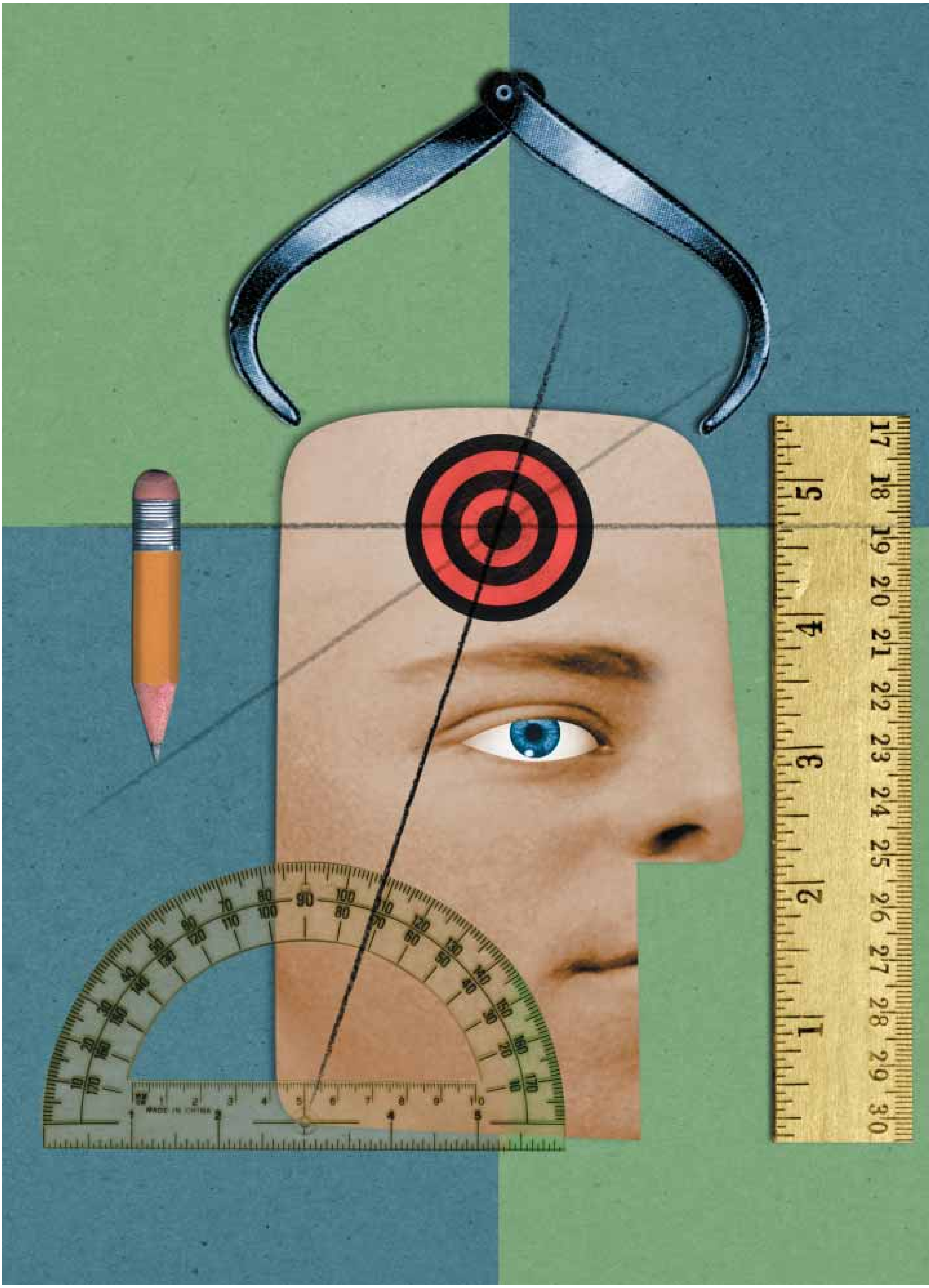

17 18 19 20 21 22 23 24 25 26 27 28 29 30
5 4 3 2 1

Responding Responsibly

To the Frenzy to Assess Learning in Higher Education

By Richard J. Shavelson and Leta Huang

The assessment of learning once again has become a mantra for higher-education reformers. This time, their fires are fueled by the National Center for Public Policy and Higher Education's (NCPPHE) award of 50 "Incomplete" grades for student learning in its state-by-state report card, *Measuring Up* in both 2000 and 2002. *Change*'s recent focus on assessment and accountability in the March/April 2001 issue and Callan and Finney's "educational capital" article in the July/August 2002 issue—together with federal prescriptions calling for more emphasis on learning assessment by accrediting agencies and state mandates for more accountability for student learning—reinforce this message.

These calls for assessment-based accountability are not going away any time soon. How can we respond to them responsibly? What measures will work? And how closely are these measures linked with what students really learn, or to the outcomes we desire?

Today, it seems, almost any measure will do. In the frenzy to find something to test college students with, reformers have suggested instruments ranging from the National Assessment of Educational Progress (NAEP—the so-called "Nation's Report Card" for K-12), to the GRE and other graduate admissions tests, to certification examinations in fields like nursing or teaching, to standardized general-education tests as varied as ACT's Collegiate Assessment of Academic Proficiency (CAAP) and ETS's Tasks in Critical Thinking, and to students' self-reports about the educational processes they experience like the National Survey of Student Engagement (NSSE).

