[cla]

COLLEGIATE LEARNING ASSESSMENT

CONCEPTUAL FRAMEWORK

The CLA measures fit within Shavelson and Huang's (2003) framework for conceptualizing, developing, and interpreting direct measures of students' learning. This framework utilizes past research on cognition and human abilities (e.g., Gustafsson & Undheim, 1996; Martinez, 2000; Messick, 1984; Pellegrino, Chudowsky & Glaser, 2001) to characterize alternative ways of measuring college students' knowledge and learning. Policy makers, educators, and instrument developers are now discussing the utility of these measures as possible tools for assessing certain higher education outcomes. There are at least three reasons why the Shavelson and Huang framework is important for this project. First, it clarifies the constructs we are and are not measuring and the higher education goals associated with them. The framework, then, guides instrument construction/selection and interpretation. Second, the framework shows where our constructs fit within a 100-year history of efforts to assess student learning in higher education and what has been measured in the past. Third, some of the visions of student learning being proposed by others for higher education initially appear to be inconsistent and contradictory. The framework shows how these visions, and ours, can be integrated.

As Shavelson and Huang note, cognitive outcomes in higher education range from domain-specific knowledge acquisition to the most general of reasoning and problem-solving abilities, yet we know that learning is highly situated and context bound. Only through extensive engagement, practice and feedback in a domain does this knowledge, interacting with prior knowledge and experience, become increasingly decontextualized so that it transfers to enhance general reasoning, problem solving and decision making in a broad domain and later to multiple domains (e.g., Bransford, Brown, & Cocking, 1999; Messick, 1984). And what is learned and to what level it transfers depends on the aptitudes and abilities that students bring with them from their prior education (in and out of school) and their natural endowments (e.g., Shavelson, Roeser et al., in press). A useful framework for linking outcomes with assessments, then, must capture this recursive complexity. It must allow us to map the proposed tests onto the knowledge and abilities that are so highly valued as cognitive outcomes in higher education.

The Shavelson and Huang (2003) framework ranges from domain-specific knowledge to what Spearman called general ability or simply "g." We prefer to refer this as "G" to avoid antiquated interpretation of g as genetically determined (see Cronbach, 2002; Kyllonen & Shute, 1989; Messick, 1984; Snow & Lohman, 1989).[1] As shown in Table 1, levels I – VI in the Shavelson and Huang framework move from "abstract/process oriented" at the top of the table to "concrete content oriented" abilities at the bottom. This ordering also corresponds to abilities that are based on "inheritance and accumulated experience" to those based on "direct experience." General abilities, such as verbal, quantitative and visual-spatial reasoning (see Carroll, 1993), build on inherited capacities and typically develop over many years in formal and informal education settings. These abilities contribute to fluid intelligence (closely allied with "G" and indirectly related to prior learning from a wide range of experiences) and crystallized intelligence (closely allied with learning experiences). "[F]luid intelligence is functionally manifest in novel situations in which prior experience does not provide sufficient direction, crystallized intelligence is the precipitate of prior experience and represents the massive contribution of culture to the intellect" (Martinez, 2000, p. 19).

[1] There are multiple theories of intelligence with Spearman at one extreme postulating a single undifferentiated general intelligence and at the other Guilford postulating 128 abilities and Gardner postulating different, independent intelligences. Shavelson and Huang do not intend to resolve this dispute (but see Carroll, 1993 or Gustoffson, 1996 for recent treatments). Rather, their intent is heuristic, providing a framework in which to locate debates and achievement tests that have been used in the past to assess student learning.

However, measures of crystallized, fluid, and general intelligence do not adequately reflect the in-college learning opportunities of students. They are included in Table 1 for completeness only.

Table 1: Schematic of the Shavelson/Huang Conceptual Framework

Level	What is measured
I	General intelligence ("G")
II	Fluid and crystallized intelligence
III	Verbal, quantitative, and spatial reasoning
IV	Reasoning, comprehending, problem solving, and decision making across broad domains (humanities, social sciences, sciences)
V	Reasoning, comprehending, problem solving, and decision making within broad domains (humanities, social sciences, sciences)
VI	Declarative, procedural, schematic, and strategic domain-specific knowledge

Shavelson and Huang acknowledge that their hierarchy oversimplifies. Knowledge and abilities are interdependent. Learning depends not only on instruction but also on the knowledge and abilities students bring to college. Indeed, instruction and abilities are likely to interact to produce learning, and the course of this interaction evolves over time so that different abilities are called forth and different learning tasks are needed in this evolution (Snow, 1994; Shavelson, Roeser et al., in press). Thus, Table 1 does not behave in strict hierarchical fashion. It is intended to be heuristic, to provide a conceptual framework for discussing and developing learning measures. Our research focuses on levels III-VI of this framework and especially on the cusp between III and IV.

