On the Links Between Students' Motivational Patterns and Their Perceptions of, Beliefs About, and Performance on Different Types of Science Assessments: A Multidimensional Approach to Achievement Validation

CSE Technical Report 573

Angela M. Haydel and Robert W. Roeser
CRESST/Stanford University

July 2002

National Center for Research on Evaluation,
Standards, and Student Testing
Center for the Study of Evaluation
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Project 1.1 Models-Based Assessment: Individual and Group Problem Solving in Science
Project 3.1 Construct Validity: Understanding Cognitive Processes—Psychometric and Cognitive Modeling
Richard Shavelson, Project Director, CRESST/Stanford University

The work reported herein was supported in part under the Educational Research and Development Center Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U. S. Department of Education, and in part by the National Science Foundation (REC9628293).

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, the U. S. Department of Education, or the National Science Foundation.

PREFACE

In 1995, Richard E. Snow wrote in CRESST's proposal to the Office of Educational Research and Improvement that his previous work showed that "psychologically meaningful and useful subscores can be obtained from conventional achievement tests" (Baker, Herman, & Linn, 1995, p. 133). He went on to point out that these subscores represented important ability distinctions and showed different patterns of relationships with demographic, "affective" (emotional), "conative" (volitional), and instructional-experience characteristics of students. He concluded that "*a new multidimensional approach to achievement test validation* should include affective and conative as well as cognitive reference constructs" (italics ours, p. 134).

Snow (see Baker et al., 1995) left hints of what he meant by "a new multidimensional approach" when he wrote, "the primary objective of this study is to determine if knowledge and ability distinctions previously found important in high school math and science achievement tests occur also in other multiple-choice and constructed response assessments. . . . A second objective is to examine the cognitive and affective correlates of these distinctions. And a third objective is to examine alternative assessment designs that would sharpen and elaborate such knowledge and ability distinctions in such fields as math, science, and history-geography" (p. 133).

We, as Snow's students and colleagues, have attempted to piece together his thinking about multidimensional validity and herein report our progress on a research program that addresses cognitive and motivational processes in high school science learning and achievement. To be sure, if Dick had been able to see this project through to this point, it might well have turned out differently. Nevertheless, we attempted to be true to his ideas and relied heavily on the theoretical foundation of his work, his conception of aptitude (Snow, 1989, 1992).

Snow called for broadening the concept of aptitude to recognize the complex and dynamic nature of person-situation interactions and to include motivational (affective and conative) processes in explaining individual differences in learning and achievement. Previous results, using a mixed methodology of large-scale statistical analyses and small-scale interview studies, demonstrated the usefulness of a multidimensional representation of high school science achievement. We identified three distinct constructs underlying students' performance on a standardized test and sought validation evidence for the distinctions between "basic knowledge and reasoning," "quantitative science," and "spatial-mechanical ability" (see Hamilton, Nussbaum, & Snow, 1997; Nussbaum, Hamilton, & Snow, 1997). Different patterns of relationships of these dimensions with student background variables, instructional approaches and practices, and out-of-school activities provided the groundwork for understanding the essential characteristics of each dimension. We found, for example, that gender differences in science achievement could be attributed to the spatial-mechanical dimension and not to aspects of quantitative reasoning or basic knowledge and facts.

Our studies, reported in the set of six CSE Technical Reports Nos. 569–574,* extend the groundwork laid down in Snow's past research by introducing an extensive battery of motivational constructs and by using additional assessment formats. This research seeks to enhance our understanding of the cognitive and motivational aspects of student performance on different test formats: multiple-choice, constructed response, and performance assessments. The first report (Shavelson et al., 2002) provides a framework for viewing multidimensional validity, one that incorporates cognitive ability (fluid, quantitative, verbal, and visualization), motivational and achievement constructs. In it we also describe the study design, instrumentation, and data collection procedures. As Dick wished to extend his research on large-scale achievement tests beyond the National Education Longitudinal Study of 1988 (NELS:88), we created a combined multiple-choice and constructed response science achievement test to measure basic knowledge and reasoning, quantitative reasoning, and spatial-mechanical ability from questions found in NELS:88, the National Assessment of Educational Progress (NAEP), and the Third International Mathematics and Science Study (TIMSS). We also explored what science performance assessments (laboratory investigations) added to this achievement mix. And we drew motivational items from instruments measuring competence beliefs, task values, and behavioral engagement in the science classroom. The second report in the set (Lau, Roeser, & Kupermintz, 2002) focuses on cognitive and motivational aptitudes as predictors of science achievement. We ask whether, once students' demographic characteristics and cognitive ability are taken into consideration, motivational variables are implicated in science achievement. In the third report (Kupermintz & Roeser , 2002), we explore in some detail the ways in which students who vary in motivational patterns perform on basic knowledge and reasoning, quantitative reasoning, and spatial-mechanical reasoning subscales. It just might be, as Snow posited, that such patterns interact with reasoning demands of the achievement test and thereby produce different patterns of performance (and possibly different interpretations of achievement). The fourth report (Ayala, Yin, Schultz, & Shavelson, 2002) then explores the link between large-scale achievement measures and measures of students' performance in laboratory investigations ("performance assessments"). The fifth report in the set (Haydel & Roeser, 2002) explores, in some detail, the relation between varying motivational patterns and performance on different measurement methods. Again, following Snow's notion of a transaction between (motivational) aptitude and situations created by different test formats, different patterns of performance might be produced. Finally, in the last report (Shavelson & Lau, 2002), we summarize the major findings and suggest future work on Snow's notion of multidimensional achievement test validation.