Indeed, in their recent *Change* article Callan and Finney call explicitly for the development of a new NAEP-like test for higher education. But however well-intentioned the chorus of reformers and policymakers may be, going forward blindly with the current testing frenzy may damage higher education more than improve it. From high-stakes testing in K-12 education, for instance, we know full well that what you test is what you get because teachers will teach to the test.

The problem is not that assessment is bad; it's the fact that it is currently being advanced as a policy "solution" in the absence of a coherent conceptual framework that would align assessments with the valued outcomes of higher education. Indeed, the common one-size-fits-all mentality is likely to reduce the diversity of learning environments that characterizes and gives strength to our national higher education system. In the midst of a past K-12 assessment controversy in Texas, Rand Corporation President Jim

---

Richard J. Shavelson is a professor of education and psychology at Stanford University. Leta Huang is a PhD student in education at Stanford University. *The authors are indebted to their colleagues in the Stanford Education Assessment Laboratory—Blake Naughton, Maria Araceli Ruiz-Primo, and Anita Suen—for their critical comments and editing of the manuscript. The authors and not their colleagues, of course, are responsible for the ideas presented here. They are grateful to The Atlantic Philanthropies for support of their research and the preparation of this article, but the ideas are their own.*

Illustration by Joyce Hesselberth/Images.com, Inc.

We believe that by linking knowledge and various kinds of abilities in a single framework, we can perhaps begin to speak to one another coherently about the relative merits of assessing different kinds of learning with different kinds of tests.

Thompson put his finger on the issue:

> One of the prime tools of effective private sector management is an accountability system that includes clear goals, a well-designed incentive structure and solid performance measures. Building this kind of system into American education is a fine idea. *But we have to recognize that the development of accurate education measurements represents an enormous challenge* (Jim Thompson, *Los Angeles Times*, November 2, 2000, emphasis added).

If an organization's outcomes are focused and readily measurable—as in some for-profit businesses—and if the distance between measurements of "outputs" (like revenues or stock prices) and valued "outcomes" (like profit or value-added) is small, performance measures have immediately productive consequences. But if valued outcomes vary substantially within an enterprise—like those of higher education—and if the distance between what is measured and what is desired is great, the opportunity for mischief can be enormous. This is because the tests themselves quickly become proxies for the goals we really value. And where this is the case, unwanted consequences may very well outweigh our good intentions.

In this article, we offer a framework for guiding policy about assessing student learning. We intend this framework to be primarily a heuristic for clarifying the purposes and goals for learning in higher education against proposed measures of those goals. Trying to locate the various arguments for or against a particular assessment in such a manner helps us avoid speaking past one another.

In creating this framework, we draw on almost a century of learning-assessment history in order to provide images of what might be envisioned concretely when we call for "assessing learning," and to remind us of past large-scale assessments (some of which, though almost forgotten today, were close to the magnitude currently proposed). We then conclude with some key points that we believe should be considered in the design of any assessment-of-learning system proposed for higher education today.

Higher Education's Goals

Institutions of higher education vary greatly. They differ with regard to their "inputs" (students, financial resources, faculty), their "processes" or the means by which they transform their inputs (class sizes, student-faculty ratios, teaching methods, and so on), the "outputs" that they measure to indicate their success, and the valued "outcomes" that they seek to ultimately bring about in their students.

These variations give rise to the uniqueness of the American higher education system as well as reflect differences in philosophy about both teaching and the appropriate balance among varied learning goals. This system provides an array of diverse educational environments that have evolved and been adapted to fit the widely varying characteristics of our students with respect to their capabilities, interests, and goals.

We'd like to focus first on the short-term, intermediate, and long-term goals of a college education, and the great variability in emphasis typically placed on these goals—cognitive, personal, social, and civic—across institutions. Among the cognitive outcomes most often cited by institutions are (A) learning skills (for example, how to read historical texts) and knowledge of particular facts in a discipline (for instance, knowing that lesions in Broca's area of the cortex are most likely to lead to expressive aphasia); (B) reasoning with or applying knowledge to solve problems (for example, the ability to understand and contribute to a debate about whether melting glaciers are due to global warming or to secular climate changes); and (C) learning about one's own learning so as to monitor progress in problem-solving and to learn in new situations.

Among the personal and social outcomes we include are empathy, caring, compassion, and self-comprehension. And among civic outcomes are the ability to balance personal and social goals appropriately, take initiative, demonstrate integrity and social responsibility, and work effectively with others and in social institutions.

Such diversity is also reflected in the public's perception of the main purposes of higher education. Judging from results of a national poll reported by NCPPHE (see Immerwahl in Resources), the public supports considerable diversity, ranking higher education's goals in the following order, in terms of the proportion of respondents who believed each to be "absolutely essential":

- Sense of maturity and [ability to] manage on [one's] own (71%)
- Ability to get along with people different from self (68%)
- Problem solving and thinking ability (63%)
- High-technology skills (61%)
- Specific expertise and knowledge in chosen career (60%)
- Top-notch writing and speaking ability (57%)
- Responsibilities of citizenship (44%)

But in today's debates about learning assessment and in state accountability systems, the goals or desired outcomes of higher education are almost always confined to cognitive output measures at the expense of personal, social, and civic outcomes. An example is that the only higher education goal established by the National Education Goals Panel in 1990 looks for increases in "critical thinking, effective communication, and problem solving" without mentioning important social, civic, and personal outcomes. And lacking an appropriate conceptual framework, participants in public policy debates about these matters also tend not to distinguish among a variety of quite different cognitive outcomes. Therefore, they speak past one another.

