Effects of Classroom Assessment on Student Motivation in Fifth-Grade Science

CANDICE STEFANOU
Bucknell University

JAY PARKES
University of New Mexico

ABSTRACT Researchers have suggested that assessment has the potential to affect learner behavior in terms of cognitive strategy use and motivation. The authors attempted to provide an understanding of the nature of the effect of particular assessment types on motivation. Students in 3 5th-grade science classes were exposed to 3 different classroom assessment conditions: traditional paper-and-pencil tests, a laboratory task format of assessment, and a performance assessment. Measures of attitudes about science, goal orientation, and cognitive engagement (J. L. Meece, P. C. Blumenfeld, & R. K. Hoyle, 1988) were used. Analyses indicate a significant effect attributable to assessment type on goal orientation only, with the traditional paper-and-pencil tests and the performance assessments fostering more task-focused orientations than the laboratory tests.

Key words: classroom assessment, fifth-grade science, student motivation

Dissatisfaction with large-scale, high-stakes multiple-choice testing in the 1980s led many educators to focus attention on the format of the assessments that students were taking. Although some educators called for abolishing the use of standardized tests (e.g., Kohn, 2000), others rallied around the “tail wagging the dog” argument acknowledging that large-scale, high-stakes tests were not going away and that a better tactic might be to co-opt high-stakes tests and make those measures “tests worth taking” (Wiggins, 1992). As a result, research was done and conjecture made on the systemic impact of educators’ using performance assessments for large-scale, high-stakes assessment (e.g., Khatrri, Reeve, & Kane, 1998).

Part of the impetus for the performance-assessment movement was the grassroots work of classroom teachers in assessing their students through projects, portfolios, and the like. There were many claims that the same benefits seen (but rarely documented) at the classroom level would accrue when these assessment formats were transferred to the large-scale, high-stakes level (cf. Wiggins, 1989). Few of those claims for large-scale, high-stakes use were empirically explored (Bond, 1995).

The goal of this study was to explore one claim for performance assessments: that they foster a more positive motivational orientation for students than supplied-response assessments do when used in the classroom. There is more than anecdotal experience to expect them to do so; there is a strong theoretical underpinning to support this argument, assuming that certain design elements are in place. In this study, we explored the effects of different classroom assessment formats on motivation by using upper elementary science classes as the context.

Systemic Impact of Standardized Testing

Because the framework of interest in this study centers on the consequences of test use (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999; Messick, 1994), the importance of consequences to valid inferences and uses of assessment results is explored through a review of studies that document the consequences of large-scale, high-stakes tests.

Studies of Consequences

The consequences of a particular assessment are important evidence of validity (AERA, APA, NCME, 1999; Messick, 1994; Shepard, 1997) and are especially important with reference to performance assessments (e.g., Shepard, 1989; Wiggins, 1989, 1992). Exploring the consequences of a particular assessment procedure is in essence a cost–benefit analysis. That is, there are intended positive benefits to be gained from a particular assessment, although there are undoubtedly also some unforeseen and sometimes detrimental outcomes. The role of an analysis of consequences is to determine whether the positive consequences (intended and unintended) outweigh the negative consequences.

Address correspondence to Candice Stefanou, 465 Olin Hall, Bucknell University, Lewisburg, PA 17837, (E-mail: cstefano@bucknell.edu)
This type of analysis is not new in the performance-assessment debate. Although test designers have acknowledged some of the benefits of using performance assessments in large-scale, high-stakes settings, the costs for doing so in dollars and in time (among other costs) have made the use of such assessments prohibitive (cf. Hardy, 1996).

A body of research focuses on the costs or detrimental effects of standardized testing on teaching practices and learning strategies (e.g., Darling-Hammond, 1990; Shepard, 1991). The argument in favor of alternative assessments as a tool for reform rests in large part on the results of studies that have been conducted to assess the effects of traditional standardized testing on instruction and achievement (e.g., Herman, 1992; Herman & Golan, 1993; McNeil & Valenzuela, 2000; Monty Neill & Medina, 1989; Smith & Rottenberg, 1991). Such studies indicate that there are a number of unintended, negative consequences as a result of an increasing overreliance on standardized assessments. Included among these adverse consequences are (a) large portions of instructional time devoted to preparing students to take the standardized tests; (b) narrowing of the curriculum to include only that which is assessed by the tests; (c) fragmentation of the curriculum, leading to an inert knowledge base for the students; (d) evolution of instructional practices, including teacher-made tests, to mimic the types of learning skills that dominate standardized tests; and (e) truncation of students’ learning strategies, resulting in the use of primarily lower order learning skills.

