The Concept of Aptitude and Multidimensional Validity Revisited

Robert W. Roeser
School of Education
Stanford University

Richard J. Shavelson
School of Education
Stanford University/CRESST

Haggai Kupermintz
School of Education
University of Colorado at Boulder/CRESST

Shun Lau, Carlos Ayala,
Angela Haydel, Susan Schultz,
Larry Gallagher, and Gisell Quihuis
School of Education
Stanford University

Richard E. Snow, reasoning from his new conception of aptitude, advocated a multidimensional approach to validating the construct of academic achievement. We briefly overview Snow’s approach and then summarize evidence from this special issue in 3 themes: (a) the multidimensional structure of science achievement, (b) the incremental predictive validity provided when both cognitive and motivational (affective and conative) constructs are used to model individual differences in achievement, and (c) the co-contributions of ability, motivational orientations, and characteristics of achievement test situations to performance differences. Overall, our studies confirmed or established (a) a multidimensional structure of science achievement scores, (b) the validity of several key motivational constructs for predicting science achievement among high school students, (c) systematic variation in relations be-

Requests for reprints should be sent to Robert W. Roeser, School of Education, Cubberley Hall, 485 Lasuen Mall, Stanford University, Stanford, CA 94305–3096. E-mail: roser@stanford.edu
tween motivational and general-ability constructs and science reasoning dimensions, assessment type, and achievement behavior (performance and anticipated choice), and (d) how alternative assessment methods (constructed response and performance assessment) shed light on the notion of multidimensional validity.

On the way to the office today, the first author stopped by a local middle school in which he was conducting a study on students' motivation to learn. It was standardized testing week in California, and students were cloistered in classrooms throughout the day, filling in bubble after bubble on a statewide achievement test. In conversation with an assistant principal overseeing the testing, it came to light that he had concerns about the validity of these exams. First, he worried that some students could only get through 15 min of the 50-min exam because they hadn't learned the material—not really a concern about test validity as students were demonstrating what they did and did not know. Second, and problematically, he worried that students' attention spans wandered, that some became disengaged when faced with challenging questions, and that still others seemed to guess randomly. Finally, we discussed how odd it was that such tests rarely resulted in feedback to students about their own knowledge—creating in essence a test with no discernible purpose in the eyes of the test taker. As we were discussing these issues, one of the test proctors emerged from a classroom incredulous that she had just witnessed some students simply marking patterned responses (e.g., A, D, A, D, A, D) rather than reading the questions.

Although our comprehension of how students' demographic backgrounds, knowledge and ability, attention and engagement, test-related perceptions and motivation, and outright guessing may affect their standardized test scores is still relatively undeveloped, this "just-in-time" life example highlights both the practical and scientific importance of considering the full range of ordered and random social, situational, and psychological processes that may contribute to students' achievement test scores. As standardized tests assume higher and higher stakes in U.S. society, a better understanding of their validity seems essential (e.g., Merrow, 2002).

The development of such an understanding was at the heart of Richard E. Snow's work, and the articles in this special issue take several steps toward increasing our understanding of aptitude complexes and the multidimensional validity of achievement tests. By and large, the aggregate results of the articles suggest that these tests measure more than what students know and can do. Though cognitive abilities were very important in predicting levels of test performance, we also found that motivational processes, perceptions of tests, and test formats were linked to individual differences in performance outcomes. Furthermore, we found that some of the most consequential outcomes of education—commitment to intellectual activities outside of school and the anticipated commitment, in this case, to study science and pursue science-linked careers—were not just related to the
knowledge and abilities of youth that are the focus of many of the standards-driven reforms today. We also found that such achievement outcomes were linked to youths' perceptions of their competencies and their valuing of the domain of science above and beyond their abilities and family background. In this final article, we discuss the overall findings of the articles in this special issue, extract general themes that address the "big ideas" Snow left us, and propose some new directions in the work on aptitude complexes and multidimensional test validity.