The stakes in these debates are high. Diverse college missions and richly varied curricula may give way to narrowly defined, one-size-fits-all assessments of highly specified skills at the expense of other important cognitive outcomes as well as valued personal, social, and civic goals. Given the reformers' focus on assessment of knowledge and skills, we want to turn our initial attention in this article to the debate over strictly cognitive output measures—leaving to a later date a consideration of important personal, social, and civic output measures.

We do so with some hesitancy. The most important college outcomes, according to the NCPPHE survey, are associated with others and self: "sense of maturity and manage[s] on own," and "ability to get along with people different from self." To be sure, some state indicator systems attempt to track this by following up with their graduates. We plan to devote another paper to these goals and how they might be measured in the short-, intermediate-, and long-term.

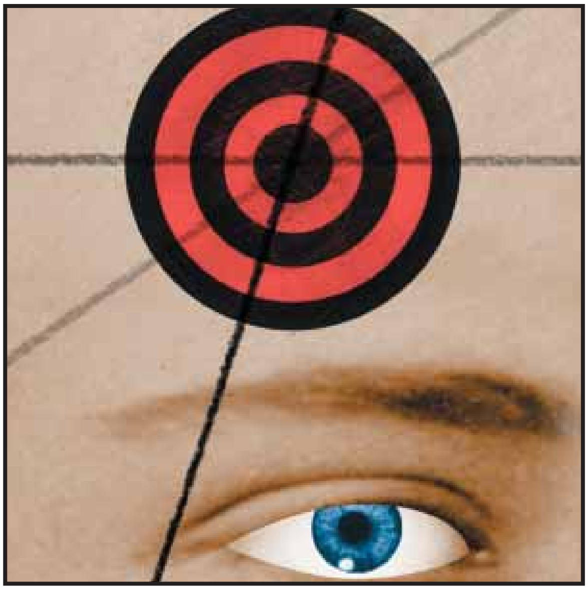
A Framework for Assessing Learning and Achievement

The debate over assessing cognitive outcomes, as we noted earlier, has yet to be informed by a conceptual framework, built on established cognitive and educational research. Such a framework would link statements of outcomes to actual student learning and achievement, which could be linked in turn to specific tests and assessments.

We believe that much of the current debate about assessing student learning in higher education arises because those who join it come to the table with different ideas about what assessment of learning really is and how it ought to be carried out. Some hold lofty goals for assessing such things as critical reasoning and writing abilities, without consideration of the knowledge domain these abilities are shown in—reminiscent of the Renaissance man or woman. Others take a narrower view by contending that "higher learning" involves building up an extensive body of factual and conceptual knowledge—an ability reminiscent of the winners of *Jeopardy*. Increasingly, moreover, state policymakers understand "learning" principally in terms of subsequent occupational success.

Tracing the history of learning assessment in higher education in America, we see different views like this emerging at different times. An emphasis on knowledge content at the turn of the 20th century, for example, gave way to increasing focus on more abstract reasoning abilities at its close.

Recent research on cognition and human abilities provides some ways to reconcile these different views. But making this link is not easy because most research on human abilities has focused principally on testing reasoning processes at the expense of content.



Nevertheless, we believe that by linking knowledge and various kinds of abilities in a single framework, we can perhaps begin to speak to one another coherently about the relative merits of assessing different kinds of learning with different kinds of tests.

Cognitive outcomes in higher education range from what we call "domain-specific" knowledge acquisition to the most general of reasoning and problem-solving skills. Yet we know from research that learning, at least initially, is highly situated and context-bound. Only through extensive engagement, practice, and feedback within a particular subject area does learned knowledge become sufficiently decontextualized to enable it to transfer to the realm of enhanced reasoning, problem-solving, and decision-making skills exercised in broader or multiple domains.

Exactly what is learned, moreover, and to what degree it transfers through generalization, depends a lot on the particular abilities that students bring with them from their prior education (in and out of school) and their own natural endowments. A useful framework for linking outcomes with assessments must capture this recursive complexity. It must allow us to map any proposed assessments onto the particular kinds of knowledge and abilities that are valued highly by multiple stakeholders as cognitive outcomes of higher education.

One possible framework for linking knowledge and reasoning outcomes together is presented in Chart 1. The framework ranges from domain-specific knowledge (shown at the bottom of the chart) to what Charles Spearman called "general ability" or simply "g" (we prefer to refer to it as "G" to avoid the antiquated interpretation of this ability as genetically determined). Toward the top of Chart 1 are theories of intelligence—with Spearman at one extreme postulating a single undifferentiated general intelligence and Guilford at the other postulating over 100 abilities or Gardner postulating different, independent intelligences.

