Using Cognitive Analyses to Understand Motivational and Situational Influences in Science Achievement

Angela M. Haydel

SRI International

Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, April 2003. Correspondence should be addressed to Angela M. Haydel, SRI International, 333 Ravenswood Ave. BN334, Menlo Park, CA 94025. This research was supported by a grant from the National Science Foundation (REC-0126352). The opinions expressed here are those of the author and not necessarily those of the National Science Foundation.

Abstract

This study explores the relationship between students' test perceptions, motivation and engagement in multiple-choice and performance assessment testing contexts. Thirty-three students, a subset of which were classified as mastery-oriented or performance-oriented, participated in three think-aloud test sessions and described both their thoughts and feelings as they completed the test items. Students also answered a series of follow-up interview questions related to their motivation during testing and perceptions of the test items. In general, students perceived performance assessments as more valid, interesting and challenging compared to multiple-choice tests. Mastery-oriented students were more likely to use strategic knowledge and help-seeking strategies than performance-oriented students. Performance-oriented students tended to rely more on declarative and procedural knowledge during problem solving. Implications of these findings for test score interpretation and assessment development are discussed.

Using Cognitive Analyses to Understand Motivational and Situational Influences in Science Achievement

Standards-based reform has challenged science educators to create learning environments and assessments that emphasize deep conceptual understanding and foster mastery of the material. In order to draw conclusions about student achievement, a critical analysis of student engagement on these assessment items is needed (Lazer, Mislevy, Whittington, & Ward, 1995). If the current tests are not appropriately assessing significant learning outcomes, then using these scores as evidence to judge student understanding is questionable. Traditionally, educators and policy makers have focused on the extent to which test scores reflect students' cognitive abilities and depth of understanding in a particular domain. However, motivation also may be a critical factor to understanding achievement. Recently, Pellegrino, et al. (2001) concluded, "For classroom or large-scale assessment to be effective, students must understand and share the goals for learning" (p. 9), illustrating the importance of understanding students' goals related to learning and achievement in a domain. Thus, students' test scores may reflect not only their knowledge of the subject matter, but also their motivation in the subject matter domain, their motivation for engaging with a particular assessment, and the nature and history of their instruction (Snow, 1994).

Understanding the relations between students' motivation, their perceptions of assessments, and their actual achievement on these tests will contribute to a more general understanding of the full spectrum of cognitive, motivational and situational processes that are implicit in students' test scores. Previous research has failed to adequately examine the interaction between the affordances of what individuals bring to a test situation, in terms of their motivational goals and beliefs, abilities and achievement history, and the affordances of various

assessment formats. Thus, when the validity of tests is disputable, educators and policymakers are inhibited from fully understanding the meaning of students' test scores.

In one approach to studying motivation, goal theorists (e.g., Ames 1992; Dweck, 1986; Dweck & Leggett, 1988; Maehr & Midgley, 1996) identified two goal orientations (mastery and performance) that differentially relate to engagement in achievement situations. Students with mastery goals engage in learning and achievement tasks as a means of improving their skills and competencies, whereas students with performance goals view achievement situations as an opportunity to prove their superior relative ability or hide their perceived sense of incompetence. Students with performance goals who focus on proving relative ability (a performance-approach orientation) tend to engage in self-regulated learning (Wolters, Yu, & Pintrich, 1996) and perform as well or better than mastery-oriented students on achievement-related outcomes such as grades or test performance (e.g., Harackiewicz, Barron, & Elliot, 1998; Harackiewicz et al., 2000; O'Neil et al., 1997). In contrast, students with performance goals who are concerned about avoiding failure (a performance avoidance orientation) disengage from challenging tasks (Stipek & Gralinski, 1996), lack persistence, use maladaptive strategies, and exhibit performance deficits (Ames & Archer, 1988; Kaplan & Midgley, 1997; Meece, Blumenfeld, & Hoyle, 1988; Middleton & Midgley, 1997; Nicholls, Cobb, Wood, Yackel, & Patashnick, 1990; Nolen, 1988; Nolen & Haladyna, 1990; Wolters, 1998).

In a series of studies O'Neil et al. (1997) examined how goal orientation influences performance on the National Assessment of Educational Progress (NAEP) mathematics reference examination by manipulating students' motivational goals through the test instructions. In one study, students in the performance-oriented condition performed better than students who had received standard NAEP instructions. This finding was inconsistent with what the authors

anticipated – mastery-oriented students were expected to outperform performance-oriented students. However, the authors had not considered the possibility that some of these students might have adopted a performance-approach orientation, which is sometimes adaptive (Harackiewicz, Barron, & Elliot, 1998; Harackiewicz et al., 2000). In addition, they did not consider the possibility that the NAEP items may be well-suited to allow performance-oriented students to feel as if they can accomplish their goal of showing what they know and can do relative to other students. That is, it may be the case that the NAEP item characteristics are more congruent with a performance orientation.

NAEP mathematics items have certain characteristics such as the types of knowledge required, the amount of structure, the level of complexity, and novelty. The items administered in by O'Neil et al. (1997) were primarily multiple choice, which suggests that they are constrained in terms of the overall structure, have a clear method of response (circle/fill-in correct response) and may be less likely to require complex problem-solving skills compared to more open-ended mathematics problems. Because performance-oriented students perceive effort as an indication of lack of ability (Dweck & Leggett, 1988), these more process-constrained types of items may be ideal for performance-oriented students who want to show what they know and can do with as little effort as possible.

