Multidimensional Validity Revisited: A Multidimensional Approach to Achievement Validation

CSE Technical Report 574

Richard J. Shavelson, CRESST/Stanford University
Shun Lau, 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).

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 573) 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.

MULTIDIMENSIONAL VALIDITY REVISITED

Richard J. Shavelson, CRESST/Stanford University
Shun Lau, Stanford University

Abstract

Richard E. Snow, working from his new aptitude theory, advocated a multidimensional approach to validating the construct of academic achievement. This report briefly describes Snow's recommended approach and then summarizes new evidence from the present studies in terms of three related themes: (a) multidimensionality of science achievement, (b) transaction between achievement and situations, and (c) multidimensional approach to construct validity. Overall, our studies established the predictive validity of several key motivational constructs for science achievement, demonstrated how the relations between these constructs and achievement varied as a function of reasoning dimensions and assessment, and suggested how alternative assessment methods (constructed response and performance assessments) shed light on the notion of multidimensional 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 for science achievement in the National Education Longitudinal Study of 1988 (NELS:88). They established three underlying dimensions, which they called basic knowledge and reasoning, quantitative science reasoning, and spatial-mechanical reasoning. Just as importantly, 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 the spatial-mechanical dimension subscore, but not on the basic knowledge and reasoning or quantitative science reasoning subscores. Curious, Snow 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 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" (see Baker, Linn, & Herman, 1995, pp. 133-134).

We did not accomplish all that Snow imagined. Rather, this study 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 selected or constructed to tap one of the three achievement dimensions: basic knowledge and reasoning (BKR), quantitative science reasoning (QS), and spatial-mechanical reasoning (SM). We included measures of cognitive ability—verbal, mathematical, spatial, and fluid—as well as measures of affect and conation.

In this report, we briefly summarize what we found out about the multidimensional validity of science achievement and the cognitive and affective correlates of achievement. We do so as a set of themes.

Three Themes

Multidimensionality of Achievement

In his new aptitude theory, Snow (1992) expanded the concept of aptitude to include motivational processes in explaining individual differences in achievement. More specifically, he posited two general pathways to describe the manner in which these cognitive and motivational resources played out (Snow, 1994; see also Stanford Aptitude Seminar, 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 in the service of accomplishing particular tasks in particular situations. The other, parallel hypothesized pathway described by Snow was the commitment pathway—a concept that denoted the process by which motivational resources were activated in the service of energizing and guiding behavior toward particular goals in a given situation. Consistent with Snow's theory, Lau, Roeser, and Kupermintz (2002) found that motivational variables predicted science achievement, even after controlling for students' cognitive ability and demographic characteristics. In particular, results of path analysis indicated that students' self-efficacy and task values had direct links to science achievement, as well as indirect links through the mediation of engagement. Moreover, the incorporation of motivational constructs into the model improved its predictive validity (amount of variance explained) for science achievement. Similarly, Kupermintz and Roeser (2002) obtained significant partial correlations between test performance and a number of motivational variables, after accounting for students' cognitive ability and demographic characteristics.

Transaction Between Achievement and Situations

A central tenet of Snow's aptitude theory was that achievement is the result of person-situation interaction. He posited (1994; see also Stanford Aptitude Seminar, 2002) 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 Kupermintz and Roeser's (2002) correlational analysis provided empirical support for the notion of person-situation interaction. First, when compared to multiple-choice total scores, constructed response total scores showed lower correlations with domain-specific motivational constructs, including self-efficacy, task value, self-regulation, and engagement in science class. Second, when compared to multiple-choice or constructed response scores, science grades showed lower correlations with situation-specific (or test-specific) motivational variables, including cognitive strategies, effort expended, mood, and energy level during test-taking situations. Overall, the findings suggest that different assessment methods for science achievement, which represent different situational demands, have differential patterns of associations with motivational processes.

In their path analytic study, Lau et al. (2002) obtained evidence that patterns of engagement depended on achievement situations. Specifically, scores on a standardized science test were associated with cognitive engagement during the test (but not with classroom engagement), whereas science grade was associated with classroom engagement (but not with cognitive engagement during the test).

Furthermore, Haydel and Roeser (2002) found that students characterized by different configurations of motivational beliefs showed different levels of science achievement. Of particular relevance to Snow's idea was the finding that group differences in achievement depended on the type of assessment. For example, the helpless group obtained the lowest scores on the multiple-choice test, whereas the intrinsic-mastery group obtained the lowest scores on constructed response test.

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. Extending previous work, Ayala, Yin, Schultz, and Shavelson (2002) examined the construct validity of the achievement dimensions (BKR, QS, and SM) in constructed response and performance tasks. The authors found that the dimensional complexity of items increased, going from multiple-choice to constructed response to performance items. Indeed, the complexity and paucity of constructed response items ($n_i = 6$) resulted in very low reliability for BKR, QS and SM subscores and for a total constructed response score. Moreover, although the performance assessments were selected to tap one of the three dimensions, respectively, they proved to be quite complex, drawing on all three dimensions to a greater or lesser extent. In the end, what became apparent is that achievement is, indeed, multidimensional.

The three reasoning dimensions capture part of that multidimensionality that we have seen in this study. But there are other aspects of achievement that only can be captured by incorporating a wider variety of achievement tests based on alternative frameworks. For example, Li and Shavelson (2001; see also de Jong & Ferguson-Hessler, 1996; Shavelson & Ruiz-Primo, 1999) distinguish among declarative ("knowing that"), procedural ("knowing how"), schematic ("knowing why"), and strategic ("knowing when") science knowledge. They have provided both cognitive 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 items, especially multiple-choice and constructed response 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.

References

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.

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.

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.

Li, M., & Shavelson, R. J. (2001, April). *Examining the linkage between science achievement and assessment.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

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.

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.

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

Snow, R. E. (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.