Working from knowledge toward "G," we find increasingly more general abilities—like verbal, quantitative, and visual-spatial reasoning—that build on innate capacities and are typically developed over many years through formal and informal educational settings. These general reasoning abilities, in turn, contribute to "fluid intelligence" (which is closely allied with "G" and indirectly related to prior learning drawn from a wide range of experiences) and "crystallized intelligence" (which is more closely allied with specific learning experiences). As Martinez puts it, "[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" (see Resources).

Using Chart 1 as a guide, we'd like to briefly describe each level of the hierarchy, beginning at its base. And as we do so, we'd like to provide examples of

We suspect that instruction and abilities interact in complex ways to yield what we call learning, and the course of this interaction itself evolves over time in a fashion that calls forth different abilities at different times.

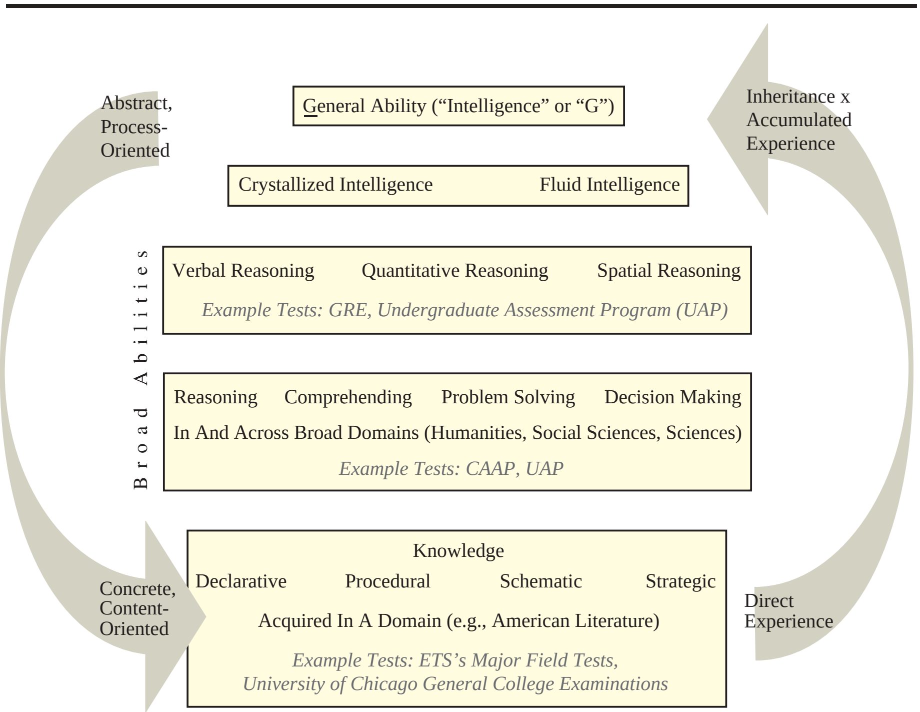
published assessments of collegiate learning at each level. Of course, what we present is clearly an oversimplification. For one thing, knowledge and abilities are interdependent. Learning depends not only on college instruction but also on the knowledge and abilities that students bring to it. Indeed, we suspect that instruction and abilities interact in complex ways to yield what we call learning, and the course of this interaction itself evolves over time in a fashion that calls forth different abilities at different times.

So different kinds of learning tasks are needed as this evolution proceeds. Consequently, what we have sketched in Chart 1 does not behave in strict hierarchical fashion in real life. Indeed, we could have flipped the figure 90 degrees. Our intent is heuristic, to provide a simple but useful conceptual framework to ground future discussions of particular assessment measures.

Domain-Specific Knowledge. By domain-specific knowledge we mean things like a knowledge of physics or of music. This is the kind of knowledge that we would especially expect to see assessed in an academic major. But domain-specific knowledge also corresponds to valued outcomes of higher education that are not limited to a particular academic discipline. Examples include "learning high-technology skills" or "specific expertise and knowledge in chosen career"—skills that were prominently included in the list of publicly valued outcomes of higher education outlined earlier.

We divide domain-specific knowledge into four types—declarative ("knowing that"), procedural ("knowing how"), schematic ("knowing why"), and strategic ("knowing when certain knowledge applies, where it applies, and how it applies"). Conceptual and empirical support for these distinctions comes from many places. For example, brain-imaging studies have found that different types of knowledge—especially declarative knowledge and procedural knowledge—are localized in different parts of the brain (see Bransford, Brown, and Cocking in Resources).

Chart 1. Framework for Cognitive Outcomes



Research in cognitive science, meanwhile, not only provides evidence of differences among declarative, procedural, and strategic knowledge, but also identifies what we have called "schematic knowledge" (often labeled a "mental model"). Similar distinctions among these various types of knowledge have been made by those establishing content standards for K-12 subject areas. And the test-development frameworks for such large-scale assessments as the Third International Mathematics and Science Study (TIMSS) and NAEP not only distinguish declarative from procedural knowledge, but may also include "application" skills that probably draw on schematic knowledge.

At the beginning of the 20th century, a good deal of attention was paid to testing declarative knowledge in colleges and universities as well as procedural knowledge of a step-by-step or algorithmic nature. The Pennsylvania Study (see Learned and Wood; see also Kandel and Pace in Resources) is a case in point. In 1928, about 70 percent of all then-enrolled Pennsylvania college seniors were tested for 12 hours on 3,482 selected-response (multiple-choice, matching, true-false) questions across all areas of the curriculum (see Table 1).