Multiple-choice items such as these may be less compatible with the mastery-oriented goal of wanting to improve mathematics skills. Thus, for mastery-oriented students there may be a lack of congruence between their goal orientation and these types of test items. Mastery students may be more likely to prefer more complex, novel and authentic tasks that provide them with the experience of feeling challenged, help them attain their goal of improving their knowledge and skills, provide valuable feedback about which concepts they have mastered.

When the problem-solving environment is incongruent with students' personal goals and capabilities, successful performance on assessments may be more difficult because engagement in the problem-solving environment will not ensure goal attainment (Ford, 1992).

It may be the case that students bring motivational goals and a set of learned skills and knowledge to a test, and that different types of tests may be better at revealing different skills and knowledge or engaging students. In order to understand what scores on assessments tell us about what students know, we should consider the congruence between students' motivational goals during testing and the demands of the assessment environment.

The domain of science presents a useful context for exploring these relationships because items on these examinations range from multiple-choice (more processed-closed) to "hands-on" performance assessments (more process-open). However, few studies have examined students' perceptions of these item formats in science. This study explores the relationship between students' perceptions of multiple-choice tests and performance assessments, their motivational goals, and their cognitive, metacognitive and affective engagement during testing.

Surveys were used to provide general information about students' perceptions of the different test items and their motivational goals. Qualitative methods, including "think-alouds" and post-test interviews with a small sample of students, were used to illustrate cognitive, metacognitive and motivational engagement on multiple-choice and performance assessment items. Students' articulations of their thought processes can provide meaningful information about their knowledge, experiences and feelings, and may reflect different cognitions related to their motivation, knowledge, and strategy use (e.g., Baxter & Glaser, 1998; Haydel, 2002; Ruiz-Primo, Schultz, Li, & Shavelson, 1999; Schutz & Davis, 2000; Schutz et al., 1998; Yepes-Baraya, 1995). The think-aloud methodology allows for greater insight into students' short-term

thought processes and experiences during problem solving and is less likely to interfere with test administration compared to follow-up questions during testing (Hamilton, Nussbaum, & Snow, 1997).

Method

Participants

Think-alouds were conducted with 33 (52% female) eighth grade students from two middle schools in the Pacific Northwest. Of these students, 6 were classified as mastery-oriented and 5 as performance-oriented because they consistently requested mastery-oriented or performance-oriented feedback about their test performance across the three days of testing on a post-test survey measure of motivation.

Procedure

Participants completed a pre-test survey that included questions about their perceptions of multiple-choice test and performance assessments. Approximately two weeks later, the think-alouds were conducted with each student individually during three consecutive days of testing. Each session took approximately 50 minutes and was audiotaped. During the first think-aloud session, students completed nine multiple-choice and four constructed-response items that focused on content related primarily to Populations and Ecosystems. During the second think-aloud session students completed one of two performance assessments. During the third think-aloud session, students completed whichever performance assessment they did not work on during the second think-aloud session.

Protocols were prepared for conducting the think-alouds and administering a set of follow-up interview questions about their science knowledge, their motivation in science and

their test perceptions. Because the goal was to capture the full range of students' cognitive, metacognitive and affective experiences during the test, students were asked to talk about both what they were thinking and feeling as they completed the test items.

Prior to beginning the first testing session, students practiced thinking aloud as they tied their shoe. (The experimenter provided a tennis shoe if the student was not wearing a shoe with laces.) Because tying one's shoe generally does not elicit much feeling, students were asked to complete a second practice think-aloud task during which they attempted to tie their shoe with only one hand. This task was intended to be perceived as a challenge that would elicit some emotion. Although most students perceived this as a fun challenge, the task may have frustrated some students. Therefore, at the completion of this practice task, students were reminded that the task was supposed to be challenging and to remember to talk about both positive and negative feelings while taking the test.

At the end of each test session, students were asked a series of questions about their knowledge, feelings and perceptions related to the test items they just completed. Specifically, if students did not provide detailed explanations of their thinking or feeling during the think-aloud, they were asked to explain their answers and how they were feeling during the post-test. More general questions about test items were asked, as well. For example, students were asked how valid multiple-choice tests and performance assessments are for measuring what they know in science and which type of test they would prefer to have more of in their science class.

Materials

Teacher Rating of Student Science Achievement. Science teachers rated each student on their science achievement using a three-point scale (high, medium, or low) based on how

students were performing in comparison to all eighth grade students they were currently teaching.

Survey Measures of Test Perceptions. Students were asked to indicate whether they believed multiple-choice tests or performance assessments were more authentic, challenging, ambiguous, effortful, valid, meaningful, anxiety producing, interesting, and difficult. In addition, students were asked which test made making comparisons between students easier and on which test they would prefer to have their science grade based. Questions were answered using a four-point scale (1 = Definitely Multiple-Choice, 2 = Probably Multiple-Choice, 3 = Probably Hands-On, 4 = Definitely Hands-On). These items were new and designed for this study.

Science Assessments. All assessment items were drawn from three examinations: the National Assessment of Educational Progress in Science (NAEP) (1996, 2000), the Third International Mathematics and Science Study (TIMSS) (1995), the TIMSS-Repeat (1999) examination, and the New Standards Science Reference Examination (1999). Both secure and released items from each examination were included in the test booklets. During the first test session, students completed a booklet composed of nine multiple-choice and four constructed-response items. During the second and third testing sessions, students completed either the "Aquarium" performance assessment (New Standards, 1999) or the Solutions" performance assessment (TIMSS, 1995). In the "Solutions" assessment students were asked to design an experiment to determine how water temperature affects the speed with which a tablet dissolves. "Solutions" may be described as process-open because students were not provided with any scaffolding or directions to help them design and conduct their investigations. Only one question required any content knowledge, so "Solutions" also may be considered relatively content-lean.