Research provides support for some of the claims about standardized testing. For example, in a survey of nearly 1,000 students in elementary school through high school, Paris, Lawton, Turner, and Roth (1991) asked students about standardized tests. As they get older, students feel “greater resentment, anxiety, cynicism, and mistrust of standardized achievement tests” (p. 16). The authors reported that students seem to get the message that the tests do not count for anything—they have no impact on grades, teachers and parents do not care about an individual student’s performance, and the tests are unrelated to what is going on in the classroom. Overall, Paris et al. reported a decrease in motivation to perform well on standardized achievement tests as students get older.

A. H. Karmos and Karmos (1984) found similar results by surveying students in Grades 6–9. In general, they found that students tended to react negatively to standardized achievement tests. A large percentage of students responded positively to statements that the tests were dumb or a waste of time. However, the authors found that students who believed that achievement tests were important and who had higher perceptions of self-efficacy did better on the tests than students who reacted negatively to the tests.

Claims About Performance Assessment

Advocates of performance assessments argue, in light of what is believed about the perceived powerful effects of the ability of standardized assessments to drive instruction, that the form of an assessment creates the conditions for the type and quality of instructional practices engaged in by educators and learning strategies and behaviors exhibited by students. The arguments offered to demonstrate the advantages of performance assessments over more traditional assessments are that performance assessments (a) are more authentic assessments; (b) use context, thus diminishing the potential overreliance on inert knowledge; (c) encourage higher order thinking rather than memorization and recall; (d) are more engaging and therefore more motivating for the student; and (e) are more valid assessments, specifically more systemically valid (Frederiksen & Collins, 1989). Wiggins (1989) argued that standardized tests rarely inform students, often tap irrelevant information, and are seldom meaningful to students. He suggested that performance assessments not only inform others about the test taker but also inform the test taker about the important issues and topics within a given field.

Furthermore, the idea that performance assessments may encourage higher order thinking (e.g., Frederiksen, 1984; Resnick, 1987; Resnick & Resnick, 1989; Shepard, 2000; Wiggins, 1992) and may be more engaging and more motivating (e.g., Meisels, Dorfman, & Steele, 1995; Wiggins, 1989) than traditional tests leads one to investigate the hypothesized relationship. The criticisms of standardized testing and advocacy for performance assessments have been generalized to the format of the tests, that is, multiple-choice items versus open-ended tasks. At this point in the understanding of the effects of testing, it is unclear whether the environment surrounding the test format (e.g., the imposed testing for accountability purposes) or the format itself is more problematic.

Performance Assessment in the Classroom

There is some supportive evidence for the claims made regarding performance assessments in the classroom. Lee (1994) found that college students have a tendency to use different study strategies depending on the format of the test to be taken; deep-level processing strategies are associated with performance assessments, and surface-level strategies are associated with traditional paper-and-pencil tests. Lu and Suen (1995) examined the effects of field dependence and field independence on performance-assessment scores and multiple-choice test scores in a college classroom. They found that there was an interaction between cognitive style and assessment type. Their results showed that field-independent students scored higher on the performance assessment than did field-dependent students, whereas no difference existed on the multiple-choice scores. This finding suggests that performance assessments may require a set of filtering strategies different from those that are required for traditional assessments.

Parkes (2000) explored the relationship between goal orientation and scores on both an objective measure and a per-
formance assessment. High school introductory Spanish students were given an objective test of grammar, vocabulary, and idioms and a performance assessment requiring a written paragraph about a day of sightseeing in Spain. Both tests covered the same text chapters. The paragraphs were scored for verb forms, vocabulary, coherence, and creativity. The students also completed a survey of their perceived control. Parkes found that a student’s perceived internal control correlated significantly with performance-assessment scores but did not correlate with objective-test scores. Elliot, McGregor, and Gable (1999) similarly found that performance goals (avoidant and approach) were related to an examination score but mastery goals were not. Furthermore, when Parkes (2000) used objective-test scores to predict performance-assessment scores, perceptions of internal control explained variance beyond that predicted by the knowledge represented on the objective test.