---

* This report and its companions (CSE Technical Reports 569, 570, 571, 572, and 574) present a group of papers that describe some of Snow's "big ideas" with regard to issues of aptitude, person-situation transactions, and test validity in relation to the design of a study (the "High School Study") undertaken after Snow's death in 1997 to explore some of these ideas further. A revised version of these papers is scheduled to appear in *Educational Assessment* (Vol. 8, No. 2). A book based on Snow's work, *Remaking the Concept of Aptitude: Extending the Legacy of Richard E. Snow*, was prepared by the Stanford Aptitude Seminar and published in 2002 by Lawrence Erlbaum Associates.

ON THE LINKS BETWEEN STUDENTS' MOTIVATIONAL PATTERNS AND THEIR PERCEPTIONS OF, BELIEFS ABOUT, AND PERFORMANCE ON DIFFERENT TYPES OF SCIENCE ASSESSMENTS*

Angela M. Haydel and Robert W. Roeser
CRESST/Stanford University

Abstract

Snow conceived of performance as emerging from a transaction between a person, and all the relevant resources that she brings to a situation, and the situation itself. When confronted by a task, for example, the person cobbles together a "complex" (combination) of her cognitive and motivational resources to meet situational demands. This research sought to examine the link between different situational demands—in this case, three different types of science achievement tests (multiple choice, constructed response, and performance assessment)—and perceptions, beliefs and performance of high school students characterized by three well-established motivational patterns—intrinsic-mastery, ego-success, and academically-helpless. We compared students across the three motivational patterns with respect to their efficacy for working on the tests, their beliefs about whether these tests were valid measures of their science knowledge, and their observed performance on the multiple-choice and constructed response tests. We found that students varying in motivational pattern could be reliably distinguished by their (a) efficacy for working on multiple-choice and constructed response tests, (b) beliefs about the validity of multiple-choice and constructed response tests in revealing their science knowledge, and (c) performance on the multiple-choice and constructed response tests when quantitative and verbal ability are controlled for. However, motivational pattern did not relate to efficacy for or validity beliefs about performance assessments. Perhaps the flexible nature of performance assessments facilitates all students' goal pursuits and performance whereas multiple-choice and constructed response tests do not.

Snow (1994; see also Stanford Aptitude Seminar, 2002) posited that a person's performance was a function of a broad set of aptitudes and the affordances and constraints of a particular situation. In this person-situation transaction, the person cobbles together a combination of cognitive and motivational aptitudes—an

*An earlier version of this report was presented at the annual meeting of the American Educational Research Association in Seattle, Washington, in April 2001 under the title *Understanding the Links Between Students' Motivational Patterns, Perceptions of, and Performance on Different Types of Science Achievement.*

“aptitude complex”—for dealing with the situation. The purpose of this study, in broad terms, is to shed light on the link between aptitude complexes,—here, motivational patterns—and perceptions, beliefs, and performance in different test situations. However, little is known about how students’ motivation and perceptions of tests relate to their actual performance on them. Students’ test scores presumably reflect not only their knowledge of the subject matter, but also their motivation in the subject matter domain and their motivation for engaging with the assessment (Snow, 1994). Thus, including motivational variables and students’ perceptions of such assessments in studies of achievement may reveal more about a fuller spectrum of cognitive and motivational processes than has typically been done. More specifically, we examine the kinds of motivational resources that some students “bring to the task” of standardized achievement tests. We also describe what we believe different achievement test situations “bring out of the child” by the nature of the affordances they provide. We then describe an empirical study in which we examined the relation between students’ configurations of motivational beliefs and their perceptions of and performance on various types of standardized achievement tests.

What the Student Brings to the Test: A Person-Centered, Motivational Approach

In this study, we draw on the work of Dweck and her colleagues to examine how motivational beliefs in the domain of science may be related to high school students’ perceptions of and performance on science achievement tests. Dweck and colleagues have identified three motivational patterns that characterize different children in achievement situations. These patterns are comprised of children’s implicit theories of intelligence, their personal goals, and their competency beliefs (Dweck, 1986; Dweck & Leggett, 1988) and include what we will call an intrinsic-mastery pattern, an ego-success pattern, and an academically helpless pattern. A large body of research has shown that children and adolescents characterized by certain of these motivational beliefs, as well as the particular configurations of multiple motivational beliefs as defined by Dweck (1986) and Dweck and Leggett (1988), show differential patterns of behavioral engagement and performance in achievement situations (see Ames, 1992, Dweck & Leggett, 1988, and Nicholls, 1984, for full descriptions). In this paper, we identify subgroups of high school students who manifest one of the three distinct patterns of motivation described by Dweck and colleagues in the particular domain of science. That is, we take a person-centered approach to defining motivation in this paper by identifying subgroups of students characterized by particular *configurations of motivational beliefs.* We then compare students characterized by different motivational patterns in terms of their perceptions of and performance on science achievement tests. Such an approach can be contrasted with a variable-centered approach in which the relations among motivational beliefs and science achievement outcomes across all students are the focus of investigation. Below, we describe each of the motivational patterns identified by Dweck and her colleagues.

Intrinsic-mastery pattern. Students characterized by the intrinsic-mastery pattern believe that intelligence is malleable and thus can grow over time. Such a belief is hypothesized to predispose such individuals to adopting goals in which the growth of competencies, and thereby intelligence, is the aim. In sum, students characterized by an intrinsic-mastery pattern are hypothesized to engage in learning and achievement tasks as a means of improving their skills, competencies, and intelligence. There is some evidence that mastery-oriented individuals enjoy learning, seek out challenges, persist during difficulties, use adaptive problem-solving and learning strategies, and show continuing motivation to learn in a subject domain outside of formal learning settings (Dweck & Leggett, 1988; Pintrich, Roeser, & DeGroot, 1994; Wolters, 1998).