ON THE MULTIDIMENSIONAL STRUCTURE OF SCIENCE ACHIEVEMENT

Snow and colleagues (Hamilton, Nussbaum, & Snow, 1997; Nussbaum, Hamilton, & Snow, 1997) found evidence of a multidimensional structure to the science achievement measures used in the National Education Longitudinal Study of 1988 (NELS:88). They established three underlying dimensions, which they called basic knowledge and reasoning, quantitative science, and spatial–mechanical reasoning. Just as important, student demographic characteristics, prior science experience (course taking, extracurricular activities), and motivation correlated with these three dimensions in different ways. For example, they found a strong gender effect on spatial–mechanical reasoning subscores, but not on the basic knowledge and reasoning or quantitative science subscores. Curious, Snow (1995) set out to test these findings in a new study, writing that 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 also occur also in other multiple-choice and constructed response assessments. A second objective was to examine the cognitive and affective correlates of these distinctions. And a third objective was to examine alternative assessment designs that would sharpen and elaborate knowledge and ability distinctions in fields such as math, science, and history–geography (Snow, 1995, p. 17; see also Baker, Linn, & Herman, 1995, pp. 131–136). The High School Study was designed and conducted to carry on and extend this aspect of Snow’s work, and the results of this study are reported in the articles that make up this special issue.

We did not accomplish all that Snow imagined or would have accomplished himself. Rather, we focused on science achievement and the cognitive ability and motivational correlates of that achievement. We measured achievement with a multiple-choice test composed of items from NELS:88, the National Assessment of Educational Progress (NAEP), and the Third International Mathematics and Science Study (TIMSS), with a constructed response test composed of TIMSS items, and with three performance assessments, each selected and/or constructed to tap, predominantly, one of the three achievement dimensions: basic knowledge and reasoning (BKR), quantitative science (QS), or spatial–mechanical reasoning (SM). We included measures of cognitive ability—verbal, mathematical, spatial, and
fluid—as well as measures of affect and conation (e.g., motivation). We turn now to a brief summary of what we found out about the multidimensional validity of science achievement and the cognitive and motivational correlates of achievement in the empirical articles in this special issue.

EMPIRICAL FINDINGS

In the first empirical article of this special issue, Ayala, Shavelson, Yin, and Schultz set out to determine whether the reasoning dimensions identified by Snow and colleagues (Hamilton et al., 1997) were unique to the NELS:88 multiple-choice test or could be found in other multiple-choice tests and also in performance assessments. Results of confirmatory factor analyses showed evidence for these three dimensions for a test composed of NELS:88, TIMSS, and NAEP multiple-choice items, as well as for the subset of NELS:88 items used on our achievement test. The researchers then administered three reasoning-dimension-linked performance assessments to a subsample of 35 students from the main study. Assessing students’ think-aloud reasoning as they moved through these tasks, Ayala et al. found that performance assessment items correlated moderately with each other and NELS:88 reasoning scores. In short, there was some evidence that the two methods partially converged on these three reasoning dimensions. However, the relationship between performance and multiple-choice scores scattered across the dimensions, most likely due to the broad reasoning and knowledge spectrum tapped by the relatively open-ended structure of the performance assessments. The suggestion arising from these results is that there exist diverse reasoning pathways to the same performance outcome—a basic concept in systems theory called “equifinality” (many paths, equal end).

The second empirical piece, Kupermintz built on the earlier taxonomic efforts of Snow, Corno, and Jackson (1996) to describe hierarchies of affective and conative processes that were linked to the three reasoning dimensions identified by Snow and colleagues (Hamilton et al., 1997). He provided a profile of partial correlations between students’ standardized multiple-choice and constructed response test scores and various affective and conation scores after accounting for students’ general ability and demographic characteristics. He ordered affective and conative constructs in relation to three distinct levels of generality: domain-specific (e.g., in relation to the domain of science), task-general (e.g., in relation to standardized tests in general), and situation-specific (e.g., in relation to a specific test). Overall, these findings demonstrated the importance of considering motivational resources at different levels of generality in relation to performance on academic tasks, as Snow expected. Grades in science, for instance, were most strongly related to domain-general motivation and engagement in science class rather than to the more test-specific constructs. In addition, students who reported high value and interest in science were more likely to achieve high scores on a multiple-choice test—especially on the basic
knowledge and spatial–mechanical dimensions. Kupermintz suggested that valuing of science is critical to learning outside of school and that such “informal learning” may lead to outcomes such as BKR and SM that reflect knowledge and skills less critically associated with classroom instruction.