Pintrich and DeGroot (1990) explored the relationships among self-efficacy, test anxiety, intrinsic value, cognitive strategy use, and self-regulation and seventh graders’ performance on seatwork, examinations and quizzes, and essays and reports. Regression analyses revealed a different set of relationships when one considered examinations and quizzes versus essays and reports. In the authors’ investigation of examinations and quizzes, (a) test anxiety showed a negative relationship; (b) self-regulation was positively related; and (c) cognitive strategy use, self-efficacy, and intrinsic value were unrelated. For the essays and reports, there was no relationship with self-efficacy, test anxiety, or intrinsic value; levels of cognitive strategy use was negatively related and self-regulation positively related. Studies such as these provide support for the claims that different psychological processes operate under the different assessment formats.

Some researchers (e.g., Meisels, Dorfman, & Steele, 1995; Wiggins, 1989) have claimed that performance assessments will be more motivating for students than other forms of assessment. There are characteristics of performance assessments that should foster positive goal orientations in students, especially when these characteristics are examined from the perspective of the achievement goal theoretic framework.

Motivational Climate and Orientation

Goal theory posits that persons enter achievement situations with different orientations. Generally speaking, a person might exhibit mastery–learning goals or performance goals (both approach and avoidant). Although slightly different names have been used in the literature, the underlying concepts are the same (cf. Ames & Archer, 1988; Midgley, Kaplan, & Middleton, 2001).

Students with a mastery–learning goal orientation are concerned with increasing their competence and believe that effort is a key mechanism in achieving mastery. Work-
individual improvement, mastery, effort, and making mistakes as part of growth (Ames, 1992).

Motivational Aspects of Performance Assessments

Blumenfeld (1992) determined that performance assessments are good candidates for research on mastery goals because of their characteristics. According to Baker, O’Neill, and Linn (1993, p. 1211), performance assessments (a) are open ended, (b) focus on higher order or complex skills, (c) use context sensitive strategies, (d) use complex problems requiring several types of performance and significant student time, (e) may require group as well as individual performance, and (f) allow for a significant degree of student choice. Given the goal theory perspective, and especially Ames’s (1992) articulation of instructional practices that foster mastery orientations, these characteristics should help with that process.

Method

Participants

Seventy-nine students in three 5th-grade science classes participated in this study. All three classes were in the same rural, northeastern U.S. elementary school. The classes were heterogeneous in terms of ability levels and gender: the actual classroom roster was determined in the previous academic year by a team consisting of school administrators and faculty. All three classes were taught by the same teacher. The sample consisted of 46 (58%) boys, 28 (35%) girls, and 10 (6%) students whose sex was not indicated. All students were either 10 or 11 years old.

Procedure

The study took place during the last half of the school year (roughly January to June) and focused on three different instructional units: (a) a unit on fresh water, (b) a unit on salt water, and (c) a special unit on The Voyage of the Mimi, a commercially available interdisciplinary unit (Bank Street College of Education, 1990). Each form of assessment was developed for each unit. The assessment type was then counterbalanced across the three classes of students so that at each unit, one class took each type of assessment, and across the three units, each class received each type of assessment. After each assessment had been administered, we asked the students to complete the questionnaire about their attitudes toward science, goals, and cognitive engagement during the preceding assessment session.

Instruments

In this study, we used portions of the Science Activity Questionnaire (reported in Meece, Blumenfeld, & Hoyle, 1988), specifically the Science Attitudes, Goal Orientations, and Cognitive Engagement subscales. The Science Attitudes subscale consisted of 12 items with observed Cronbach’s alpha of .76 when administered after the first instructional unit, .86 when administered after the second instructional unit, and .87 when administered after the third instructional unit. Higher scores on the Science Attitudes subscale indicate more positive experiences with science. The Goal Orientations subscale consisted of 12 items with observed Cronbach’s alphas after each instructional unit of .56, .36, and .62, respectively. Some items on the Goal Orientations subscale require reverse scoring. When the subscale is scored with items given the appropriate score and given the direction of the reverse scoring, higher scores indicate more mastery-learning goal orientations. The Cognitive Engagement subscale consisted of 15 items with observed Cronbach’s alphas after each instructional unit of .55, .71, and .74, respectively. Some items on the Cognitive Engagement subscale required reverse scoring. When the subscale was scored with items given the appropriate score and given the direction of the reverse scoring, higher scores indicated more self-regulated and engaged behaviors.