These tests were built on the assumption "that to deserve recognition, knowledge must be a relatively permanent and available equipment of the student; that it must be so familiar and so sharply defined that it comes freely to mind when needed and can be depended on as an effective cross-fertilizing element for producing fresh ideas" and that "a student's knowledge, when used as adequate evidence of education…should represent as nearly as possible the complete individual" (see Learned and Wood in Resources).

The success of the Pennsylvania Study led to the College Sophomore Testing Program, in which testing was first conducted in May of 1932 at 140 institutions across 38 states, with 18,134 college students participating. The test battery from which an institution could choose consisted of an intelligence test and no fewer than 12 Cooperative Tests including English, Literary Acquaintance, General Culture, General Science, General Mathematics, Foreign Language, Contemporary Affairs, Physics, Chemistry, Zoology, Botany, and Geology.

As tests of domain-specific knowledge evolved, they began to focus on types of knowledge other than simply declarative and procedural algorithms. Perhaps the most ambitious and well-known attempt to broaden knowledge-domain testing was that of Benjamin Bloom, the University of Chicago's Examiner. Bloom and colleagues (see Bloom, *et al.*, in Resources) built a now well-known taxonomy of cognitive objectives that served as a basis for creating specific tests that went beyond factual and conceptual knowledge to include application, synthesis, analysis, and evaluation.

Student achievement in different content domains was measured by the test set developed and used by the University of Chicago Examiner's Office. What is important (and remains unique) about these examinations is that some provided students completing a course of study say, in physics, with original-source reading material a few weeks before the examination. This reading material formed the content basis for the examination. The test itself was then focused on the student's understanding of the provided material—something that most students could not demonstrate without having actually taken the course of study.

In some sense, our characterization of knowledge types as declarative, procedural, schematic, and strategic harkens back to Bloom's taxonomy. This is certainly true of its intent: to characterize the kinds of domain-specific achievement tests that might be developed, and to expand content-testing to include things like test questions whose answers require students to draw upon schematic knowledge and strategic knowledge to address novel problems or tasks.

But most tests of domain-specific knowledge in higher education still focus on declarative knowledge, as exemplified by ETS' Major Field Tests (MFT). For example, in the Psychology test, students are asked: "As children become older, they are increasingly able to take others' needs and points of view into consideration. Piaget has called this trend in children's thinking: (A) conservation/(B) object constancy/(C) decentration/(D) informal operations/(E) specific and nonspecific transfer" (*Note: Question reprinted with permission of Educational Testing Service, the copyright owner.* Online at www.ets.org/hea/mft/discipline.html).

Table 1. Sample Test Items from the Pennsylvania Study Examination (1930-32)

IV. GENERAL SCIENCE, *Part II*

Directions: In the parenthesis after each word or phrase in the right-hand column, place the number of the word or phrase in the left-hand column of the same group which is associated with that word or phrase.

14.	1. Unit of work	Calorie
	2. Unit of potential difference	Dyne
	3. Unit of electrical current	Erg
	4. Unit of heat quantity	H. P.
	5. Unit of power	Volt
	6. Unit of force	Ampere
	7. Unit of pressure	B. T. U.
		Atmosphere
		Foot-pound
		Watt

V. FOREIGN LITERATURE, *Multiple Choice*

9. Sophocles' *Antigone* is a depiction of 1 the introduction of laws into a barbarous state, 2 the prevailing of sisterly love over citizenly duty, 3 idyllic peasant life, 4 the perils of opposing oneself to Zeus

10. Of Corneille's plays, 1 *Polyeucte*, 2 *Horace*, 3 *Cinna*, 4 *Le Cid*, shows least the influence of classical restraint

VI. FINE ARTS, *True-False*

1. Greek architecture prior to contact with the Romans made no use of the dome

10. The slow movements of Beethoven's symphonies are somewhat inferior to the rest of those compositions

VIII. MATHEMATICS

Directions: Each of the problems below is followed by several possible answers, only one of which is entirely correct. Calculate the answer for each problem; then select the printed answer which corresponds to yours and put its number in the parenthesis at the right.

5. If two sides of a triangle are equal, the opposite angles are
(1) equal (2) complementary (3) unequal (4) right angles

Source: Learned and Wood, 1938.

Broad Abilities. Broad abilities consist of particular complexes of cognitive processes ("thinking") that underlie what we generally call verbal, quantitative, and spatial reasoning—as well as comprehension, problem-solving, and decision-making skills within domains (and more generally, across domains). Broad abilities are developed well into adulthood in non-school as well as school-based settings through the repeated practice of domain-specific

We believe that what is needed are clear distinctions between achievement in a domain-specific area of study and demonstrations of more general abilities, and that tests of each ought to be included in the assessment of learning.

knowledge in conjunction with prior learning and previously established general-reasoning abilities. Consequently, these developed abilities are neither solely innate nor fixed in their capacity (see Messick in Resources).