The third empirical article, by Lau and Roeser, builds on Snow’s (1989) idea of two parallel pathways that described psychological-level contributions to achievement outcomes—a performance and a commitment pathway. The goal of this study was to examine how cognitive and motivational factors associated with these two hypothesized pathways, respectively, contributed incrementally to the prediction of achievement outcomes that included science test performance, classroom grades, and anticipated choices, of science majors in college and science-related careers later in life. Using hierarchical regression analyses to test a series of paths flowing from demographic, ability, motivational, and situational engagement predictors to these outcomes, they found that students’ cognitive abilities were the strongest predictor of their performance in science as measured by standardized test scores; that motivational processes enhanced the predictive validity for science test scores and grades beyond the variance accounted for by ability; and that motivational processes were the strongest predictors of students’ commitment to science in the form of situational engagement and anticipated choices of science-related college majors and careers. Furthermore, there was evidence that situational engagement in different situations—in a test, in the classroom, and with science outside of school—provided one set of pathways by which psychological resources of the self (e.g., aptitude resources) related to achievement outcomes—like performance and anticipated choice.

In the final empirical article by Haydel and Roeser, person-centered analyses were used to address aspects of Snow’s idea that differences in the “fit” between aptitude resources and situational characteristics were related to differences in test performance. Person-centered analyses accord nicely with Snow’s theoretical notion that the organization of individuals, aptitudes in relation to situational demands is critical to understanding differences in their achievements. Such analyses focus on subgroups, defined by (organized) profiles of variables, as the unit of focal interest. Subgroups were created from Dweck’s (1986) motivational typology. Students were characterized as “mastery oriented,” “ego oriented,” “helpless,” or “unclassified” by this typology. The groups were then compared on their demographic composition and their test-related perceptions, engagement, and performance on multiple-choice and constructed response items.

The basic argument of this article was that the characteristics of achievement tests, especially closed-ended multiple-choice tests, provided a good fit with the motivational dispositions of ego-oriented students (e.g., motivated by a sense of science intelligence as stable, high confidence in their science intelligence, and a desire to prove this ability relative to others); less of a fit with the motivational dispositions of mastery-oriented students (e.g., motivated by a sense of science intelligence as mal-
leable and a desire to improve their science knowledge and skills); and a poor fit or a “mismatch” with the motivational dispositions of helpless students (e.g., motivated by a sense of science intelligence as stable, low confidence in their science intelligence, and a desire to avoid demonstrations of incompetence). After controlling for ability, the hypothesis was that motivational patterns would be linked to differences in test engagement and performance, with ego-oriented students doing the best, helpless students doing the worst, and mastery-oriented and unclassified students falling somewhere in between.

Haydel and Roese’s results provided only limited support for their hypothesis. Consistent with hypotheses, helpless students were the most test-anxious and did worse on the science achievement test compared to students characterized by an ego orientation after controlling for level of cognitive ability. In addition, female students were more likely than male students to be characterized by a helpless orientation, whereas male students compared to female students were more likely to be characterized by an ego-oriented pattern. Unexpectedly, students characterized by a mastery orientation did slightly better than the ego-oriented students on the test, and they reported greater engagement with the test than other groups. Overall, these results suggested the possibility that performance differences, after accounting for ability, were in part due to the fit of the ends of the testing situation with individuals’ capability beliefs and personal achievement goals. However, these results also raised the possibility that students whose motivational orientation did not necessarily fit with that of the situation (mastery-oriented students) nonetheless may engage with and perform well on such tests. Once again we see the notion of equifinality in these results—that there exist multiple person-centered profiles of motivation that can predispose toward the same performance outcomes. It may well be that some individuals adapt and “goal switch” in relation to situational demands in testing situations, even if such situations do not “match” certain broad-band motivational or personality dispositions of the individual outside of the testing situation.

In the next section, we reflect on the mosaic of findings from these studies in relation to some of the overarching themes in Snow’s research program (see Shavelson et al., this issue). These themes include (a) the multidimensional nature of achievement, (b) the transaction between person and situation in shaping individual differences in performance, and (c) a multidimensional approach to construct validation.

