to remember that the validity of a test must be measured against the purpose the test was used (Dietel, Herman, & Knuth, 1991). A good test directly measures how well a student performs on the tasks that the test was intended to measure. Although this seems obvious, there are times when a test does not validly measure what is intended at all. So, the question which must be answered is which assessment most validly measures the performance of science students?

First, we should define what we are comparing. Although 'Performance Based Assessment' can mean different things to different groups, I will use the general description used by Kind (1999). He describes a performance based assessment as one that:

- Requires students to use higher level thinking skills.
- Places the problem within a real world context.
- Assesses the process and reasoning used to solve the problem as well as the solution attained.

Often, these will appear as labs or projects. There may be more than one right answer. Students may have to work alone, but in many instances, they will collaborate in groups. Students are required to develop a plan, collect and organize data, and formulate conclusions based on the evidence collected. This assessment will require students to incorporate information from many areas (Kind, 1999).

On the other side of this spectrum is a more traditional test. This gives questions or problems on a discreet concept, without context. These questions can be multiple choice (on standardized tests, this is most likely the format) or written answer. Students

nearly always work alone, without any collaboration. Higher level thinking skills may be used, but often to a lesser degree. The process used is rarely assessed in any real way, since often there is one right answer. Often, problem solving is involved with something practiced repetitively.

The Debate

. Dietel, Herman, & Knuth (1991) make the statement; 'Current evidence about the nature of learning makes it apparent that instruction which strongly emphasizes structured drill and practice on discrete, factual knowledge does students a major disservice.' As we change our attitudes about learning, we have put an increasing emphasis on strategies like inquiry learning. The student now plays a more active role and the lessons are placed in a more 'real world' context. Students are asked to explore and problem solve as scientists. More and more, we ask students to use higher level cognitive skills in the classroom. As teaching becomes more contextual, involving more active dynamics between students, teachers, and the community, the assessments should too (Fusco & Barton, 2001).

As this shift takes place, our assessments must shift from passive test takers to active participants, from discrete items on tests to cross disciplinary problems, from single test assessments to samples of work over a period of time, from individuals taking tests alone to collaborative group settings. Performance based assessments provide a perfect environment for a student to demonstrate his skill in scientific problem solving. Students must incorporate concepts from different fields. Traditional assessments, do not give the same opportunity for students to exercise their higher level thinking (Dietel, Herman, & Knuth, 1991).

Performance based assessments give a better opportunity for the evaluator to see how much the student understands about the concepts. A traditional test may never ask students to apply concepts in a new setting. Since a performance based test requires students to use knowledge from multiple disciplines, he must demonstrate an understanding as to how various discrete concepts relate.

If we examine the California standards, it is possible to imagine a series of discrete narrow facts that the students are expected to learn. I believe most teachers

would expect that the intent is a much greater level of understanding. Bos, Kulper, & Plomp (2001) analyzed results of the TIMMS tests in the Netherlands. They looked for possible causes for the lower scores on the performance tests than the written test. What was uncovered was a discrepancy between the intended curriculum and the implemented curriculum. The teachers were not expecting students to use the concepts taught in a way that required higher order thinking skills, yet that is what the writers of the curriculum had in mind. We may have the reverse situation in California. The STAR test does not require higher order thinking. In fact, there is very little, if any, problem solving of any kind on the test. We are not testing our students' ability to apply what they have learned, to perform investigative problem solving, yet I believe that many teachers would feel this is the intent of the standards.

Another common complaint about traditional assessments is that they show cultural biases which make them unfair for minority students. Placing the assessment within a real world context is seen as being a way around this problem (Fusco & Barton, 2001), (Dietel, Herman, & Knuth, 1991). By using a context that is meaningful to the students, they are more likely to grasp the concepts and may be more motivated. What is important is that the rubric for scoring the assessment takes into consideration the background (prior knowledge, personal experience, etc.) of the students.

A challenge and an advantage of performance assessment is that there is no single answer. A rubric must be made to score the assessment. If done well, the rubric will allow the evaluator to consider a great deal about the students understanding of science, his reasoning skills, as well as organization and planning. Kind (1999) discusses the rubric used in scoring the TIMMS assessments. Scores are given for plan design, data gathering, conclusion, explanation, and for changes made in the plan (as they learn information, do they adjust and can they explain their adjustments?). A great emphasis is placed on providing evidence supporting any conclusions. This becomes a powerful tool in evaluation the level of understanding of the students. Not just their understanding of the concepts involved, but also of the nature of science.

To illustrate the difference between a more traditional test and performance based test, I will take examples from the released California Standards test questions (CDE,

2005) and a performance assessment submitted by the Kentucky Department of Education (1992).

California Standards Test: acids and bases

1. Equal volumes of 1 molar hydrochloric acid (HCl) and 1 molar sodium hydroxide base (NaOH) are mixed> After mixing, the solutions will be
    a. Strongly acidic
    b. Weakly acidic
    c. Nearly neutral
    d. Weakly basic

2. Of four different laboratory solutions, the solution with the highest acidity has a pH of
    a. 11
    b. 7
    c. 5
    d. 3

.Kentucky Performance Test: This is a summary of the 3 page packet.

Problem: Design an experiment to test which of four antacids would be most effective for neutralizing acid.

Students are given 30 minutes to design and carry out their experiments in groups. They are given four antacids, an acid solution, pH paper, a 4 well tray, and stirring rods.