Broad abilities play out in educational achievement together with different types of knowledge. As Messick puts it, "In educational achievement, cognitive abilities and ability structures are engaged with knowledge structures in the performance of subject-area tasks. Abilities and knowledge combine in ways guided by and consistent with knowledge structures to form patterned complexes for application and action."

As tasks become increasingly broad—moving from a specific knowledge domain like sociology, to an area of study such as social science, to everyday problems in social situations—general abilities exercise increasingly greater influence over performance than do disciplinary knowledge structures and domain-specific abilities. We believe that many of the most valued outcomes of higher education are associated with the development of such broad abilities. For example, two important goals identified in the NCPPHE survey were "improved problem solving and thinking ability" and "top-notch writing and speaking ability."

Table 2. Sample Area Test Questions

ACT Collegiate Assessment of Academic Proficiency (CAAP)—
Science Reasoning (2002)

Given a passage contrasting Aristotle's and Galileo's explanations of the physics of falling objects, answer the following question (among others):

2. A book dropped from a height of 1 meter falls to the floor in t seconds. To be consistent with Aristotle's views, from what height, in meters, should a book 3 times as heavy be dropped so that it will fall to the floor in the same amount of time?
   A. 1/9
   B. 1/3
   C. 1
   D. 3

Note: Reproduced with permission from ACT, Inc., the copyright owner.

ETS Academic Profile—*Social Sciences* (1992)

After reading a passage on economic growth and territorial expansion, answer the following question (among others):

4. In using the phrase, "community of disease" in line 14 above, the author of the passage most likely intends to indicate that

   (A) members of a particular society generally develop resistance to diseases that occur frequently in their society
   (B) only members of the same society are likely to be susceptible to certain diseases
   (C) the exposure of diverse peoples of the world to the same diseases constitutes a link between these peoples
   (D) the devastating effect of disease is a unifying factor among the people who suffer from it

Note: Reprinted by permission of Educational Testing Service, the copyright owner.

Most of the assessments of collegiate learning currently in vogue—as well as some developed in the mid-20th century—are designed to measure such broad general abilities. The domains of most—for example, ACT's Collegiate Assessment of Academic Proficiency (CAAP), ETS' Tasks in Critical Thinking, ETS' Academic Profile, and ETS' now-defunct Undergraduate Assessment Program's (UAP) Area Tests—have been constructed broadly at the level of knowledge areas like the sciences, social sciences, and humanities (see Table 2).

Nevertheless, many of them (for example, the CAAP and the UAP) address an area like science through discrete sets of questions on physics, biology, and chemistry. Because too few items are available in each discipline to produce reliable discipline-level knowledge scores, a broader aggregate science area score is provided even though the questions themselves address disciplinary knowledge. The "science area score" on these tests thus falls somewhere between domain-specific knowledge and general reasoning abilities.

Other tests are more generic—focusing, for example, on critical reasoning and writing abilities (see Table 3). Examples include the GRE's new Analytical Writing Section, the College-BASE, the Academic Profile, CAAP, and the critical thinking test used in the Cooperative Study of Evaluation in General Education. Indeed, many tests of broad abilities contain both area-level, and general reasoning and writing tests.

Intelligence: Crystallized, Fluid, and General. In the upper parts of Chart 1 we find a number of general reasoning abilities that develop over significant periods of time through experience (including school) in interaction with innate capacities. These are the most general of abilities, and account for consistent levels of performance across quite different contextual situations. R.B. Cattell, for example, argued that intelligence involves both fluid and crystallized abilities: "Both these dimensions reflect the capacity for abstraction, concept formation, and perception and eduction [*sic*] of relations" (see Gustafsson and Undheim in Resources). The fluid dimension of intelligence "…is thought to reflect effects of biological and neurological factors" and includes speed of processing, visualization, induction, sequential reasoning, and quantitative reasoning. It is most strongly associated with performance on unique or novel tasks. The crystallized dimension is thought to reflect acculturation (obtained especially through education) and involves language and reading skills like verbal comprehension and language development, and school-related numeracy, as well as school-based achievement.

These reasoning abilities are somewhat removed from the immediate effects of college instruction. They are developed over a long period of time, both in school and out. Nevertheless, there is some evidence of short-term college impact on these abilities (see Pascarella and Terenzini in Resources). In current assessments, these abilities typically are interpreted in terms of crystallized and fluid reasoning abilities (as in the GRE's treatment of Verbal and Quantitative abilities, respectively).

The Undergraduate Assessment Program, in turn, provided measures of both verbal ability ("…knowledge of words and ability to comprehend reading materials from a variety of sources") and quantitative ability ("…understanding of, and ability to reason with, mathematical symbols in the solution of problems"), based originally on the GRE. There is also considerable overlap between the questions on, for example, the Academic Profile Area Tests and the GRE Verbal and Quantitative tests.

The most general of all abilities is "G"—the stuff that fuels thinking, reasoning, decision making, and problem solving, and accounts for the consistency of performance across vastly different (and not-so-different) situations. General intelligence involves induction "and other factors involving complex reasoning tasks" (see Gustafsson and Undheim in Resources). While education might *ultimately* be aimed at cultivating general intelligence, changes in this commodity due to college instruction would be expected typically to be quite small, and far removed from the

Table 3. Sample Questions from Critical Reasoning and Writing Tests

GRE Analytical Writing Issue Topic (2002)

Present your perspective on the issue below, using relevant reasons and/or examples to support your views.