Students were assessed through traditional paper-and-pencil (P&P) tests, traditional science laboratories (lab), and performance assessments (PA). The traditional P&P tests were designed by the classroom teacher and consisted of multiple-choice, true–false, matching, fill-in-the-blank, and concept definition essay questions. The teacher designed the traditional P&P tests to capture the conceptual knowledge apart from the student’s ability to put that knowledge to use. We designed the laboratory assessments and the PAs jointly with the classroom teacher (see Appendix for a sample P&P test, laboratory assessment, and PA). The laboratory assessments were designed along the lines of typical laboratory experiments that require students to actively manipulate materials and reach conclusions based on observations. The steps that the students needed to take were clearly defined, and the observations the students needed to make were specified. The intent of the laboratory assessment was to serve as an assessment type that falls in between the traditional P&P test and the more open-ended PA. We designed PAs by using the attributes of performance-based assessments delineated by Baker, O’Neill, and Linn (1993) and reviewed earlier. In the sample laboratory assessment and PA provided in the Appendix, we illustrated the differences between the 2 items using the same subject-matter platform.

In addition to the quantitative data, we obtained qualitative data in the form of taped interviews with each of the fifth-grade science classes several days after the third instructional unit. All three interview sessions were conducted in the classroom with the entire class participating. The sessions were conducted on the same day in three consecutive class periods. We asked the following questions: (a) Having experienced all three assessment types, which type of assessment did the students prefer? (b) How could the PA be made better; that is, what recommendations
would the students have for making these tests "tests worth taking" (as they have been called by Wiggins, 1992)? And following conversation during the interviews, we asked, (c) If grades were not a factor, would the students' preferences change? We collected qualitative data to assist in gaining a better understanding of the interplay between assessment type and student motivation than what might have been captured by the inventory.

Analyses

Any effect on the dependent variables (attitudes, goal orientations, and cognitive engagement) caused by time of year and topic of the unit should have been negated by the counterbalance. Likewise, the counterbalance combined both between-subjects and within-subjects comparisons across different assessment types. Therefore, the data were collapsed so that one multivariate analysis of variance was performed with the three dependent variables (attitudes, goal orientations, and cognitive engagement) and assessment type, the one within-subjects independent variable. Some students were lost because of the repeated measures design. Students were eliminated from the study if they missed either the PA test or the laboratory test on the scheduled test date. Although students could easily make up the traditional P&P test by taking it on another day, they could not make up the PA or laboratory tests in this way because the tests were designed to be carried out in groups. Therefore, once a student missed either the PA or the laboratory test, he or she was automatically given the traditional P&P test as the make-up test and was accordingly excluded from the analysis. This left an operational sample of 57 students who each had one traditional P&P test, one PA, and one laboratory test.

Results

There was a statistically significant multivariate effect (Wilks's $\lambda = .769$). $F(6, 51) = 2.551, p < .05$. Follow-up univariate $F$ tests revealed no statistically significant effect on science attitudes, $F(2, 112) = 2.557, p > .05$, or cognitive engagement, $F(2, 112) = 0.922, p > .05$. There was, however, a statistically significant effect for goal orientation, $F(2, 112) = 6.013, p < .05$. Using the Bonferroni multiple-comparisons procedure for pairwise comparisons, we found that the mean differences of goal orientation were statistically significantly higher in the P&P and PA conditions than in the Lab condition, although there was no statistically significant difference between the PA and P&P conditions. The respective means are reported in Table 1.

Qualitative analyses revealed that the students were concerned primarily with grades, and their preference for assessment type was overwhelmingly in favor of P&P, the form that they knew best and that they believed would help them achieve their grade. When asked for a show of hands for which type of assessment they preferred, 54 out of 79 students indicated the traditional P&P test; 23 students indicated a preference for the laboratory; and only 2 students preferred the PAs. They preferred the P&P test because they knew how to study for it; they knew how to take it; they preferred the breadth of the P&Ps over the depth of the PAs; and, above all, they knew what the teacher wanted from them. When asked why they preferred the P&P test, the students pointed to the comfort that the familiarity with the P&P test provided for them. For example, students responded.