Details about the items "Aquarium" task cannot be described in this paper because the items must remain secure according to the New Standards regulations. However, the "Aquarium" assessment may be described as moderately content-rich and process-constrained. In a majority of the questions students were required to use content knowledge, but the procedures for conducting tests of water samples were clearly specified for students.

Motivational Construal of Testing Context. At the end of each test, students completed a survey in which they were asked to indicate the kind of feedback they would like about their performance on the test items. Responses reflect either a mastery orientation (information about "what concepts you demonstrated good knowledge and understanding of") or a performance orientation (information about "how you did compared to other students of your age on the test"). Students were characterized as having a mastery goal orientation or a performance goal orientation based on their responses to a post-test survey. Students who consistently selected the mastery-oriented option over the three days of testing were classified as mastery-oriented, and students who consistently selected the performance-oriented option were classified as performance-oriented.

Coding Procedures

Think-aloud audiotapes were transcribed and segmented by test item. The start of a segment was defined as where the student began discussing a particular test item and the end of a segment was defined as when the student finished discussing that test item. In order to develop the coding categories, segments were reviewed initially to determine the range of statements made by students during the think-alouds and post-test interviews. Based on this overview of the transcripts, coding categories were developed to reflect the range of cognitive, metacognitive and

affective statements made by students. See Table 1 for a list of coding categories and definitions.

Raters participated in training sessions prior to coding each item. During the training sessions raters discussed the content of the item and practiced applying the coding categories to segments of students' transcripts from the think-alouds and post-test interviews associated with that particular item. Discrepancies between raters were discussed and resolved during practice sessions. At the end of the practice sessions, raters were randomly assigned a list of student identification numbers indicating which transcripts they had to code for a particular item.

Two raters coded at least 50% of all papers, while the remaining papers were coded by only one rater. Consistency across raters was determined by calculating the sum of agreement out of the total number of entries between two raters. The agreement for each coding category reported in these analyses is 85% or higher. Because agreement was sufficiently high, ratings from either Rater 1 or Rater 2 were used for double-coded papers.

Only a selection of student work was coded due to the extensive time it took to code each item. For the multiple-choice items, three were selected that addressed skills related to interpreting graphs. Three additional multiple-choice items were also coded. Items in each performance assessment were highly correlated, suggesting that similar knowledge and skills were measured in each of these items. Therefore, three out of the seven "Aquarium" items were coded (although one of these items has two parts—doing an experiment and analyzing results), and three out of the four "Solutions" items were coded. The one item omitted from the Solutions performance task required content knowledge with which most students participating in the think-alouds were not familiar and therefore could not discuss their thoughts in detail.

Missing Data. Most of the 33 students selected to participate in the think-alouds were present on all testing days. One student was absent during the "Aquarium" performance assessment. Due to the constraints of the think-aloud design and the teachers' and students' schedules, a make-up session was not possible. Because sample sizes were small, data from this student was still included in analyses for those items that were completed.

Data was also considered missing if students completed an item, and their response was inaudible on the audiotape, or if students chose to skip an item but did not have time to return to it before the end of the test session. Rather than indicating that evidence of the codes was not present, these students' work was noted by raters as inaudible or missing.

Not all students had time to complete post-test interviews. In addition, the researcher did not always have time to get through all of the interview questions directly related to items completed by students. Therefore, students who did not complete all of the interview questions on a given testing day had missing data for responses based on the post-test interview data. Generally, there is more missing post-test data for the performance assessment items because the performance assessments took longer to complete, and so less time remained for the interviews.

Results

Teacher Rating of Student Science Achievement

Because engagement may reflect differences in prior science knowledge and skills, it is worthwhile to note that the mean teacher rating of science achievement for students participating in the think-alouds was 2.30. There was no difference in teacher rating of science achievement for mastery-oriented and performance-oriented students.

Cognitive and Metacognitive Engagement During Testing

Results related to cognitive and metacognitive engagement for all students participating in the think-alouds are presented in Table 2. While the average percentage of students who used declarative knowledge or schematic knowledge did not differ by item format, more students used procedural knowledge on performance assessment items. A greater percentage of students used strategic knowledge (e.g., planning, monitoring progress, rereading) on performance assessment items than on multiple-choice items. Despite being in a test situation, some students who participated in the think-alouds felt comfortable asking for help when uncertain about how to answer an item. (Additional help from the researcher related to identifying a correct solution was not provided.) More students used this strategy of asking for help while working on the performance assessment items.

Table 3 shows the average percentage of mastery- and performance-oriented students who made statements about cognitive engagement, strategy use and metacognition on the multiple-choice and performance assessment items. On multiple-choice items, approximately the same percentage of mastery- and performance-oriented students used declarative and schematic knowledge, although performance-oriented students were more likely than mastery-oriented students to use procedural knowledge on these items. On performance assessment items, more performance-oriented students made declarative and procedural knowledge statements than mastery-oriented students, while approximately the same percentage of students in both groups used schematic knowledge.