Ego-success pattern. Students characterized by what we call an ego-success pattern, in contrast, believe that intelligence is fixed. This belief is hypothesized to predispose such individuals to adopt goals in which the proving of one’s fixed ability or the hiding of one’s fixed inability is the aim. However, ego-success students are defined as those who, despite these beliefs, have high confidence in their abilities. Thus, such individuals are hypothesized to view achievement situations as opportunities not necessarily to improve skills and competencies, but rather to prove superior relative ability. Both the constituent beliefs and this motivational pattern have been linked to negative affect during learning and the use of non-optimal learning strategies, but did not necessarily result in diminished performance compared to mastery-oriented students (Dweck & Leggett, 1988; Pintrich, et al., 1994; Wolters, 1998).

Helpless pattern. Dweck (1986) described students characterized by a helpless motivational pattern as those who, like the ego-success students, believe that their intelligence is fixed. Again, this belief is hypothesized to predispose such individuals in achievement situations to the pursuit of goals in which proving ability or hiding inability relative to others is a central aim. What differentiates helpless from ego-success students is, in part, their lack of confidence in their abilities. Because these students lack confidence in their abilities, they are occupied with the goal of hiding their perceived sense of incompetence. Research has linked the motivational components of helplessness, as well as students characterized by the full spectrum of the helpless belief system, to increased negative affect in achievement settings, avoidance of challenge seeking, failure to persist on tasks in the face of challenge, and performance deficits (Dweck & Leggett, 1988).

The work of motivational theorists is useful in identifying what it is that students "bring to the tasks" associated with the classroom and school, including tasks like taking a standardized achievement test in science.

What the Task Brings Out of the Student

Motivational patterns can affect how individuals engage in academic situations; environmental factors also contribute to how a person will behave. Hamilton (1994) provided empirical evidence that assessments differ in terms of the affordances and limitations of the item formats in the domain of mathematics. Hamilton showed how the affordances of the item formats influenced students' preferences for working on items. Those students who preferred the multiple-choice test thought it was easier, believed that they were more likely to get the right answer by eliminating responses or working backwards from the possible choices, or believed that the format prevented them from making "stupid mistakes." In contrast, students who preferred the open-ended tasks enjoyed them because the items were challenging, different or "fun." In addition, many students preferred the extended tasks because they believed that the tasks allowed them to be creative and relate their ideas to the real world.

Intrinsic-mastery, ego-success, and helpless students may be better suited to perform in different types of test situations, given their differential approaches to achievement settings. The affordances and limitations of test situations can either facilitate or frustrate students' habitual goal pursuits in relation to performance in a given domain. As noted by Ford (1992), motivation and environment are intimately linked:

> [E]ffective functioning requires a motivated, skillful person whose biological and behavioral capabilities support relevant interactions with an environment that has the informational and material properties and resources needed to facilitate (or at least permit) goal attainment. If any of these components is missing or inadequate, achievements will be limited and competence development will be thwarted. (p. 69)

Snow (1994) too has argued for an approach to understanding students' achievement that includes cognitive, motivational, and situational variables. He posits that different learning and assessment situations may require the "assembly" of different abilities, skills, knowledge, and motive dispositions on the part of students if they are to be successful. In addition to requiring certain resources, we believe it may be fruitful to consider how certain assessment formats either afford or constrain the attainment of particular achievement-related goals. For instance, it may be that the highly structured, rather unambiguous testing situation provided by multiple-choice assessments affords ego-success oriented students a perfect opportunity for them to attain their goal of wanting to demonstrate their superior relative ability. In this case, we might also expect that ego-success students would also feel particularly efficacious in the presence of these kind of assessments. In addition, because such tests allow them to meet their achievement goals, they may be more likely to perceive such assessments as valid in terms of allowing them to demonstrate what they know in science.

In contrast, it may be that the less structured constructed response tests and the even more ambiguous performance assessments, in which students are required to bring more of their own skills, knowledge, and interests in the topic to bear on the assessment task, provide a better opportunity for intrinsic-mastery–oriented students to attain their goals of developing and expanding their knowledge and understanding. If this were the case, then we would also expect such students to feel more efficacious in working on constructed response test items and performance assessments, and also to perceive these kinds of assessments as valid in terms of allowing them to demonstrate what they know in science.

Finally, it may be that helpless students perceive all assessment situations with some trepidation because they represent the real possibility that such students' sense of incompetence relative to others will be revealed and reinforced. That is, standardized achievement tests may represent situations in which helpless students see their goal of hiding their inability relative to others as unattainable. This may cause anxiety or frustration that then interferes with performance. We hypothesized that such students will feel less efficacious in all types of assessment situations, and may not believe that any such situation allows a valid assessment of what they know in science.

We explored these hypotheses further by examining the relationship between students' general motivational patterns in the domain of science and their specific perceptions of multiple-choice items, constructed response items and performance assessments. We also examined students' performance on multiple-choice and constructed response items. In this study, high school students completed a survey including (a) Dweck's measures of implicit intelligence beliefs, goal orientation, and confidence beliefs, (b) questions about efficacy for working on multiple-choice items, constructed response items, and performance assessments, and (c) questions about their beliefs about whether each of these formats was a valid measure of science knowledge. The questionnaire data were used to determine high school students' motivational patterns in science. Motivational patterns in science were constructed based on students' responses to implicit intelligence beliefs, goal orientation, and confidence beliefs measures. One-way analyses of covariance were used to explain how motivational patterns related to students' responses on efficacy beliefs, perceived validity, assessment preference, and assessment performance measures.