THREE THEMES

Multidimensionality of Achievement

In his new concept of aptitude, Snow (1992) included motivational processes in explaining individual differences in achievement. More specifically, he posited two general pathways to describe the manner in which these cognitive and moti-
vational resources played out (Snow, 1994; see also Corno et al., 2002). The first was what he called a “performance pathway”—a concept that denoted the dynamic process by which cognitive resources were activated, retrieved, assembled, and executed to accomplish particular tasks in particular situations. The other, parallel pathway was the “commitment pathway”—a concept that denoted the process by which motivational resources were activated to energize and guide behavior toward particular goals in a given situation.

Consistent with Snow’s theory, Lau and Roeser (this issue) found that motivational variables improved the prediction of science achievement scores after accounting for students’ cognitive ability and demographic characteristics. Specifically, students’ self-efficacy and task values had direct links to science achievement, as well as indirect links through the construct of situational engagement. These authors noted that one of the main touch-points of these two hypothesized pathways is students’ perceived competence, which is derived from their own abilities, but also from social experiences (e.g., Bandura, 1997). Kupermintz (this issue) also found significant partial correlations between different reasoning dimensions in science test performance and a number of motivational variables, after accounting for students’ cognitive ability and demographic characteristics. It is also important to note that person-level aptitudes were often stronger correlates of science achievement than social address characteristics of the students in this sample (gender, parental education, and ethnicity).

It is interesting to extend Snow’s thinking concerning the spectrum of psychological processes along the performance and commitment pathways that “show through” in visible behaviors by thinking about how “achievement” includes both behavioral commitments and performances, choices and test scores (see also Eccles, 1984). Any teacher, educator, or parent knows that a young person’s engagement in the classroom (e.g., doing homework, paying attention), on a test (e.g., mental and emotional investment of energy), or in free-time intellectual activities (e.g., watching a science-related television program) are all achievements in their own right. Results from Lau and Roeser (this issue) suggest that such achievements are affected indirectly by individuals’ cognitive ability (through their perceived competence) and directly by motivational processes such as their perceived competence, interest in, and valuing of the subject matter. These kinds of engagement, in conjunction with adolescents’ cognitive and motivational aptitudes, also relate to more distal forms of what are commonly known as achievements—performances on tests and in the school grade book. In addition, they relate to another form of achievement not commonly considered, intentions to study science in college, major in a science, and seek a science-related career in the future. Such intentions, research has shown, are (probabilistically) critical for the actualization of such achievement choices later in life, even though not all such intentions bear fruit with the passage of time (Eccles, 1993). Perhaps most interesting is how commitments and performances
interact reciprocally over time—with long-term commitment necessary to develop expertise in a domain, and with expertise fueling future commitment (Ericsson & Charness, 1994; Ericsson & Lehman, 1996).

In summary, one of the interesting ideas that arise from Snow’s model of the performance and commitment pathways is that it is important to conceive not only of the manifold psychological processes that eventuate in achievement, but also of the varied, consequential, proximal and distal behaviors that define the landscape of achievement in a given intellectual domain. Perhaps by becoming more explicit about the array of desirable achievement outcomes that include but transcend performance on tests, educational discussions of curricular and instructional reforms could be enriched—for instance, how can we think about reforms that not only enhance performance and learning, but also cultivate individuals’ long-term commitments to, and free-time choice of, intellectual activities (e.g., Meier, 2000).

Person by Situation Transactions in Achievement

A central tenet of Snow’s aptitude concept was that achievement is the result of the quality of attunement between aptitudes on the person side and demands and opportunities on the situational side. He (1994; see also Corso et al., 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, a person cobbles together a combination of cognitive and motivational aptitudes—an “aptitude complex”—for addressing relevant task- and situation-specific goals (e.g., performance).