On both multiple-choice items and performance assessment items, more mastery-oriented students used strategic knowledge. Mastery-oriented students also were somewhat more likely than performance-oriented students to make statements regarding having the knowledge and

skills needed to solve both multiple-choice and performance assessment problems. On multiple-choice items, mastery-oriented students made more statements about lacking necessary knowledge or skills, but on performance assessment items, performance-oriented students made statements about lacking necessary knowledge or skills.

The average number of students who sought help on the multiple-choice items was relatively equal during multiple-choice items, on average. Even though performance-oriented students were more likely to be aware that they lacked particular knowledge and skills need to solve performance assessment items than mastery-oriented students, fewer of these students asked for help.

Perceptions of Multiple-Choice Tests and Performance Assessments

Findings from questions on the pre-test survey showed that students believed that multiple-choice tests and performance assessments have different qualities. Table 4 displays perceptions of multiple-choice and performance assessment items for students participating in the think-alouds. Students characterized performance assessments as having features associated more with mastery and challenge, whereas multiple-choice tests were perceived as being associated more with performance-oriented features such as evaluation and competition. Students reported that performance assessments are more challenging, require more thinking, are more authentic, and are more valid compared to multiple-choice tests. Students also indicated that multiple-choice tests are easier to use for the purpose of making comparisons among students. Most students believed that multiple-choice tests are less difficult, less anxiety producing, and require less effort than performance assessments. Students' perceptions of these items did not differ by motivational orientation.

Table 5 shows the distribution of students' statements about their perceptions and feelings about multiple-choice tests and performance assessments based on analyses of the think-alouds. Generally, students said little about their perceptions of items during testing, although they reported slightly more positive feelings during the performance assessments. Interestingly, students were just as likely to describe negative feelings and perceptions about performance assessment items. Paralleling the survey data, students have positive perceptions about performance assessments related to authenticity and interest, but these are accompanied with anxiety and uncertainty about how to get the right answer. Negative statements about multiple-choice items during think-alouds were not balanced by as many positive statements.

Table 6 shows that students who adopted either a mastery goal orientation or performance goal orientation in relation to the tests were more likely to report negative feelings than positive feelings during multiple-choice items. In fact, mastery-oriented students did not report any positive feelings about multiple-choice items, and were slightly more likely to describe these items as difficult. Contrary to expectations, mastery-oriented students also reported more negative than positive feelings during the performance assessments, while the opposite was evident for performance-oriented students.

Typically, students concentrated on responding directly to the questions during the think-aloud rather than reflecting about their feelings and perceptions. Post-test interviews provided students with the opportunity to express additional feelings and perceptions of test items. Students' perceptions and feelings based on post-test interviews are presented in Table 7. After completing the items, students, on average, were more likely to describe the multiple-choice items as easy compared to the performance assessment items, paralleling the findings from the survey. While students made both positive and negative statements about multiple-choice and

performance assessment items after testing, they were about twice as likely to report negative feelings about multiple-choice items than about performance assessment items.

During the post-test interviews, mastery students reported more positive feelings when reflecting on multiple-choice items and performance assessment items compared to performance oriented students. (See Table 8.) Negative feelings about multiple-choice items were relatively comparable for both groups, but higher for performance-oriented students for performance assessments. Mastery-oriented students were more likely to describe multiple-choice items as difficult than easy, while the opposite pattern was found for performance-oriented students. However, mastery-oriented students were much more likely to describe performance assessment items as easy compared to performance-oriented students.

In post-test interviews many students remarked that multiple-choice tests were easier because the right answer is available among several options and there is a 25% chance of guessing the right answer. For instance, when asked about what he thought about multiple-choice questions, one student said:

> You don't really need to know the answer, you just have to have an idea because when you look at the multiple choice then if you just somewhat know then it'll pop out at you, but if have to write it down then you have to know exactly what it is. And so it's [multiple-choice] a little easier.

After completing the multiple-choice items, another student commented:

> They were pretty easy. If they put answers that made, if they put a lot of answers that made really good sense, they would have been harder. But you can usually go through and eliminate at least one, usually two, of the answers because they don't make any sense to the questions.

Following up with this student, the interviewer asked her to explain what makes a good multiple-choice question, and she responded:

> If it makes you have to stop and think about it. Like sometimes, sometimes multiple-choice questions, um, when teachers make up tests and stuff, they'll put, they'll put like

> two answers that make sense and then they'll put like popcorn and cheese, stuff like that. Stuff that doesn't, doesn't even fit the question at all. And that's no fun because, that, I mean it's not a challenge at all. And, I mean, it may be for some people, but it's not for me. And so just a bunch of answers that make sense and actually make you think about it, that's a good multiple-choice question. Because stupid ones are dumb and a waste of time.

So this student seems to enjoy being challenged on tests, but admits that because oftentimes multiple-choice questions are written poorly, they are easy to answer.

In addition to responding to specific questions about the test items in the post-test interviews, students were asked questions about the validity of different kinds of test questions and their preference for working on different kinds of test questions in general. To assess students' perceptions of item validity, they were asked, "In what ways, if at all, did this test let you show what you know about science?" To examine students' format preferences, students were asked, "If you had to choose, which type of test questions would you prefer to have more of in your science class?"

All students, regardless of motivation group, believed that performance assessments provided more valid test questions. For instance, one student stated:

> I like [multiple-choice questions] because you don't get as many wrong but um like you don't really get to show like I mean your teachers knows that you knew, you've got the right [answer] but your teacher doesn't know if you just circled it randomly or if you actually got it right and like it's, like if you get your test back and…if it's like a multiple choice question test then it's like well yeah but how much do I really know, you know, if you like forget which ones you guessed on.