Method

Participants

Of the total 491 students participating in the High School Study (see Lau, Roeser, & Kupermintz, 2002), we had adequate data for our purposes on 443 10th- and 11th-grade students (50.8% female). The 443 students were ethnically diverse: 48.6% European-American, 27% Latino, 7.9% African American, 7.9% Asian American, and 8.4% other ethnic groups. Approximately two thirds of their parents attended 4 or more years of college. Participants were drawn from courses in Earth science, biology, physics, and chemistry.

Instrumentation

We drew items from the Beliefs and Attitudes Toward Science Survey (see Lau et al., 2002; Shavelson et al., 2002) and constructed a variety of motivation measures (Table 1). Motivational patterns as defined by Dweck and her colleagues consisted of participant ratings on an implicit-theories-of-intelligence scale, a self-confidence-in-ability scale, and a goal-orientation item. These procedures and measures (implicit intelligence theories, goal orientation, self-confidence in intelligence) used to create "helpless," "ego-success," and "intrinsic-mastery" motivational patterns were based on previous work by Dweck and her colleagues (e.g., Dweck & Henderson, 1989; Henderson & Dweck, 1989; Hong, Chui, Dweck, Lin, & Wan, 1999; Quihuis, 1998). The Beliefs and Attitudes Toward Science Survey also included measures of students' efficacy for working on multiple-choice, constructed response, and

Table 1

Descriptive Statistics

	n	*M*	*SD*
Measures of motivational patterns			
Implicit Intelligence Theories Scale	438	4.54	1.039
Self-Confidence in Intelligence Scale	410	2.13	4.90
Measures of pretest assessment perceptions			
Multiple-Choice Efficacy Beliefs	438	6.63	1.90
Constructed Response Efficacy Beliefs	435	6.57	1.94
Performance Assessment Efficacy Beliefs	432	6.74	2.12
Multiple-Choice Perceived Validity	436	5.67	2.02
Constructed Response Perceived Validity	435	6.10	2.03
Performance Assessment Perceived Validity	432	6.79	2.31
Preference for Performance Assessments	439	3.46	1.30
Measures of achievement			
General Ability (math and verbal achievement combined)	406	10.43	3.16
Science Multiple-Choice Achievement	343	16.17	5.65
Science Constructed Response Achievement	342	4.22	1.99
Measure of posttest assessment perceptions			
Test Effort	358	6.24	2.36

performance assessment science items, measures of validity beliefs with respect to each of these assessment formats, and a measure of students' assessment format preference. On subsequent school days, several weeks after the completion of the survey, students completed mathematics, verbal, and science achievement tests. Immediately following the completion of the science achievement test, which took place several weeks later, students completed a posttest survey. The only item used in our analyses from the posttest survey is a measure of students' effort during test taking. (See Appendix for selected example items.)

Implicit intelligence theories scale. The implicit intelligence theories scale consisted of a set of three items that measure the extent to which individuals believe their intelligence in science is fixed or malleable (e.g., "I can't change how smart I am in science"). Participants were asked to indicate the degree to which they agreed with each statement using a 6-point Likert scale ranging from 1 (*strongly agree*) to 6 (*strongly disagree*). The reliability coefficient for the three-item scale was .68. Students who scored above or below the grand mean for the sample were categorized into incremental and entity groups, respectively. These procedures differ from those used by Dweck. In Dweck's studies, students were categorized into incremental and entity groups if they fell within one half of a standard deviation above or below the grand mean for the sample, respectively. The criterion for classifying students as incremental or entity was loosened for this study in order to include more students in the motivational groups. We believe this modification does not substantially change the meaning of these motivational groups.

Goal orientation. The goal orientation measure was a forced-choice question with four options. Students who chose “science problems that I’ll learn from, even if they’re so hard I’ll get a lot wrong” were classified as learning oriented. Students who selected any of the other three options (e.g., “science problems that are hard enough to show that I’m smart”) were classified as ego oriented. Thirty-eight percent of participants were classified as learning-oriented.

Self-confidence in intelligence scale. The self-confidence in intelligence scale consisted of four items created from a set of eight items in the survey measuring participants’ beliefs about their intellectual ability in science. For each of four items in the survey participants selected a statement indicating high confidence or low confidence (e.g., “I usually think I’m intelligent in science” versus “I wonder if I’m intelligent in science”). For whichever sentence participants chose, they rated how true the statement was for them. By combining the high or low confidence items selected, and participants’ ratings of how true the statement was for them, the eight items were transformed into four Likert scale items. The four-item self-confidence in intelligence scale had a reliability of .81. A median split was then used to classify participants as high or low confidence.

Efficacy for multiple-choice items. Participants were provided with six examples of multiple-choice questions that they did not have to answer. Following the examples of multiple-choice questions, participants were asked to rate how confident they were that they could correctly answer questions that were similar to the sample questions using an 11-point Likert scale ranging from 0 (*cannot do at all*) to 10 (*certain can do*). The items in the scale were based on Bandura’s (1986) work on self-efficacy. The six-item multiple-choice efficacy scale had a reliability of .87.

Validity beliefs about multiple-choice items. Participants were also asked to indicate how well they thought the multiple-choice example questions allowed them to demonstrate their knowledge of science using a 11-point Likert scale ranging from 0 (*not at all*) to 10 (*very well*). This validity beliefs question was asked three times following the sets of example problems. The three-item multiple-choice validity beliefs scale had a reliability of .76.

Efficacy for constructed response items. Participants were provided with six examples of constructed response questions that they did not have to answer. Following the examples of multiple-choice questions, participants were asked to rate how confident they were that they could answer questions that were similar to the sample questions using an 11-point Likert scale ranging from 0 (*cannot do at all*) to 10 (*certain can do*). The items in the scale were based on Bandura's (1986) work on self-efficacy. The six-item constructed response efficacy scale had a reliability of .90.