Several findings from both the Kupermintz and the Lau and Roesser articles (this issue) provide empirical support for the notion of considering carefully the level of specificity of the processes in the person and the situation that are most likely to be consequential for a particular kind of (situated) outcome (e.g., test performance, grade, choice of a free-time activity). For instance, both articles provided evidence that classroom-level measures of engagement were most closely linked to classroom grades, whereas test-specific-level measures of engagement were most closely linked to achievement test scores. Motivational processes, like cognitive abilities, exist at different levels of specificity in relation to domains, situations, and time scales. Clearer conceptualization and measurement of narrow- to broad-band motivational processes is one implication of this work that is already receiving attention in the field (e.g., Elliot & Church, 1997), though infusing a focus on narrow- to broad-band motivational processes with a nuanced view of achievement situations is as of now still underdeveloped.

At a slightly deeper level of analysis, Snow (1994) was searching not only for a viable taxonomy of both proximal and distal, narrow- and broad-band cognitive, and affective–conative constructs relevant to achievement, but also for a
means of conceiving of the dynamic relation between such constructs and features of achievement situations that resulted in variation in outcomes. He followed the path of many systems-oriented theorists who have been interested in the development of human motivation, abilities, and personality in context (e.g., Bowlby, 1988; Eccles & Midgley, 1989; Lazarus, 1991; Pervin, 2000; Ryan, 1992; Tharp & Gallimore, 1988). In his own work, Snow (1994) arrived at notions of attunement, or its lack, as a means of describing the dynamic interplay between characteristics of persons and characteristics of situations. That is, he proposed that successful task performance resulted when individual resources and situational demands were more or less in harmony, or attuned, or fit with one another (or when an individual or an instructional treatment could “adapt” to the other in real time); and that unsuccessful task performance resulted when the two were in conflict, or ill-attuned, or “mismatched” with one another (or when an individual or an instructional treatment was unable to adapt to the other in real time). This idea is reminiscent of Vygotsky’s (1978) zone of proximal development, but it is viewed from both sides (e.g., transactionally). To the extent the resources of the individual do not extend into the zone of possible task solutions, successful performance will not occur; and to the extent that the characteristics of the task do not extend into the zone of proximal development of the learner, successful performance will not occur.

In Haydel and Roesser’s article (this issue), this notion of attunement was conceptualized in two ways—attunement between the perceived capability of the person and the kinds of task demands represented by different item formats; and attunement between the kinds of domain-general achievement goal orientations that students “bring to” testing situations and the kinds of opportunities for goal fulfillment afforded by those situations. There was some evidence that those who would be expected from a cultural and motivational psychological standpoint to do best on multiple-choice-like tests did in fact perform at a (relatively) high level—White male students who had high confidence in their science ability and who were interested in proving their superior ability relative to others. Unlike female students and members of traditionally targeted minorities in the United States, male students who are White do not have to navigate cultural stereotypes about their presumed inability to perform well on intellectual tests, especially in domains like math or science. Thus, notions of “stereotype threat” do not drain off these students’ motivational resources when they are confronted with achievement test situations (e.g., Steele, 1997). In this sense, such tests fit with their sex and ethnic background, as it were. Similarly, those persons with positive perceptions of their capability, net of their actual ability, do not labor under anxieties about performing in such situations and thereby can devote more attentional resources to the task at hand (e.g., Dweck & Wortman, 1982). In this sense, their perceived capabilities fit with tests in which items get progressively harder. Finally, ego-oriented individuals may see tests as a valid opportunities to
meet their habitual goals of wanting to prove their ability relative to others, especially because standardized tests are rarely, if ever, used for formative feedback.

All of these kinds of fit or attunement between person and situation were found in mirror image in the situation of the helpless students. Here we found evidence for perhaps gender-linked and motivational mismatches that were associated with poorer performance on the multiple-choice items. For instance, females were over-represented in this group, raising the possibility that implicit stereotype threats about their inability in science contributed to their poor test scores (Steele, 1997). In addition, these students were characterized by the highest levels of test anxiety and lowest levels of perceived capability, suggesting that attentional resources were likely to be devoted to performance worries as much as to the task at hand (Dweck & Wortman, 1982). Finally, these students held a domain-specific achievement goal orientation in which hiding inability relative to others was the aim—an unlikely goal to be fulfilled in a norm-referenced testing situation. All of these processes represent ways in which testing situations can be ill-attuned to person characteristics and thereby, potentially, undermine performance.