Another student commented:

> [Hands-on tests] let you show a lot better then multiple choice or just questions because you have to carry out what you write and you have to just know what you're doing and it's more of a life experience sort of.

When asked about which type of test they would prefer to have more of in their science class, students' responses reflected their motivation group in some ways. Mastery students

preferred the performance assessments. The reasons they provided were that they were "fun" and helped them to "learn more". Learning was important to these students even in the context of a test. Performance-oriented students provided more mixed responses with respect to test preference. Students who preferred performance assessments did so for most of the same reasons as the mastery students. However, performance-oriented students were also likely to say that they preferred performance assessments because they were "easier." If performance-oriented students preferred multiple-choice items, it was because they thought that it was "easier to get a better grade" on multiple-choice items.

Discussion

Characterizing Science Achievement Tasks

Findings indicated that students perceive that multiple-choice tests and performance assessments differ in terms of key features such as *complexity/information burden*, *ambiguity*, and *authenticity* (Doyle, 1983; Snow, 1994). Students believed that performance assessments were more challenging, more difficult and required more effort, which suggests that they believed that performance assessments were more complex than multiple-choice items. Additional evidence related to complexity comes from the post-test surveys and think-alouds. Students used more procedural and strategic knowledge on performance assessment items. Strategy use may have been greater during the performance assessments because the performance assessment items required students to plan an investigation, manipulate equipment and explain experimental findings. Similarly, in the performance assessments students had to carry out their investigations, which required more procedural knowledge.

Students reported that it was more difficult to know the right answer on performance assessments, which implies that the process for completing performance assessments may be

more ambiguous for students. This also suggests that the procedures used to evaluate student work on performance assessment are less clear to students. Another indication that students perceive performance assessments as ambiguous is that help-seeking increased while students worked on the performance assessments.

These findings illustrated not only that students perceive some distinctions between multiple-choice tests and performance assessments, but also that students may view these two types of tests as representing different aspects of the school culture. School "culture" may be related to learning or to demonstrating competence (Maehr & Midgley, 1996). For students in this study, performance assessments seem to represent that dimension of school that is focused on learning. Learning-focused environments are about being challenged, working hard, and hopefully, engaging in an activity that has some meaning or purpose outside of the classroom (Maehr & Midgley, 1996). Despite being an evaluation of performance, these findings suggested that students perceive performance assessments as more congruent with a learning-focused school culture. Perhaps because performance assessments allow students to try out their ideas, manipulate laboratory equipment, and often have more than one right answer, students are more likely to view them as learning-oriented.

Assessments are used to evaluate learning, and thus also provide opportunities for comparisons between students. As a result, assessments can shift focus away from learning towards competition and a performance orientation. Students seem to construe multiple-choice tests as more related to this aspect of school culture. Not only do they believe that multiple-choice tests allow for easier comparisons among students, but when grades become relevant, students prefer multiple-choice tests, rather than the performance assessments that they find more authentic, valid, interesting and fun. If the goal is to do better than other students or to

avoid failing, then it makes sense that students would want to be evaluated with an assessment on which they feel most confident they can achieve.

Motivation in Assessment Contexts

The theoretical approach to goal orientation adopted in this study is one in which personal goals are viewed as context-dependent (Ames & Archer, 1998, Maehr & Midgley, 1996). That is, a student's goal orientation may change depending on the environment. If a mastery-orientation is salient in a classroom environment, students are more likely to adopt a mastery-orientation to their work in that class.

Typically, studies of motivation are situated within the classroom context. Thus, general classroom measures of goal orientation can be related to achievement outcomes. In this study, however, the assessments were not embedded within classroom activities. Rather, the administration of these assessments was more similar to that of NAEP or TIMSS – the "drop-from-the-sky" approach to summative assessment. Because the assessments are not integrated into the classroom activities, students' classroom motivational goals are not necessarily relevant in this assessment context. Therefore, the measures often used to identify students' goal orientation were not used in this study. Instead, students' responses on the post-test surveys were used in order to obtain a measure of goal orientation that is associated more closely with the testing context. Students who consistently indicated on this survey that they wanted mastery-oriented or performance-oriented feedback about their performance across each of the testing days were classified as mastery-oriented or performance-oriented, respectively.

One limitation to this method was the number of mastery-oriented and performance-oriented students could not be identified until after completion of the testing. As a result, the sample size for each group was smaller than anticipated. In addition, the performance-oriented

option was not specifically approach- or avoidance-focused. Given, previous work on motivation differences in engagement may be expected from individuals with a performance-approach or performance-avoidance goal orientation, and would be worthwhile to explore in future studies that examine the relationship between motivation and engagement on assessments. Despite these limitations, the analyses yielded interesting patterns about the relationship between students' motivational orientation and their test perceptions, feelings, cognitive engagement and metacognitive engagement in relation to multiple-choice and performance assessment items.