Experience taking the test:

I prefer the regular test because we have more experience on them than the other.

I think the written test was a lot easier because it was like the ones we already had. You just had to give the answers to the questions.

It seems like there are more questions and you can study for that, but really you don’t need to study for the lab test. But the written ones—I don’t know. I just know them better.

Knowing what the teacher expects:

The written test, they kind of [like] tell you more of [like] what they want to see really. You can tell what the teacher is looking for but it’s not so clear in the other ones.

I like the written test better because it [like] gives you the directions and you don’t have to be so specific about things besides the essays. And you’re learning about many different things instead of just one thing.

Breadth of coverage:

On the written test, it covered everything that we learned. On the lab or the performance test, it was only about one or two things. So we studied everything instead of just that.

The PA required the students to come up with the rules of what steps to take when on their own, causing great concern about taking the wrong step and thereby negatively influ-

<table>
<thead>
<tr>
<th>Variable</th>
<th>Science attitudes</th>
<th>Goal orientation</th>
<th>Cognitive engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Paper and pencil</td>
<td>3.100</td>
<td>0.414</td>
<td>3.214</td>
</tr>
<tr>
<td>Laboratory</td>
<td>3.008</td>
<td>0.454</td>
<td>3.035</td>
</tr>
<tr>
<td>Performance test</td>
<td>3.149</td>
<td>0.391</td>
<td>3.213</td>
</tr>
</tbody>
</table>
encising their grade. Examples of student concerns about PA and laboratory assessments follow:

Ambiguity regarding expectations and how to proceed:

On the performance test, or whatever it was, there were no directions and you could have done something [like] what the teacher didn’t want. Like something different and on the written test there were directions and you knew what you were doing.

You didn’t get directions and it was [like] hard to do.

I didn’t like performance assessments because I didn’t know where to start or anything. I didn’t have a set of rules. I couldn’t set my goal. My goal is to get an A.

I didn’t like the performance test because if you couldn’t pick out what the rules were or if the rules you made up were wrong then everything you did would be wrong.

Depth rather than breadth of coverage:

I didn’t like it because you studied all things then it was only on one thing.

Unfamiliar with task demands:

With the lab and the other one, it all depended on how you explained it. By writing and I’m not very good at writing. The other one—it’s how much you know.

Discussions of the relative merits of PAs and laboratory tests revealed that there were some features that the students found attractive. Generally speaking, the students found certain aspects of working in teams to be both positive and negative and the interactive, hands-on nature of the tasks to be positive as well.

Benefits of working in teams:

I prefer the lab test. It’s a lot easier to get answers because you’re working in a group. And it’s a lot easier to get them in the beginning so you have them for the other part where you have to write down your results and all that. And you can get together and if you don’t know something maybe somebody else in your group does and they can help you out with it.

I like to work with partners so you can give your opinion and everyone else gives their opinion so you can get the general idea.

Hands-on aspects:

It’s something new and different and you can actually see for yourself what happens when you put two ends of a magnet together or something.

I like it more because you get to interact with your work instead of just [like] writing true/false. You got to do it.

I don’t like to wear my hand out because it starts aching. And it’s more fun and you get to be involved with your work instead of having to do it all [the experiment] in your head like with the written test.

I like doing hands on activities and I really hate writing.

I like the performance assessment because it’s just like the lab but you get to think of the rules and stuff—how to do it.

I liked it because it was more fun. You actually got to do something. I like it because you get to fool around. I mean that’s the purpose—you experiment and things.

When grades were removed from the equation, a different picture emerged; students were much more likely to take willingly the risks of deciding how to proceed on their own. Issues of challenge also emerged in the interviews. Although students found the familiar P&P tests preferable, they also indicated that the PAs presented them with more of an intellectual challenge, with challenge being viewed in a positive light.

For example, students’ preoccupation with grades is evident in the following comments (made multiple times by many different children):

I prefer the regular kind of test because you get better grades on them.

I like the regular test better because it was too hard to get a 100, because like on the graphs you’d always get a 3 and you’d call that average. You had to be really outstanding to get a 5.

When the children were asked to consider how they might feel if grades were no longer assigned, they responded.

The performance assessments are more challenging.

I think it would be a little different there because the project with the experiment would be more fun than sitting down and just writing essays and filling out true or false, fixing the wrong answer.