"The best ideas arise from a passionate interest in commonplace things."

Note: Reprinted by permission of Educational Testing Service, the copyright owner.

ACT CAAP Writing Essay (2002)

Your college administration is considering whether or not there should be a physical education requirement for undergraduates. The administration has asked students for their views on the issue and has announced that its final decision will be based on how such a requirement would affect the overall educational mission of the college. Write a letter to the administration arguing whether or not there should be a physical education requirement for undergraduates at your college. (www.act.org/caap/sample/essay.html)

Note: Reproduced with permission from ACT, Inc., the copyright owner.

ACE's Cooperative Study of Evaluation: Test of Critical Thinking (1952)

In items 32 through 34 each item gives part of an argument, followed by five sentences. One of the five sentences completes the argument in such a way as to justify the conclusion. Select this one sentence in each case.

34. Russelson is not an American. It follows, therefore, that he is not truly democratic.

1. Americans are truly democratic.
2. All Americans are truly democratic.
3. Only Americans are truly democratic.
4. Some Americans are truly democratic.
5. Some non-Americans are not truly democratic.

Note: Reprinted by permission of the American Council on Education.

substance of higher education under the control of colleges and universities.

Some Ironies of Typical Practice. One irony made evident by this discussion is that learning and knowledge are highly domain-specific—as, indeed, is most reasoning. Consequently, the direct impact of college is most likely to be seen at the *lower* levels of Chart 1—domain-specific knowledge and reasoning. Yet, in the formulation of most college goal statements for learning—and consequently in choices about the kinds of tests to be used on a large scale to hold higher education accountable—the focus is usually in large part on the *upper* regions of Chart 1.

Indeed, some states identify performance on the GRE and other graduate admissions tests as measures to be used to index learning in college. This simply misplaces the type of assessment results for which colleges should be held accountable. Certainly we should reasonably expect some collegiate effects on broader abilities in the long run. But as one moves up the hierarchy, multiple life experiences—before, inside and outside of college—come into play.

At the most general level, the effects of college may be there, but they are probably at their weakest because these abilities are founded on a lifetime of experience. In contrast to current practice, we believe that what is needed are clear distinctions between achievement in a domain-specific area of study and demonstrations of more general abilities, and that tests of each ought to be included in the assessment of learning.

A second irony, as others have pointed out, is that higher education tends to be organized so that general education requirements and survey courses occur at the beginning of college, while specialized in-depth courses in the disciplinary major happen later. Yet from all we know, college is more likely to have direct effects on the latter, while the intellectual demands of breadth and integration of knowledge expected in statements of general education outcomes draw on reasoning processes developed throughout a stu-

Resources

■ Bloom, B. S., *et al.*, *Taxonomy of Educational Objectives: The Classification of Educational Goals*, by a Committee of College and University Examiners, New York: Longmans, Green, 1956.

■ Bransford, J. D., A. L. Brown, and L. L. Cocking, *How People Learn: Brain, Mind, Experience, and School*, Washington, DC: National Academy Press, 1999.

■ Carroll, J. B., *Human Cognitive Abilities. A Survey Of Factor-Analytic Studies*, Cambridge, England: Cambridge University Press, 1993.

■ Cattell, R. B., "Theory of Fluid and Crystallized Intelligence: A Critical Experiment," *Journal of Educational Psychology*, 1963, Vol. 54, pp. 1-11.

■ Cronbach, L. J., ed., *Remaking the Concept of Aptitude: Extending The Legacy of Richard E. Snow*, Mahway, NJ: Erlbaum, 2000.

■ Gustafsson, J. E., and J. O. Undheim, "Individual Differences in Cognitive Functions," in R. Calfee and D. Berliner, eds., Handbook of Educational Psychology, New York: Macmillan, 1996, pp. 186-242.

■ Heubert, J. P., and R. M. Hauser, eds., *High Stakes Testing for Tracking, Promotion and Graduation*, Washington, DC: National Academy Press, 1999.

■ Immerwahl, J., *Great Expectations: How Californians View Higher Education*, San Jose, CA: National Center for Public Policy and Higher Education and Public Agenda, August 2000 (Table 3, National column).

■ Kandel, I. L., "Examinations and Their Substitutes in the United States," *Bulletin Number Twenty-Eight*, New York: The Carnegie Foundation for the Advancement of Teaching, 1936.

■ Learned, W. S. and B. D. Wood, *The Student and His Knowledge: A Report to the Carnegie Foundation on the Results of the High School and College Examination of 1928, 1930, and 1932*, Boston: The Merrymount Press, 1938, No. 29.

■ Martinez, M. E., *Education as the Cultivation of Intelligence*, Mahway, NJ: Erlbaum, 2000.

■ Messick, S., "The Psychology of Educational Measurement," *Journal of Educational Measurement*, Vol. 21, No. 3, 1984, pp. 215-237.

■ Pace, C. R., *Measuring Outcomes of College: Fifty Years of Findings and Recommendations for the Future*, San Francisco: Jossey-Bass, 1979.

■ Pascarella, E. T., and P. T. Terenzini, *How College Affects Students: Findings and Insights from Twenty Years of Research*, San Francisco: Jossey-Bass, 1991.