Relating Motivational Goals, Test Perceptions and Engagement on Achievement Tasks

Results from the think-alouds helped to distinguish between how mastery-oriented and performance-oriented students approach problem solving. Mastery-oriented students were more likely to use strategic knowledge while working on multiple-choice and performance assessment items, while performance-oriented students used more procedural and declarative knowledge. Performance-oriented students were more likely to discuss information about the basic knowledge and skills related to the problems, but less likely to be less reflective about their strategic approach. Although planning prior to problem solving and monitoring throughout problem solving may take more time, they should lead to more systematic problem-solving. Interestingly, performance-oriented students wanted to get a good grade, but did not want to invest this extra effort. It may be the case that mastery-oriented students are more interested in the challenge of developing a problem-solving strategy and taking the time to monitor or double-check their work. Perhaps because performance-oriented students view effort as an indication of lack of ability (Dweck & Leggett, 1998), they may hesitate to put the extra effort into planning and monitoring. Instead they may prefer to show that they know more than others by stating simple facts and procedures. Analyses of post-test interviews also showed that performance-

oriented students take into account the amount of effort they think they have to put into the task to succeed (or avoid failing) as a factor when determining which assessment they prefer. Many performance-oriented students preferred the multiple-choice items because they were perceived as being easier to solve.

Another finding that distinguishes mastery-oriented from performance-oriented students is related to help-seeking strategies. Mastery-oriented students were approximately twice as likely to ask for help on the performance assessments compared to the performance-oriented students. Help-seeking could be an indication that these students lacked confidence in their problem-solving abilities. However, because mastery-oriented students also reported that these performance assessments items were relatively easy, this interpretation of help-seeking may not be appropriate. Rather, it may be the case that mastery-oriented students were less likely to view help-seeking as an indication of lack of ability as performance-oriented students might. Asking for help is more likely to make others aware that one does not know how to solve a problem, which may be why performance-oriented students are more hesitant to do so. Help-seeking may not be the most appropriate strategy to apply in a testing context (i.e., when help cannot be provided), but in a classroom learning activity help-seeking certainly could increase learning. If performance-oriented students are reluctant to seek help during classroom activities, they may miss out on important learning opportunities.

Findings provided some evidence that a mastery orientation may be more congruent with a performance assessment testing context than with a multiple-choice testing context. A larger percentage of mastery-oriented students reported that performance assessment items were more easy than difficult, which may indicate that they feel more confident working on performance assessments. Mastery-oriented students were more likely to apply the strategic knowledge

needed to solve the more complex performance assessment items, as well. Post-test interviews also revealed that mastery-oriented students preferred performance assessments to multiple-choice items.

It may also be the case that multiple-choice tests are more congruent with a performance orientation. In post-test interviews, performance-oriented students were more likely to report that they felt more positively about multiple-choice items than performance assessments, that they preferred working on multiple-choice items, and that multiple-choice items were easy.

Implications for Score Interpretation and Assessment Development

Findings from this study help to inform our understanding of students' test scores and may have implications for assessment development, as well. These results provide evidence that students' test scores may be not only a function of what they know but also their motivational orientation during testing. That is, test scores may reflect the use of different types of knowledge for students with different motivational orientations. When students adopt a mastery-orientation, they are more likely to engage in higher levels of cognitive processing, especially on the more challenging and complex performance assessment items. They are also more likely to feel positively while working on performance assessments and prefer them to the multiple-choice items. The knowledge and skills that performance-oriented students bring to bear tend to be primarily declarative, procedural or schematic. These types of knowledge and skills are certainly useful in problem solving, but may not help them to engage in as systematic problem-solving as the mastery-oriented students.

Generally, different item formats provided information about different knowledge and skills. The multiple-choice and performance assessments used in this study elicited approximately the same amount of declarative and schematic knowledge, while the performance

assessments elicited more procedural and strategic knowledge. Help-seeking increased on the performance assessments which indicates that these open-ended formats are more difficult for students, perhaps because they do not have enough experience designing experiments and collecting data independently in their science classes. In order to evaluate greater variety of knowledge and skills, assessments need to include multiple item formats.

Moreover, a teacher who is constructing an assessment for use in her classroom may want to consider whether the characteristics of the assessment and the feedback she provides about student performance on the assessment are congruent with the goal orientation she promotes in her classroom. For example, if students perceive performance assessments as having more learning-oriented characteristics, performance assessments may be more appropriate to use in a class where mastery-oriented goals are emphasized. However, in mastery-oriented classroom feedback about student achievement, even on a performance assessment, must be provided such that students learn what concepts they have and have not mastered. Feedback that consists merely of an overall score reported for each student would not be consistent with a mastery-orientation.

In summary, this study is an exploratory examination of the relationship between test perceptions, motivation and engagement in different assessment contexts. Although limited sample sizes inhibit the generalizability of these results, detailed information about the knowledge, skills and beliefs of the students presented in this paper suggest that further research relating these concepts would improve our understanding about the cognitive and motivational processes that contribute to test performance.

References

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260-267.

Ames. C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271.

Baxter, G. P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Issues and Practice, 17*, 205-226.

Baxter, G. P., Elder, A. D., & Glaser, R. (1995). *Cognitive analysis of a science performance assessment*. CSE Technical Report 398. Los Angeles, CA: University of California, National Center for the Study of Evaluation, Standards, and Student Testing.

Carnevale, A. P. & Kimmel, E. W. (1997). *A national test: Balancing policy and technical issues*. Princeton, NJ: Educational Testing Service.

deJong, T., & Ferguson-Hessler, M. G. M. (1986). Cognitive structures of good and poor novice problem solvers in physics. *Journal of Educational Psychology, 78*, 279-288.

Doyle, W. (1983). Academic work. *Review of Educational Research, 53*, 159-199.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256-273.