Validity beliefs about constructed response items. Participants were also asked to indicate how well they thought the constructed response example questions allowed them to demonstrate their knowledge of science using a 11-point Likert scale ranging from 0 (*not at all*) to 10 (*very well*). This validity beliefs question was asked three times following the sets of example problems. The three-item constructed response validity beliefs scale had a reliability of .83.

Efficacy for performance assessments. Participants were provided with two pictorial examples of performance assessments that they did not have to answer. Participants were asked to rate how confident they were that they could answer questions in hands-on assessments that were similar to the sample questions using an 11-point Likert scale ranging from 0 (*cannot do at all*) to 10 (*certain can do*). The items in the scale were based on Bandura's (1986) work on self-efficacy. This short, two-item performance assessment efficacy scale had a reliability of .90.

Validity beliefs about performance assessment items. Participants were asked to indicate how well they thought questions such as the example questions allowed them to demonstrate their knowledge of science using a 11-point Likert scale ranging from 0 (*not at all*) to 10 (*very well*). This performance assessment validity beliefs item had a mean of 6.8 and a standard deviation of 2.3.

Mathematics and verbal achievement. The mathematics test included multiple-choice items from the National Education Longitudinal Study ([NELS] 1988, 1990, 1992), and the verbal test included multiple-choice items from the Standardized Achievement Test (SAT). For the purposes of our analyses, the mathematics and verbal achievement measures were combined into a single measure called "general ability."

Science achievement test. The science test consisted of both multiple-choice and constructed response items selected from NELS:88, the National Assessment of

Educational Progress (NAEP), and the Third International Mathematics and Science Study (TIMSS).

Test effort. Following the science achievement test, students completed a posttest survey in which they were asked how hard they tried on the test. This item was answered on 11-point Likert scale ranging from 0 (*I didn't try at all*) to 10 (*I tried my hardest*).

Results and Discussion

We addressed three questions with our data: (a) What is the distribution of students across Dweck's three motivational groups? (b) What is the relation between these motivational patterns and perceptions and beliefs? (c) What is the relation between these patterns and science achievement?

Motivational Patterns

Using procedures similar to those outlined in Dweck and her colleagues' previous studies, we found that 188 participants (42.4% of the total sample) fit into Dweck's model. Of these 188 participants, 62 (33.0%) were classified as helpless, 65 (34.6%) were classified as ego-success, and 61 (32.4%) were classified as intrinsic-mastery. Following from Dweck's theory, students with an incremental view of intelligence and a mastery goal orientation were classified as intrinsic-mastery. Students with an entity view of intelligence, a performance goal orientation, and high confidence were classified as ego-success. And students with an entity view of intelligence, a performance goal orientation, and low confidence were classified helpless.

Girls were overrepresented in the helpless group (N = 43), χ^2 (df = 2, N = 185) = 13.67, p < .01. This finding supports previous research suggesting that bright girls are more likely to exhibit the helpless pattern (Dweck, 1986). Most of these girls were taking chemistry, physics, biology, and other advanced science courses. In addition, boys were overrepresented in the ego-success group (N = 40). This finding illustrates that boys in these kinds of science classes tend to believe that intelligence is fixed, have a performance orientation, and feel more confident than girls do about their abilities. There were no significant gender differences for students with an intrinsic-mastery pattern. Further analyses revealed that gender was not a significant predictor of perceptions of efficacy and validity, preference for different assessments, or performance on the multiple-choice science achievement test when prior achievement was controlled for. Because gender was not a significant predictor of any outcome, it was removed from subsequent analyses.

Relationship Between Motivational Patterns and Assessment Perceptions

Controlling for math and verbal achievement, the analyses showed that both the intrinsic-mastery and the ego-success students felt more efficacious about multiple-choice and constructed response items than did the helpless students. Contrary to hypotheses, the ego-success students felt more efficacious than the intrinsic-mastery students on the constructed response items, and there were no differences in the means among the three groups of students regarding efficacy for working on performance assessments (Table 2).

The pattern of results for validity beliefs is similar to that for efficacy (Table 3). Controlling for math and verbal ability, the analyses showed that both the intrinsic-mastery and ego-success students believed that multiple-choice and constructed response items are more valid than helpless students did. However, the lack of mean differences between groups on this measure is contrary to what we may have expected given the differential goal pursuits of the three motivational groups. For both the constructed response items and performance assessments, we would have

Table 2

Adjusted Means and *F*-Values for Efficacy Beliefs by Motivational Pattern Controlling for Math and Verbal Achievement

	Motivational pattern				*F*-values	
Efficacy beliefs adjusted mean scores	Intrinsic-mastery (n = 49)	Ego-success (n = 61)	Helpless (n = 50)	*df*	Motivational pattern (partial eta sq.)	Math / verbal achievement composite (partial eta sq.)
Multiple choice	6.57_a (2.04)	7.25_a (1.41)	5.85_b (1.53)	155	11.73*** (.13)	44.60*** (.22)
Constructed response	6.46_c (2.19)	7.22_b (1.39)	5.65_a (1.77)	154	12.01*** (.13)	29.78*** (.16)
Performance assessment	6.42 (2.21)	7.06 (1.56)	6.27 (2.31)	152	2.48 (.09)	14.33*** (.03)

Note. Means in the same row accompanied by different subscripts differ at $p < .05$ using the Byrant-Paulson generalized studentized range statistic to examine mean comparisons. Standard deviations for mean scores of each group are shown in parentheses below adjusted mean scores.