Another interesting finding from Haydel and Rooser was that the mastery-oriented students, whose achievement goal orientations in science seemed incongruous with, or at least irrelevant to, standardized testing situations, did well on and engaged the most deeply with the science test compared to other students. This finding may provide an example of how important it is to bring together Snow's ideas on transactions. That is, this finding raises the possibility that it is critically important to identify the right level of processes to consider in relation to the relative attunement of the person and the situation. Students who were "mastery oriented" in the domain of science may well have been able to switch to situation-specific ego performance goals when testing, or at least to "adapt" their motivational resources to the characteristics of the situation they were in by "blending" their pursuit of mastery and ego goals in that situation. This in turn may have provided a form of motivational insurance (Ford, 1992) that allowed for striving for superior performance and engaging deeply in the content of the exam simultaneously. Mastery-oriented students in fact received slightly better scores, on average, than any other group. By using more dynamic measurement procedures, teasing apart the relative importance of conative and affective constructs at different levels of analysis in the relative fit or mismatch between person and situation may be possible (e.g., Haydel, 2002). This in turn could lead to more nuanced understanding of person-situation transactions in differences in achievement.

**Multidimensional Approach to Construct Validation**

Snow wanted to extend the notion of multivariate achievement by incorporating into a measure of science achievement not just multiple-choice items but also constructed response (open-ended) items and performance assessments. Ex-
tending previous work, Ayala et al. (this issue) examined the construct validity of the achievement dimensions (BKR, QS, and SM) in multiple-choice and performance tests. They found evidence of convergence for multiple-choice and performance tests on the dimensions. However, as expected, the dimensional complexity increased from multiple-choice to performance items. Although each of the three performance assessments was selected to tap, predominantly, one of the three dimensions, they proved to be quite complex, each drawing on all three dimensions to a greater or lesser extent. In the end, what was reinforced was the multidimensional character of achievement test scores.

The three reasoning dimensions capture part of that multidimensionality that we have seen in this study. But there are other aspects of achievement that can be captured only by incorporating a wider variety of achievement tests based on alternative frameworks. For example, Li and Shavelson (2001; see also Shavelson & Ruiz-Primo, 1999) distinguished among declarative ("knowing that"), procedural ("knowing how"), schematic ("knowing why"), and strategic ("knowing when") science knowledge. They have provided both cognitive (think-aloud) and factor analytic evidence that links different kinds of science achievement items to at least the first three of these kinds of knowledge. Without doubt, the reasoning dimensions underlie the use of these types of science knowledge in achievement situations, but perhaps with a richer set of multiple-choice, constructed response, and performance assessment items, the multidimensionality Snow saw, would be more comprehensively described. With this expanded definition and measure of science achievement would come new studies helping us understand the role of students' demographic, experiential, and motivational characteristics in the achievement-situation interaction.

**FUTURE DIRECTIONS**

Snow's new concept of aptitude is challenging in the same way the old interaction model of behavior as a function of the person and situation was challenging (Lewin, 1936). The aptitude complex that a person cobbles together in performing a task depends on both the person's psychological characteristics and personal biography and the affordances and constraints of the situation. If this weren't complicated enough, aptitudes must change as an individual progresses through a task and, thus, so too does the nature of the affordances and constraints change as an individual moves through the problem space.

Mapping this set of dynamics requires a level of methodological sophistication that seems daunting. Needless to say, our studies have only scratched the surface of Snow's new concept of aptitude and its implications for validating interpretations of achievement test scores. Moreover, they were, not surprisingly, limited methodologically. The research design and instrumentation developed in
this study were, indeed, a first attempt to operationalize Snow’s ideas. We ambitiously combined measures of aptitude, achievement, motivation, and demographics in the study. In the end, reality sobered grand vision. Access to schools and the inevitable absences and missing data reduced sample sizes on different instruments so that, from one analysis to another, sample sizes vary and are not always the size we intended or desired. Due to time limitations, some instruments had fewer items than we would have wished with the result being lower reliability than intended. Moreover, we modeled the data in a wide variety of ways, and consequently, many hypotheses were tested, increasing the likelihood of concluding that an effect exists when it may not have been there. With the hope that others learn from our work and improve on it, here we briefly suggest a few possible directions for future studies.