If there were no grades, and you were going to do the project one . . . the project one was more fun.

However, the children were not necessarily in favor of eliminating grades. Some expressed concerns that the elimination of grades would create a situation in which students would not have a reason to work hard and then would not be acknowledged for the work they did.

I like grades. It makes people stand out more. People who get good grades are going to stand out more.

I think that if somebody had a pretty bad report card and they wanted to go higher it and then they had this test where there was no grade and they got everything right they would feel pretty bad because then they wouldn’t get a higher grade.

I think it would make a big difference because you wouldn’t really care because your parents wouldn’t see the grade on the paper.

Discussion

The results of the quantitative analysis indicate that goal orientation may be influenced by assessment type. However, the picture that emerges from these data is not definitive enough to allow us to say that PA alone fosters a mastery goal orientation. In looking at the data, it is clear that both
PA and P&P tests had a tendency to foster a mastery goal orientation more for these students. But why both? And why especially P&P tests? These are the types of tests that are thought to mimic standardized tests and therefore should orient students toward performance goals. The tests should have been least likely to foster a mastery goal orientation. Even the laboratory assessments seemed to allow students more opportunities to become actively engaged with task requirements. This result should have led to more cognitive engagement and to a more mastery–task goal orientation. The hypothesized relationships do not clearly emerge simply on the basis of the quantitative data.

However, when we combine the quantitative results with the qualitative information, a somewhat clearer picture emerges. Students seemed to believe that they needed the familiar format to provide them with information that they had mastered the concepts. They also seemed to believe that this format was the best way for them to demonstrate to the teacher that they knew the material. This pattern seems to represent the recent emergence of the differentiation between performance-approach and performance-avoidant goals. Midgely et al. (2001) synthesized the literature in this area to point out that sometimes a performance-approach orientation is not harmful to learning and perhaps even helpful. What seemed to be emerging in this study was a performance-approach orientation from the students regarding P&P tests.

It is clear that these students equated a good grade with competence and ability. Because they knew the format of P&P only as a testing tool, they believed that their best chances of demonstrating competency was through this format. They did not appear to interpret any of the more open-ended classroom activities (not inclusive of testing situations) as developmental precursors to competence in performing on PAs. Nor did they appear to appreciate the difference between knowing science and doing science. The knowing part is what was important, and the only mechanism that they believed served as a vehicle for demonstrating that knowledge was the traditional P&P test.

In addition, the risk involved because of the ambiguity imposed by the openness of the task seemed to be more than these children were willing to take. Their concerns about being able to predict what the teacher wanted of them appeared to far outweigh any excitement about their being able to decide for themselves how to demonstrate understanding of scientific principles. The impression that one is left with after listening to the interviews is not that the children were necessarily reluctant to strike out on their own because of a perceived lack of ability or confidence in their ability. Rather, their reluctance seemed in large part to emanate from concerns about how such activities would affect their grades.

If we remove receiving a grade from the equation for these children, they appear to have viewed PA in a much more favorable light. Once the high stakes (high being relative to the individual at a particular time) component was removed, these children expressed much of what is hypothesized about the interaction of PA with students’ behaviors and motivation. Then the students were much more willing to take risks to try to identify the problem and attempt alternate solution paths. They stated that not only might PA be more interesting than P&P tests but also that it was actually more challenging. Of course, the concern is whether challenge is viewed as something negative and something to be avoided (i.e., frustration) or whether challenge is interpreted in the sense of being something positive and engaging. When asked for an interpretation of challenge, these students defined challenge as more of the intellectual challenge that is thought to foster task orientation. They were not referring to task difficulty and challenge equated with frustration. These students were indicating that under low-stakes conditions (in their case—no grade attached), PA may be more preferable to the other forms of assessment presented to them in this study. These students indicated that PA was more interesting, more engaging, and more intellectually challenging than the traditional P&P tests. However, when one attaches the stakes (adds the grade factor back in), students’ perceptions of which format they prefer may very well revert to the original preference for the P&P tests.

What seems to be emerging is that it may not be so much the assessment format per se that influenced the goal orientations of students but the assessment format in interaction with the stakes or consequences attached to the results of the assessment. In other words, challenge is fine (and may even carry a different definition under low-stakes conditions—that is, intellectual challenge), as long as it is not academically costly. Choice and autonomy are fine, as long as there is no perceived penalty. If one attaches stakes to an assessment, then the students’ goal orientations seem to change.