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 3

Adjusted Means and *F*-values for Perceived Validity by Motivational Pattern (Controlling for Math and Verbal Achievement)

Perceived validity adjusted mean scores	Motivational pattern: Intrinsic-mastery (n = 49)	Motivational pattern: Ego-success (n = 61)	Motivational pattern: Helpless (n = 50)	*df*	*F*-values: Motivational pattern (partial eta sq.)	*F*-values: Math / verbal achievement composite (partial eta sq.)
Multiple choice	6.14_a (1.89)	6.07_a (1.94)	5.08_b (1.88)	155	5.07** (.06)	5.44* (.03)
Constructed response	6.22_a (2.16)	6.69_a (1.56)	5.27_b (1.92)	154	8.29*** (.10)	14.98*** (.09)
Performance assessment	6.58 (2.44)	6.76 (1.70)	6.45 (2.31)	152	0.29 (.00)	10.71** (.07)

Note. Means in the same row accompanied by different subscripts differ at $p < .05$ using the Byrant-Paulson generalized studentized range statistic to examine mean comparisons. Standard deviations for mean scores of each group are shown in parentheses below adjusted mean scores.

*$p < .05$. **$p < .01$. ***$p < .001$.

expected intrinsic-mastery students, more so than the other groups, to believe that these measures are valid.

Despite differences in how students perceived the assessments, students exhibiting different motivational patterns did not indicate that they preferred one type of test when asked directly on the survey (Table 4).

Table 4

Adjusted Means and *F*-values for Test Preference by Motivational Pattern (Controlling for Math and Verbal Achievement)

Test preference adjusted mean scores	Motivational pattern: Intrinsic-mastery (n = 49)	Motivational pattern: Ego-success (n = 61)	Motivational pattern: Helpless (n = 50)	*df*	*F*-values: Motivational pattern (partial eta sq.)	*F*-values: Math / verbal achievement composite (partial eta sq.)
Preference for performance assessment	3.58 (1.21)	3.33 (1.14)	3.45 (1.47)	155	0.48 (.01)	0.18 (.00)

Note. Means in the same row accompanied by different subscripts differ at $p < .05$ using the Byrant-Paulson generalized studentized range statistic to examine mean comparisons. Standard deviations for mean scores of each group are shown in parentheses below adjusted mean scores.

*$p < .05$. **$p < .01$. ***$p < .001$.

Relationship Between Motivational Patterns and Assessment Performance

One-way analyses of covariance revealed that motivational pattern was significantly related to achievement on multiple-choice items from the science test when math and verbal achievement were controlled for (Table 5). As expected, ego-success and intrinsic-mastery students performed better than helpless students on the multiple-choice test; however, this difference is only significant between the ego-success and helpless students.

The same pattern of results did not occur for the constructed response items. First, a two-way analysis of covariance revealed that boys (M = 4.42) performed better than girls (M = 3.65) on the constructed response items, F (1, 111) = 6.67; $p <$ .05. However, gender did not explain as much of the variance in test scores as motivational patterns or math / verbal achievement (Table 6). In addition, ego-success students outperformed the intrinsic-mastery and helpless students, and this difference in scores was only significant between the ego-success and intrinsic-mastery students.

Effort During Test Taking

Here we examine the relation between motivational pattern and the effort students put into test taking (Table 7). Intrinsic-mastery students tried least hard and the difference in effort during test taking is statistically significant between the intrinsic-mastery and ego-success students. This pattern of differences in effort between students paralleled the results of performance on the constructed response items (see Table 6).

Table 5

Adjusted Means and *F*-values for Multiple-Choice Science Achievement by Motivational Pattern (Controlling for Math and Verbal Achievement)

	Motivational pattern				*F*-values	
Science achievement adjusted mean scores	Intrinsic-mastery (n = 37)	Ego-success (n = 52)	Helpless (n = 38)	*df*	Motivational pattern (partial eta sq.)	Math / verbal achievement composite (partial eta sq.)
Multiple choice	16.04_{ab} (5.96)	17.37_{a} (5.33)	14.45_{b} (4.56)	123	6.44** (.10)	127.64*** (.51)

Note. Means in the same row accompanied by different subscripts differ at $p < .05$ using the Byrant-Paulson generalized studentized range statistic to examine mean comparisons. Standard deviations for mean scores of each group are shown in parentheses below adjusted mean scores.

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 6

Adjusted Means and *F*-values for Constructed Response Science Achievement by Motivational Pattern (Controlling for Math and Verbal Achievement)

	Motivational pattern				*F*-values		
Science achievement adjusted mean scores	Intrinsic-mastery (n = 34)	Ego-success (n = 47)	Helpless (n = 35)	*df*	Motivational pattern (eta sq.)	Math / verbal achievement composite (eta sq.)	Sex (eta sq.)
Constructed response	3.35_b (2.17)	4.89_a (1.59)	3.59_{ab} (1.92)	114	4.90** (.08)	6.78* (.06)	44.04*** (.29)

Note. Means in the same row accompanied by different subscripts differ at p < .05 using the Byrant-Paulson generalized studentized range statistic to examine mean comparisons. Standard deviations for mean scores of each group are shown in parentheses below adjusted mean scores.

*p < .05. **p < .01. ***p < .001.

Table 7

Adjusted Means and *F*-values for Effort During Test Taking by Motivational Pattern (Controlling for Math and Verbal Achievement)

	Motivational pattern			*F*-values	
	Intrinsic mastery (n = 34)	Ego-success (n = 51)	Helpless (n = 37)	Motivational pattern (partial eta sq.)	Math / verbal achievement composite (partial eta sq.)
Effort during test taking	5.50_b (2.36)	6.92_a (2.29)	6.08_{ab} (2.22)	3.99* (.06)	0.059 (.00)

Note. Means in the same row accompanied by different subscripts differ at p < .05 using the Byrant-Paulson generalized studentized range statistic to examine mean comparisons. Standard deviations for mean scores of each group are shown in parentheses below adjusted mean scores.