Our study of Snow’s aptitude concept took a distal approach. We studied patterns of relationships between variables across all students and patterns of variable relationships within subgroups of like students to attempt to infer something about aptitude complexes, their associations with performance and choice, and the meaning of achievement test scores. Needless to say, more and better studies are needed. Such studies could refine the variables we have used and add new ones to test specific combinations of cognitive, conative, and affective variables on complex achievement measures varying in what Snow called “referent generality”: immediate (e.g., multiple choice, constructed response, performance assessment), intermediate (extracurricular activities), and distal (e.g., choice of academic major, career).

In addition, such studies could adopt the kind of person-centered approach employed by Haydel and Roese (this issue). How do individuals (if they do at all) with similar aptitude profiles engage in and perform tasks, and how do they differ from individuals with other distinct profiles? Are there long-term consequences for the personal educational careers of students with different profiles (e.g., Quihuis, 2002)? We believe that a taxonomy of individuals who display specific aptitude profiles in relation to specific achievement situations and outcomes would provide a nice complement to the kind of taxonomy of cognitive, conative, and affective processes that Snow, Corno, and Jackson (1996) began to outline (e.g., Roese & Eccles, 2000).

Another obvious set of studies that would help us to better understand the evolving transaction between persons and situations would track, quantitatively, changes in aptitude complexes and performance over time, something we were unable to do in the present set of studies. There is a need to track the person–situation transition over time microscopically. This means that we need to develop methods for tracking, ideographically, the relative contribution of cognitive, conative, and affective factors that an individual brings into play while performing a task at different stages of the problem-solving process. It also means developing procedures for tracking a person’s “problem–space” representation of the
task as it evolves over time. Methods such as "think-alouds" (Ericsson & Simon, 1993)—concurrent verbalization during task performance—provide one possibility for micro-level analysis (e.g., Ayala, Yin, Shavelson, & Vanides, 2002; Haydel, 2002; Li & Shavelson, 2001; Yin, Ayala, & Shavelson, 2002), but other methods such as discourse analysis might play an important role as well. Finally, we need to develop ways of connecting these microlevel analyses to overall performance across tasks (e.g., Li & Shavelson, 2001).

Snow's idea of multidimensional science achievement needs to be explored further as well; we only scratched the surface in this regard in these studies. We need a "theory" of science achievement, as Cronbach recognized 40 years ago, to guide the construction of tests and validation of test score interpretations (Li & Shavelson, 2001). We need to find cost-efficient ways of developing and implementing constructed response tests and performance assessments, as well as a whole array of cognitively sensitive probes of students' understanding (Pellegrino, Chudowsky, & Glaser, 2001; White & Gunstone, 1992) to supplement the heavy diet of multiple-choice and short-answer questions so current in today's testing climate. We have evidence in this special issue that test format, among other things, does matter for student perceptions of, engagement with, and performance on such tests. However, we need also to broaden the concept of achievement to involve distal yet critical indicators of valued education outcomes such as pursuit of further education and career choice.

Without doubt, Snow's ideas have challenged this group of researchers. We hope the articles in this issue have done some justice to the important and interesting conceptual ideas that Snow left behind. It is said in the Tao Te Ching that the journey of a thousand miles starts from beneath your feet. For those of us involved in this project, these articles represent a first step toward exploring Snow's legacy. It is our hope that these articles motivate first steps among a new generation of scholars interested in the complex organization of aptitudes and situational characteristics that, together, may form the seedbed from which individual differences in academic performance emerge.

ACKNOWLEDGMENTS

Carlos Ayala is now at California State University, Sonoma. Angela Haydel is now at SRI, International. Larry Gallagher is now at American Institutes for Research. Gisell Quihuis is now at EQM Children and Family Services.

This research was funded in part by the National Center for Research on Evaluation, Standards, and Student Testing, under the Educational Research and Development Centers Program, PR/Award R305B90002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.
The findings and opinions expressed in this article do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the United States Department of Education.

REFERENCES