Ford, M. E. (1992). *Motivating humans: Goals, emotions, and personal agency beliefs*. Newbury Park: Sage Publications.

Hamilton, L. S., Nussbaum, E. M., & Snow, R. E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education, 10*, 181-200.

Harackiewicz, J. M., Barron, K. E., Elliot, A. J. (1998). Rethinking achievement goals: When are they adaptive for college students and why? *Educational Psychologist, 33*, 1-21.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M., & Elliot, A. J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology, 92*, 316-330.

Haydel, A. M. (2002). *Measuring more than we know? An examination of the motivational and situational influences in science achievement.* Unpublished doctoral dissertation, Stanford University.

Lazer, S.; Mislevy, R. J.; Whittington, K.; & Ward, W. C. (1997). Measuring cognitive skills. In E.G. Johnson, S. Lazer, & C.Y. O'Sullivan (Eds.), *NAEP reconfigured: An integrated redesign of the National Assessment of Educational Progress (4.1-4.52).* Princeton, NJ: Educational Testing Service.

Li, M. (2001). *A framework for science achievement and its link to test items*. Unpublished doctoral dissertation, Stanford University.

Maehr, M. L., & Midgley, C. (1996). *Transforming school cultures*. Boulder, CO: Westview and Harper Collins.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientation and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*, 514-523.

Middleton, M., & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An unexplored aspect of goal theory. *Journal of Educational Psychology, 89*, 710-718.

National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.

Nicholls, J. G., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Assessing student's theories in mathematics: Individual and classroom differences. *Journal for Research in Mathematics Education, 21*, 109-122.

Nolen, S. B. & Haladyna, T. M. (1990). Personal and environmental influences on students' beliefs about effective study strategies. *Contemporary Educational Psychology 15*, 116-130.

Nolen, S. B. (1988). Reasons for studying: Motivation orientations and study strategies. *Cognition and Instruction, 5*, 269-287.

O'Neil, H. F., Sugrue, B., Abedi, J., Baker, E.L., & Golan, S. (1997). Final Report of Experimental Studies on Motivation and NAEP Test Performance. *CSE Technical Report 427.* Los Angeles, CA: Center for the Study of Evaluation; CRESST.

Pellegrino, J., Glaser, R., & Chudowsky, N. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (1999). On the cognitive validity of interpretations of scores from alternative concept mapping techniques. *CSE Technical Report 503.* Los Angeles, CA: Center for the Study of Evaluation; CRESST.

Schutz, P. A., & Davis, H. A. (2000). Emotions and self-regulation during test taking. *Educational Psychologist, 35*, 243-256.

Schutz, P. A., Davis, H. A., Schwanenflugel, P. J., & Axelrod, J. (1998, April). *Personal theories of emotion and emotional regulation during test-taking.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Snow, R. (1994). Abilities in academic tasks. In R. J. Sternberg, & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3-37). Cambridge: Cambridge University Press.

Stipek, D., Gralinski, H.J. (1996). Children's beliefs about intelligence and school performance. *Journal of Educational Psychology, 88*, 397-407.

Wolters, C. A. (1998). Self-regulated learning and college students' regulation of motivation. *Journal of Educational Psychology, 90*, 224-235.

Wolters, C. A., Yu, S. L., & Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences, 8*, 211-238.

Yepes-Baraya, M. (1995). *Task analysis of science performance tasks and items: Identifying relevant attributes*. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 388 676)

Table 1

Coding Categories and Definitions

Coding Categories	Definitions	Examples
Cognitive Engagement		
Declarative Knowledge	"Knowing that"; domain specific factual knowledge	"A water lily is a plant."
Procedural Knowledge	"Knowing how"; set of steps or actions that can be carried out to interpret diagrams or tables, manipulate tools, record data, and make calculations	"All of the x's are at 60 on the graph."
Schematic Knowledge	"Knowing why"; using multiple pieces of declarative knowledge, multiple pieces of procedural or multiple pieces of declarative and procedural knowledge to interpret data and diagrams, explain concepts, and make predictions	"Plants give off oxygen and the fish breathe the oxygen."
Strategy Use and Metacognition		
Strategic Knowledge	Rereading parts of the problem, rethinking how to solve the problem, stopping to organize thoughts, checking progress, correcting reading, planning	"First, I'll put water in the cups. Then I'll take the temperature of the water."
Awareness of Having Knowledge, Skills or Experience	Expressing awareness of having the knowledge, skills or prior experience needed to solve the problem	"We've worked on a problem like this before in class."
Awareness of Lacking Knowledge, Skills or Experience	Expressing awareness of lacking the knowledge, skills or prior experience needed to solve the problem	"I don't know how to solve this problem."
Help Seeking	Asking the interviewer questions about problem-solving content or procedures	"What does primary consumer mean?"
Perceptions and Feelings		
Item Difficulty	Description of test item difficulty	"This problem is hard."
Positive and Negative Feelings	Feelings about self related to mood or energy level.	"This is fun!"

Table 2

Average Percentage of Students Who Made Statements Related to Cognitive Engagement, Strategy Use and Metacognition on Multiple-Choice and Performance Assessment Items during Think-Alouds

	Multiple Choice (n_p = 28.33, n_{items} = 6)	**Performance Assessment (n_p = 27.50, n_{items} = 8)**
Cognitive Engagement		
Declarative Knowledge	22%	21%
Procedural Knowledge	45%	56%
Schematic Knowledge	25%	29%
Strategy Use and Metacognition		
Strategic Knowledge	44%	65%
Awareness of Having Knowledge/Skills	2%	2%
Awareness of Lacking Knowledge/Skills	11%	11%
Help-Seeking	9%	22%

Note. The number of students reported at the top of each column is the average number of students who completed each group of items (multiple choice or performance assessment).