*p < .05. **p < .01. ***p < .001.

Conclusions

The findings confirmed previous research indicating that adaptive motivational patterns, such as the intrinsic-mastery and ego-success patterns, relate to perceptions about assessments that are associated with better achievement. In addition, these motivational patterns were associated with more positive perceptions of multiple-choice and constructed response tests. When asked about their efficacy for performing on multiple-choice tests, intrinsic-mastery and ego-success students felt more confident than helpless students.

The differences in validity beliefs between the intrinsic-mastery, ego-success and helpless students on multiple-choice and constructed response problems can be explained by considering more deeply what it means to have a helpless motivational pattern. Ford (1992) suggested that the helpless pattern, much like the hopeless pattern, encompasses not only negative beliefs about one's abilities, but also negative beliefs about one's environment. The environment is not perceived as responsive to these individuals' needs. For this reason, perhaps, our helpless students considered these items to be less valid measures of their scientific knowledge.

Though the multiple-choice performance results were consistent with hypotheses, the results pertaining to performance on the constructed response items were surprising initially. We expected intrinsic-mastery students to outperform students classified as ego-success or helpless; however, the intrinsic-mastery students scored most poorly on these items. However, our results are less puzzling when one considers these constructed response items in the context in which they were completed. When constructed response items are placed in achievement test settings, showing how much one knows in comparison to others is the focus, rather than learning and developing one's knowledge of material.

Thus, it makes sense that intrinsic-mastery students would disengage in such settings and perform more poorly. In an achievement test situation, completing these items may be viewed less positively because they force these intrinsic-mastery students to put more effort into a task that seems meaningless to them.

The sum of these findings across all efficacy, validity, and performance measures suggests that neither multiple-choice nor constructed response items afforded helpless students the opportunity to pursue their goals. This finding was consistent with hypotheses. However, our findings also suggest that neither constructed response items nor multiple-choice items (to a lesser extent) afforded intrinsic-mastery students the opportunity to pursue their goals.

Unlike the findings for multiple-choice and constructed response items, there were no differences in helpless, ego-success, and intrinsic-mastery students' efficacy for or validity beliefs about performance assessments. This may suggest that the flexibility and affordances of performance assessments provided a "fit" with each type of motivational pattern. For example, the authenticity, in terms of being a more meaningful science activity, and open-endedness of performance assessments may enable intrinsic-mastery students to feel as if they can explore their ideas. Though helpless students were expected to feel less positively about performance assessments compared to the other groups, the "discovery" or "active-learning" feature of these assessments may help to engage helpless students rather then discourage them. Perhaps all students were less likely to consider performance assessments as invalid because their hands-on nature makes them appear to be a more authentic measure of science ability. Given that performance assessments afford students the opportunity to make choices and allow them to apply scientific knowledge towards understanding real-world events, it makes sense that students who had different motivational patterns did not differ in their responses to performance assessments. The flexible nature of these tests may make for a more meaningful, engaging, and comfortable testing environment for all students compared to more traditional tests.

However, this finding is not supported by students' responses to the test format preference question. Overall, motivational pattern did not relate to preference, and even within each motivation pattern, students appeared to favor performance assessments only slightly. We believe that this finding is a result of the phrasing of the question. Students were asked whether they preferred performance assessments in comparison to both multiple-choice and short-answer tests. This presents a problem because performance assessments are composed primarily of short-answer items, and ultimately we were not able to adequately distinguish preference for multiple-choice tests compared to constructed response tests, preference for multiple-choice tests compared to performance assessments, *or* preference for constructed response tests compared to performance assessments.

Implications

In summary, this study supports Snow's theory of situations by person-aptitude complexes. That is, it allows us to begin thinking about how students come to a testing situation with different goals, motives and beliefs and how the testing situation facilitates or frustrates the opportunity for these students to meet their goals. In addition, this research provides greater insight into the validity of assessments. Performance on different types of tests is influenced by both ability and motivation in a particular domain, a finding that, again, supports Snow's theory. This research also allows us to conceptualize motivation in a more situated manner.

Though several studies have examined how the classroom structure and teachers' goals influence students' motivation, our study expands this notion to consider the test situation as a context in which motivation may differ or be enacted in different ways depending on the item format.

In this study we explored some of these theoretical and practical issues with students who represent three "extreme" motivational groups as a first step towards empirically examining Snow's theoretical propositions. Future research in this area may consider how assessment formats afford or constrain the goal pursuits of students who are classified as having multiple goals. First, it may be the case the "multi-goal" students are more typical students, thus enabling us to establish a more authentic understanding about how the average student engages in particular assessments. It is also relevant to examine whether having multiple motivational patterns (for example a combination of intrinsic-mastery and ego-success patterns) provides some sort of "motivational insurance" so that when the testing situation does not facilitate the engagement of one motivational pattern, another one may be enacted.

Future research also must examine more deeply how students with different motivational patterns actually engage in these various assessment tasks. Unfortunately, the survey data do not allow for a deep enough understanding about what students think and feel when they are actually taking the assessments. In addition, because all of these students did not necessarily actually take the three types of assessments prior to completing the survey, it is difficult to determine whether students really understood how the assessments differ. Thus, their efficacy, validity, and preference judgments may vary somewhat after actually having taken the different assessments. Think-alouds and posttest interviews with students may provide deeper insight into intrinsic-mastery, ego-success, and helpless students' representations of the test items, the strategies they use, and explanations they give while completing the items, as well as the affect they experience while working on the items.