■ Pellegrino, J. W., N. Chudowsky, and R. Glaser, eds., *Knowing What Students Know: The Science and Design of Educational Assessment*, Washington DC: National Academy Press, 2001.

■ Present and Former Members of the Faculty, *The Idea and Practice of General Education: An Account of the College of The University of Chicago*, Chicago: The University of Chicago Press, 1950.

■ Snow, R. E., "Abilities in Academic Tasks," in R. J. Sternberg, and R. K. Wagner, eds., *Mind in Context: Interactionist Perspectives on Human Intelligence*, Cambridge, England: Cambridge University Press, 1994, p. 337.

■ Snow R. E., and D. F. Lohman, "Implications of Cognitive Psychology for Educational Measurement," in R. Linn, ed., *Educational Measurement third ed.*, New York: Macmillan, 1989, pp. 263-331.

dent's coursework (if not a lifetime). But we do wonder what would happen if the normal course of study were stood on its head.

Some Propositions About Assessing Learning Responsibly

The current frenzy for assessing learning is well intentioned, but it needs a coherent sense of direction. In conclusion, we take a small step toward addressing that need. We do so in the form of a set of six propositions.

1) *Assess personal, social, and civic abilities as well as cognitive ones.* Policymakers, educators, and the public hold multiple goals for higher education. As important as cognitive goals on the list, stakeholders also value personal, social, and civic goals. Yet the sole focus of current learning assessment policy is on cognitive goals (a trap that, admittedly, this article falls into as well). This omission might be excusable for mandatory K-12 education. We think it is inappropriate for higher education. The conceptualization of learning needs to be broadened to include these abilities—in both the assessment arena and in broader discussions of public policy for higher education.

2) *Encourage real dialogue and greater agreement on the content of assessments.* Current proposals for measuring cognitive aspects of student learning lack conceptual coherence. If we listen closely to the cacophony of voices urging the assessment of collegiate learning, we will notice that they frequently speak past one another. Some focus on learning as declarative knowledge and call for NAEP-like assessments in higher education—ignoring the different types of knowledge that students should develop in an academic discipline.

Others propose the use of graduate admissions examinations like the GRE, which purport to measure broad general abilities like verbal and quantitative reasoning or analytic writing across disparate disciplines—even though cognitive research emphasizes the situated nature of such abilities and the bulk of collegiate education rarely focuses explicitly on developing broad abilities. Still others emphasize testing "higher-order" thinking in or across mid-range areas like the humanities, social sciences, and sciences, through the use of available instruments like the CAAP, the Academic Profile, and Tasks in Critical Thinking.

3) *Recognize that what we test and make public will greatly influence what is taught and what is learned.* We see this in the states that have actually adopted a form of assessment-based accountability in higher education (for example, Florida and South Dakota). If what we test for does not closely tap what we value—that is, if the outputs we measure do not closely link to the outcomes we really want out of higher education—we run the risk of distorting what colleges do. Many current proposals for assessing learning run exactly this risk. In launching assessments of learning in higher education, we need to be careful. We may in fact get what we measure. And we might not like what we get.

4) *Achieve clarity in the debate about what to assess through use of a conceptual framework.* The particular conceptual framework for assessing learning we present is but one of a number of possibilities that may prove helpful in framing policy conversations about assessment. Our scheme tries to portray where different voices fall within a wide array of current policy positions and assessments. It also enables us to ask questions about how closely tied particular kinds of assessments are likely to be to the learning experiences of students in our colleges and universities. The further away from the disciplinary base of Chart 1 you go, the more likely it is that factors other than what is taught in college will come into play.

On the other hand, if these intellectual abilities indeed become the long-range goals of higher education, what and how we teach needs to be altered significantly. Regardless of how this debate comes out, let's at least be conceptually clear about what we are assessing so that the learning we get through collegiate teaching and learning is what we want.

5) *Develop multiple and varied assessments.* At the college level, one size will not fit all institutions—even if we focus solely on the cognitive domain. If they are to improve students' learning, assessments most likely will need to be multi-level. They will need to tap into discipline-based knowledge and reasoning in the academic major but also look at closely related and taught abilities at the next-higher level of Chart 1—perhaps in the context of general education. Moreover, once we recognize the need to assess personal, social, and civic learning, this caveat will apply even more. We need to develop multiple and varied—as well as sound—assessments of these valued abilities.

6) *Distribute meaningful feedback on assessment results to all stakeholders.* Learning assessments are powerful instruments for educational change, but they have great potential for mischief. Alone, they will not improve learning. But they allow policymakers and pundits to point blame from far-removed bully pulpits. For these assessments to have their intended consequences—the improvement of learning and public accountability for making progress toward improved learning—we must figure out how to provide useful feedback based on their results to all of the stakeholders in higher education, from policymakers through academic leaders, to teaching faculty.

* * *

The issue is one of establishing an *improvement* process, not just a measurement process. The challenge, then, lies ultimately not just in assessing learning meaningfully and accurately, but in what we do with the results. Accountability, under these circumstances, too often involves just pointing fingers. *Real* accountability ought to embrace assuming meaningful responsibility for improving the knowledge and abilities that we value as educational organizations and as a society. C