Students must write procedures, results, and discussion as a group.

Individually, students must prepare charts and graphs to illustrate their results. Plus analysis questions asking them how they determined the antacid effectiveness, how they would change the procedures if they were doing this by themselves, and how this information would be useful in everyday life.

. It is apparent to me that the level of thought necessary to accomplish the multiple choice test is fairly limited, whereas the performance test requires greater analytical thought. They both test the same standards, yet they really do not test the same skills. A person getting two right out of two on the California test knows a few facts. A person doing well on the performance test knows how to think scientifically, and they probably would have gotten the multiple choice right based on their experiences.

It is not my intent to state the shortcomings of the California Standards Test, but, I am certain that on any test, I would see the same level of questions as this. We would know more about our students' abilities and we would have far more meaningful data if California kept the Golden State exam and dropped the STAR.

Is there another side to this argument? If there weren't, then why is the state and districts using standard assessment? The arguments against performance based assessment are vague. If there is data saying that it isn't effective in testing higher level thinking skills, I have not found it. I will base my arguments on personal experience and discussions with cohort members.

Time is probably the big issue. Some of these assessments can take hours for the students to do. I have seen some that are ongoing projects lasting weeks. Most that I have seen are easily done in 2 hours or less. I do not believe it is necessary to make these assessments last weeks unless that is the intent of your curriculum.

The tests take longer to score also. Every test will be different and must be evaluated based on rubric. I am sure this is one reason that California gives a multiple choice assessment. Graders spent weeks on the Golden State Exam. I have not met any grader of the Golden State Exam who felt they had wasted that time. My argument to the state would be that since we are placing such high stakes on the results of this test, the investment involved in the difficulty grading them would be outweighed by having a more complete picture of our students performance. For the individual teacher, the time grading these should not be much different than grading any lab report or test (not multiple choice). Certainly there is room for introducing this type of assessment occasionally at least. So, still, the time is worth it.

For the individual teacher, another argument against is preparing the assessments. For this, I suggest looking at the PALS website. Performance Assessment Links in Science Has hundreds of assessments, pre planned with rubrics and teacher guides (Quellmalz, Hinojosa, Hinojosa, Schank, 2000). Many of these still require equipment preparation, but this is again no different than any other lab. Teachers should gradually introduce these into the curriculum so as to have time to evaluate their usefulness and fully incorporate them into the curriculum.

Although I have stated that these provide for inclusiveness for minorities, that does not mean that there can't be difficulties. A poorly designed rubric or an assessment that does not provide a relevant context can create bias instead of eliminating it (Dietel, Herman, & Knuth, 1991). As with any test, great care must be given to be sure it is fair.

The last argument against performance based assessments has to do with our present obsession with the standards. Since we have based so much on the California Science standards, the state apparently finds it necessary to be sure that all standards are tested every year. It seems perfectly reasonable to have a state test that does not test the same standards every year. The response would likely be that this might unfairly penalize a school with a program with one or two weak areas. A suggestion might be to average performances over a number of years, or, go back to what the Golden State Exam did, and offer a part of the test which is multiple choice in nature, but still involved in interpreting data, drawing conclusions, and applying knowledge to a situation.

Position

I strongly believe that performance based assessments are of great value to science education and should be used by the state, districts and individual teachers to evaluate students and curriculum. Although this is not a controversial position, we are heading away from this instead of toward it.

For the individual teacher, making performance based assessments shouldn't require much modification of labs or projects. Even in classrooms where many labs are very structured, I can see the possibility of progressing students to the point where these assessments would be effective and useful.

At the state or district level, performance assessments could be even more valuable. As we continue to attempt to improve science education compared to that in other countries, it makes sense to use assessments which are comparable to those used in international comparisons (Kind, 1999). In our eagerness to ensure the standards are met, we have forgotten the intent of the standards would be that students understand and apply those concepts to real situations.

Bibliography

Bos, K., Kuiper, W., & Plomp, T. (2001). TIMSS results of dutch grade 8 students in international perspective: Performance assessment and written test. *Studies in Educational Evaluation, 27*(1), 79-94. Retrieved December 18, 2005, from ERIC database.

California Department of Education (2005). Released *Questions California Standards Test Chemistry*. http://www.cde.ca.gov/ta/tg/sr/documents/rtqchem.pdf

Dietel, R. J., Herman, J. L., & Knuth, R. A. (1991). *What does research say about assessment?* [On-line] Available: http://www.ncrel.org/sdrs/areas/stw_esys/4assess.htm

Fusco, D., & Barton, A. C. (2001). Representing student achievements in science. *Journal of Research in Science Teaching, 38*(3), 337-354. Retrieved December 18, 2005, from ERIC database.

Kentucky Department of Education (1992) *The Effectiveness of Antacids*. Performance Assessment Links in Science Website. Available on line: http://pals.sri.com/tasks/9-12/Antacids/

Kerins, C. T. (2000). *Performance assessment on a platter--but will teachers partake?*. U.S.; Illinois: December 18, 2005, from ERIC database.

Kind, P. M. (1999). Performance assessment in science--what are we measuring? *Studies in Educational Evaluation, 25*(3), 179-194. Retrieved December 18, 2005, from ERIC database.

Quellmalz, E., Hinojosa, T., Hinojosa, L., Schank, P. (2000). PALS Final Project Report Available on line: http://pals.sri.com/papers/finalreport/