Finally, this study has important implications for the types of tests educators use to evaluate their students. Currently, there is a strong movement towards the use of performance assessments. This research suggests that performance assessments may be more accurate measures of students' science ability because they provide a more positive testing environment, regardless of students' motivational patterns. Like the intrinsic-mastery and ego-success students, helpless students perceived these performance assessments as valid measures of science ability and felt just as confident that they could successfully complete these assessments.

References

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

Ayala, C. C., Yin, Y., Schultz, S., & Shavelson, S. (2002). *On science achievement from the perspective of different types of tests: A multidimensional approach to achievement validation* (CSE Tech. Rep. No. 572). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Baker, E. L., Linn, R. L., & Herman, J. L. (1995). *Institutional grant proposal for OERI Center on Improving Student Assessment and Educational Accountability: Integrated assessment systems for policy and practice: Validity, fairness, credibility, and utility*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

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

Dweck, C. S., & Henderson, V. L. (1989). *Predicting individual differences in school anxiety in early adolescence.* (ERIC Document Reproduction Service No. ED310884)

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, CA: Sage Publications.

Hamilton, L. S. (1994). *An investigation of students' affective responses to alternative assessment formats*. (ERIC Document Reproduction Service No. ED376203)

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

Haydel, A. M., & Roeser, R. W. (2002). *On the links between students' motivational patterns and their perceptions of, beliefs about, and performance on different types of science assessments: A multidimensional approach to achievement validation* (CSE Tech. Rep. No. 573). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Henderson, V. L., & Dweck, C. S. (1989). *Theories of intelligence: Background and measures.* (ERIC Document Reproduction Service No. ED312057)

Hong, Y., Chiu, C., Dweck, C. S., Lin, D. M., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*, 588-599.

Kupermintz, H., & Roeser, R. (2002). *Another look at cognitive abilities and motivational processes in science achievement: A multidimensional approach to achievement validation* (CSE Tech. Rep. No. 571). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Lau, S., Roeser, R. W., & Kupermintz, H. (2002). *On cognitive abilities and motivational processes in students' science engagement and achievement: A multidimensional approach to achievement validation* (CSE Tech. Rep. No. 570). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328-346.

Nussbaum, E. M., Hamilton, L. S., & Snow, R. E. (1997). Enhancing the validity and usefulness of large-scale educational assessments. IV. NELS:88 science achievement to 12th grade. *American Educational Research Journal, 34*, 151-173.

Pintrich, P. R., Roeser, R. W., & DeGroot, A. M. (1994). Classroom and individual differences in early adolescents' motivation and self-regulated learning. *Journal of Early Adolescence, 14*, 139-161.

Quihuis, G. (1998). *Mexican-American students' academic motivation: Achievement-related beliefs contributing to adaptive or maladaptive achievement behavior and affect.* Unpublished manuscript, Stanford University.

Shavelson, R., & Lau, S. (2002). *Multidimensional validity revisited* (CSE Tech. Rep. No. 574). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Shavelson, R., Roeser, R., Kupermintz, H., Lau, S., Ayala, C., Haydel, A., & Schultz, S. (2002). *Conceptual framework and design of the High School Study: A multidimensional approach to achievement validation* (CSE Tech. Rep. No. 569). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Snow, R. E. (1989). Cognitive-conative aptitude interactions in learning. In R. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.), *Abilities, motivation, and methodology* (pp. 435-474). Hillsdale, NJ: Lawrence Erlbaum Associates.

Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist, 27*, 5-32.

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.

Stanford Aptitude Seminar [Corno, L., Cronbach, L. J. (Ed.), Kupermintz, H., Lohman, D. F., Mandinach, E. B., Porteus, A. W., & Talbert, J. E.]. (2002).

Remaking the concept of aptitude: Extending the legacy of Richard E. Snow. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Appendix

Implicit Intelligence Theories Scale (α=.68)
How well I do in science depends on how much intelligence I was born with.
I can't change how smart I am in science.
I have to be really smart to do well in science.

Goal Orientation Measure
I would probably choose (circle only one answer)

Science problems that aren't too hard, so that I don't get many wrong.
Science problems that I'll learn from, even if they're so hard that I'll get a lot wrong.
Science problems that are fairly easy so I'll do well.
Science problems that are hard enough to show that I'm smart.

Self-Confidence in Intelligence Scale (α=.81)
Choose *one* of these:

I usually think I'm intelligent in science.
I wonder if I'm intelligent in science.

How true is this sentence for you?

Choose *one* of these:

When I get new science school work, I'm usually sure I will be able to learn it.
When I get new science school work, I often think I will not be able to learn it.

How true is this sentence for you?

Choose *one* of these:

I'm not very confident about my intellectual ability in science.
I feel pretty confident about my intellectual ability in science.

How true is this sentence for you?

Choose *one* of these:

I'm not sure I'm smart enough to be successful in science.
I'm pretty sure I'm smart enough to be successful in science.

How true is this sentence for you?

Multiple-Choice Efficacy Scale (α=.87)
When you take a multiple-choice test in science, how confident are you that you can correctly answer questions like these that:

ask about your knowledge of basic science facts?
ask you to think about basic science facts?
ask you to work through scientific problems?
ask you to do some calculations?
ask you to interpret graphs, figures, or other kinds of diagrams?
require that you visualize objects?

Constructed Response Efficacy Scale (α=.90)
When you take a short-answer test in science, how confident are you that you can correctly answer questions like these that:

ask about your knowledge of basic science facts?
ask you to think about basic science facts?
ask you to work through scientific problems?
ask you to do some calculations?
ask you to interpret graphs, figures, or other kinds of diagrams?
require that you visualize objects?

Performance Assessment Efficacy Scale (α=.90)
When you do a hands-on investigation in science, how confident are you that you can correctly answer questions like these that:

ask you to plan how to solve scientific problems?
ask you to actually carry out an investigation to solve the scientific problem?