**Limitations of the Study**

The strength and clarity of the quantitative results may have been affected by the fluctuating and sometimes low reliability of the scales used. The reliability estimates reported by Meece et al. (1988) were all within acceptable ranges; the lowest reported value was .75. Several factors could have affected our reliability estimates; they varied within subscales and across administrations of the instrument. Although we anticipated some fluctuation given the different nature of each form of assessment for the different instructional units, at some points the reliability estimates were disconcertingly low. In addition to the design of the study, the size of the final sample also may have affected the reliability estimates. Larger samples typically produce more stable estimates. This result further tempers our interpretation of and confidence in the quantitative data alone.

Clearly, this study is a first step in understanding the relationship of assessment type and student motivation. Researchers need to do much more work to fully understand
the effects of assessment type on student behavior and on instructional practices to support desired student behaviors. However, the data indicate that the relationship may not be as linear as may have been believed. Nor is the relationship necessarily clearly depicted without a qualitative look at the interaction from the students' perspective.

REFERENCES


APPENDIX
Sample Test Items

Science Grade 5: Sample Paper and Pencil Test Questions

(NOTE: all items in the appendix have been sampled from among the tests administered in this study. This appendix does not represent an intact test.)

Example multiple-choice questions:
A humpback whale's migration on the east coast takes them from
   a. California to Mexico.
   b. Massachusetts to the Caribbean.
   c. Massachusetts to Africa.
   d. California to South America.
A fathom is equal to
   a. 6 feet.
   b. 10 feet.
   c. 1 foot.
   d. 12 feet.
The heat from the sun's rays cause ocean water to
   a. condense.
   b. precipitate.
   c. evaporate.
   d. transpire.
The temperature of ocean water is highest
   a. in the deep ocean zone.
   b. at the bottom.
   c. in the thermocline.
   d. at the surface.

Example Matching Exercise:

1. salinity  A. layer between fresh water and salt water
2. density  B. the amount of dissolved salts in water
3. shoreline  C. scientist who studies the ocean
4. seamounts  D. the amount of space between water molecules
5. oceanographer  E. underwater mountains
6. brackish  F. boundary where the land and ocean meet

Example True/False items

A census is the process for counting certain organisms.
A hydrophone measures the temperature of the water at different depths.
Anne is the marine biologist on "The Mimi".
Female humpback whales sing long songs.

Example short answer questions

List two indications that a storm was approaching in Episode 8.
List two pieces of information the radio tag gave the crew about the whale.
Why don't whales get hypothermia in cold water?
What is the reason we shiver when we are cold?

Other item types

Please put a "B" for benthos, "N" for nekton, or "P" for plankton on each line below to identify the organism.

Whale shark
Starfish
Killer
Kelp
Squid

Choose the best answer from the words below and write it on the line.

Abyssal plains  trenches  Submarine canyons  mid-ocean ridges  Continental shelf

Deep V-shaped valleys cut through the continental shelf and the continental slope are called ____________________
Long, narrow cracks along the ocean floor are called ____________________
Large, flat areas on the ocean floor are called ____________________
Flat area of the ocean floor covered by shallow water where you swim is called the ____________________
Mountain ranges located under the ocean are called ____________________
Science Grade 5: Planet Earth Unit (Lab)

Objective: To follow directions to conduct an experiment to compare rates at which different substances dissolve in fresh water. Students will make predictions and observations, will use appropriate units of measurement, will report results, and will form a conclusion about "What happens when different substances are added to water?"

Key Competencies:
- Follow directions to perform the experiment
- Make credible predictions
- Use appropriate units of measurement
- Make accurate observations
- Report observations accurately
- Arrive at a logical conclusion based on results of observations

Key Concepts and Content:
- Many substances form a solution when mixed with water; some do not.
- Some substances go into solution faster than others.

Science Skills Needed:
- Observing
- Measuring
- Organizing
- Inferring
- Predicting
- Experimenting
- Communicating

Materials Needed: For the Class:
- water
- table salt
- granulated table sugar
- corn starch

For Each Group:
- 3 clear plastic cups
- 3 plastic straws or stirrers
- a plastic teaspoon
- 3 pieces of tape or sticky labels
- a graduated measuring cup
- data sheets
- crayons or colored pencils

Performance Task I: In groups, students will follow directions to conduct an experiment observing the rate at which a substance dissolves in water. They will work as a group up to the point of writing a conclusion.