By domain-specific knowledge, Shavelson and Huang refer to knowledge of specific subjects, such as chemistry or engineering. This is the kind of knowledge we would expect to see assessed in students' learning in an academic major. Domain-specific knowledge corresponds to such valued outcomes of higher education (goals) as are typically labeled "learning high-tech skills" or "specific expertise and knowledge in chosen career." Shavelson and Huang (2003) divided domain-specific knowledge into the following four types: declarative ("knowing that"), procedural ("knowing how"), schematic ("knowing why"), and strategic ("knowing when, where and how"—knowing when certain knowledge applies, where it applies, and how it applies). Tests of domain knowledge are appropriate measures of student learning in a major and should be included in the assessment of student learning. Such tests may be published, such as the GRE's area tests. Yet the GRE tests are no longer widely used in most academic majors for a number of reasons including, among others, their fit with the department's particular definition of the major. We also know that students' knowledge in their academic majors is tested extensively by individual instructors and, in some cases, in a capstone course or by an integrated examination. We believe that capitalizing on the availability of such tests provides an opportunity to assess domain-specific knowledge in context. We plan to explore some simple, straightforward ways of doing this, such as by using a pretest, intermediary, and final exams in core (or capstone) courses in the major to examine gain-score effect sizes with and without adjusting for SAT (or ACT) scores.

Broad abilities are complexes of cognitive processes ("thinking") that underlie verbal, quantitative and spatial reasoning, comprehending, problem solving and decision making in a domain, and more generally across domains. These abilities are developed well into adulthood through learning in and transfer from non-school as well as school experiences, repeated exercise of domain-specific knowledge in conjunction with prior learning and previously established general reasoning abilities. As the tasks become increasingly broad—moving from a knowledge domain to a field such as social science, to broad everyday problems—general abilities exercise greater influence over performance than do knowledge structures and domain-specific abilities. Many of the valued outcomes of higher education are associated with the development of these broad abilities. For example, two important goals identified in the National Center for Public Policy and Higher Education (Immerwahl, 2000) survey were "improved problem solving and thinking ability," and "top-notch writing and speaking." Assessments of learning currently in vogue, as well as some assessments developed in the mid-20th century, tap into these broad abilities. Most have focused primarily at the level of the sciences, social sciences, and humanities. The science area score falls between domain specific knowledge and general reasoning abilities. Other tests are more generic, focusing on critical writing and reasoning. Some examples are the GRE's Analytic writing prompts, the College-BASE, the Academic Profile, CAAP, UAP Field Tests, and the 90-minute tasks used in this study. Indeed, many tests of broad abilities contain both area (e.g., sciences) and general reasoning and writing tests.

ACT and the Educational Testing Service have both offered "general education" measures in reading, writing, and mathematics (such as for "rising junior" exams). While few would argue that college students should be proficient in these areas, there is little evidence that scores on such measures are sensitive to the effects of different types of educational programs. For example, 23 institutions participated in perhaps the most comprehensive longitudinal study of learning at the college level to date (Pascarella, Bohr, Nora, & Terenzini, 1995). This study, which produced one of the richest higher education databases ever assembled and generated over 50 published articles, "found little evidence to suggest that attending an academically selective four-year institution had much impact on growth in critical-thinking skills during the first three years of college" (Pascarella, 2001, p. 22).

A cross-sectional study at 56 institutions found that most of the improvement in skills occurs in the first two years of college (Flowers, Osterlind, Pascarella, & Pierson, 2001). However, both of these studies relied on multiple-choice tests of general education skills. There were no open-ended measures (even in writing) and the tests used did not ask students to apply their abilities to realistic and inherently engaging complex tasks. In short, different types of outcome measures—and ones that appear to be more aligned with what colleges say they are trying to do—may lead to different results.