Table 3

Average Percentage of Mastery and Performance Goal-Oriented Students Who Made Statements Related to Cognitive Engagement, Strategy Use and Metacognition on Multiple-Choice and Performance Assessment Items during Think-Alouds

	Mastery (n_{MC} = 5.00) (n_{PA} = 5.75)	**Performance** (n_{MC} = 3.00) (n_{PA} = 3.13)
Multiple Choice (n_{items} = 6)		
Cognitive Engagement		
Declarative Knowledge	17%	17%
Procedural Knowledge	30%	83%
Schematic Knowledge	20%	22%
Metacognitive Engagement		
Strategic Knowledge	60%	39%
Awareness of Having Knowledge/Skills	3%	0%
Awareness of Lacking Knowledge/Skills	13%	6%
Help-Seeking	10%	11%
Performance Assessment (n_{items} = 8)		
Cognitive Engagement		
Declarative Knowledge	22%	32%
Procedural Knowledge	50%	60%
Schematic Knowledge	30%	32%
Metacognitive Engagement		
Strategic Knowledge	87%	64%
Awareness of Having Knowledge/Skills	7%	0%
Awareness of Lacking Knowledge/Skills	11%	16%
Help-Seeking	26%	12%

Note. The number of students reported at the top of each column is the average number of students who completed each group of items (multiple choice or performance assessment).

Table 4

Students' Perceptions of Multiple-Choice Tests and Performance Assessments

	Multiple Choice	**Performance Assessment**
More challenging?	28.6%	71.4%
Requires more thinking?	23.8%	76.2%
More authentic?	14.3%	85.7%
More interesting?	14.3%	85.7%
Requires more effort?	10.0%	90.0%
More difficult?	23.8%	76.2%
More anxiety producing?	38.1%	61.9%
Shows what you know (more valid)?	42.9%	57.1%
Harder to know the right answer?	33.3%	66.7%
Easier to use to make comparisons among students?	61.9%	38.1%
Do better on?	70.0%	30.0%
Want grade based on?	57.1%	42.9%

Table 5

Average Percentage of Students Describing Perceptions and Feelings about Multiple-Choice and Performance Assessment Items during the Think-Aloud

	Multiple Choice (n_p = 28.33, n_{items} = 6)	**Performance Assessment (n_p = 27.50, n_{items} = 8)**
Positive Feelings	1%	4%
Negative Feelings	4%	4%
Items Described as "Hard"	1%	0%
Items Described as "Easy"	1%	0%

Note. The number of students reported at the top of each column is the average number of students who completed each group of items (multiple choice or performance assessment).

Table 6

Average Percentage of Mastery and Performance Goal-Oriented Students Describing Perceptions and Feelings about Multiple-Choice and Performance Assessment Items during the Think-Aloud

	Mastery (n_{MC} = 5.00) (n_{PA} = 5.75)	**Performance (n_{MC} = 3.00) (n_{PA} = 3.13)**
Multiple Choice (n_{items} = 6)		
Positive Feelings	0%	6%
Negative Feelings	3%	11%
Items Described as "Hard"	3%	0%
Items Described as "Easy"	0%	0%
Performance Assessment (n_{items} = 8)		
Positive Feelings	4%	8%
Negative Feelings	7%	4%
Items Described as "Hard"	2%	0%
Items Described as "Easy"	0%	0%

Note. The number of students reported at the top of each column is the average number of students who completed each group of items (multiple choice or performance assessment).

Table 7

Average Percentage of Students Describing Perceptions and Feelings about Multiple-Choice and Performance Assessment Items during Post-Test Interviews

	Multiple Choice (n_p = 22.00, n_{items} = 6)	**Performance Assessment (n_p = 20.00, n_{items} = 8)**
Positive Feelings	20%	16%
Negative Feelings	24%	13%
Items Described as "Hard"	16%	13%
Items Described as "Easy"	54%	41%

Note. The number of students reported at the top of each column is the average number of students who completed each group of items (multiple choice or performance assessment). The number of students for the post-test interviews may differ from those reported in the think-aloud analyses because not all students had time to complete the post-test interviews.

Table 8

Average Percentage of Mastery and Performance Goal-Oriented Students Describing Perceptions and Feelings about Multiple-Choice and Performance Assessment Items during Post-Test Interviews

	Mastery (n_{MC} = 3.33) (n_{PA} = 4.38)	**Performance (n_{MC} = 2.83) (n_{PA} = 1.88)**
Multiple Choice (n_{items} = 6)		
Positive Feelings	25%	18%
Negative Feelings	20%	18%
Items Described as "Hard"	35%	18%
Items Described as "Easy"	10%	76%
Performance Assessment (n_{items} = 8)		
Positive Feelings	17%	13%
Negative Feelings	14%	20%
Items Described as "Hard"	17%	13%
Items Described as "Easy"	57%	33%

Note. The number of students reported at the top of each column is the average number of students who completed each group of items (multiple choice or performance assessment). The number of students for the post-test interviews may differ from those reported in the think-aloud analyses because not all students had time to complete the post-test interviews.