Performance Task II: Individually, students will write a conclusion to the experiment. They will base their conclusion on the observations and recordings from the experiment.

Directions Given to Students

Directions for the Groups:
You are going to add table salt, sugar and corn starch to cups of water. What do you predict will happen to the substances in each cup?

Salt:

Sugar:

Corn Starch:

Fill three clear plastic cups with about the same amount of water. Leave about an inch at the top so you can stir. Label each so you know which gets salt, sugar, and corn starch.

How many teaspoons will you add to each cup? Circle 1 2 or 3.

Each cup gets the same number.

Drop in the number of teaspoons you circled above of each substance into the appropriate cup. After adding the substances, sit and observe for two minutes. What can you see happening?

Stir each cup 10 times using the same technique. Did any dissolve?

Repeat stirring each 10 times and then observing all of them. Continue to do this until the substance disappears. Mark out the number of times stirred before a substance disappeared if it did so.

Salt:

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
</tr>
</tbody>
</table>
| 130| 140| 150| 160| 170| 180| 190| 200

Sugar:

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
</tr>
</tbody>
</table>
| 130| 140| 150| 160| 170| 180| 190| 200

Corn Starch:

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
</tr>
</tbody>
</table>
| 130| 140| 150| 160| 170| 180| 190| 200

Make a bar graph of these results.

Did each dissolve:
Which dissolved fastest?
Write a conclusion that you could make based on the results of your experiment.

**Science Grade 5: Planet Earth Unit (PA)**

**Objective:** To follow directions to conduct an experiment to compare rates at which different substances dissolve in fresh water. Students will make predictions and observations, will use appropriate units of measurement, will report results, and will form a conclusion about “What happens when different substances are added to water?”

**Key Competencies:**
- Make credible predictions
- Use appropriate units of measurement
- Make accurate observations
- Report observations accurately
- Arrive at a logical conclusion based on results of observations

**Key Concepts and Content:**
- Many substances form a solution when mixed with water; some do not.
- Some substances go into solution faster than others.

**Science Skills Needed:**
- Observing
- Measuring
- Organizing
- Inferring
- Predicting
- Experimenting
- Communicating

**Materials Needed: For the Class:**
- water
- table salt
- granulated table sugar
- corn starch

**For Each Group:**
- 3 clear plastic cups
- 3 plastic straws or stirrers
- a plastic teaspoon
- 3 pieces of tape or sticky labels
- a graduated measuring cup
- data sheets
- crayons or colored pencils

**Performance Task I:** In groups, students will follow directions to conduct an experiment observing the rate at which a substance dissolves in water. They will work as a group up to the point of writing a conclusion.

**Performance Task II:** Individually, students will write a conclusion to the experiment. They will base their conclusion on the observations and recordings from the experiment.

**Directions Given to Students**

**Directions for the Groups:**
Will salt, sugar and corn starch mix completely and dissolve when added to water? We know some substances will dissolve in water and others will not. Why is that? What are the things that could possibly make a difference in whether the substance will dissolve or not?

List some things that might make a difference in the ability of salt, sugar or corn starch to dissolve in water.

ALL THINGS BEING EQUAL, meaning that you do the same thing for each substance each time...............  

What do you predict will happen when you add salt to water?  
What do you predict will happen when you add sugar to water?  
What do you predict will happen when you add corn starch to water?  

You have been given some equipment to use to create a way of determining if your predictions were right. Conduct an experiment and observe what happens to see if you were right.

You have been given:  
- 3 clear plastic cups  
- 3 plastic straws or stirrers  
- a plastic teaspoon  
- 3 pieces of tape or sticky labels  
- a graduated measuring cup  
- data sheets  
- crayons or colored pencils

Available to the whole class is:  
- water  
- table salt  
- granulated table sugar  
- corn starch

You will need to describe your experiment (what you are doing).  
You will need to record your observations (tell what happens as it happens).  
You will need to present your results both in words and in a graph.  
You will need to reach a conclusion (tell what happens to the three substances when added to water).  
You will need to communicate that conclusion to the teacher.