Richard J. Shavelson, Stanford University

THE COLLEGIATE LEARNING ASSESSMENT

In its 2006 report, the Spellings Commission exhorted higher education institutions to measure their effectiveness by gathering quality data to assess student learning. The commission specifically identified the Collegiate Learning Assessment (CLA) and the Measure of Academic Proficiency and Progress (MAPP) as viable instruments for gathering such data. Since the release of the report, considerable attention has been given to the CLA by the commission's chair, Charles Miller, and a number of higher education associations; likewise, the number of colleges and universities using the CLA has more than doubled during the past year or so. The CLA is unlike other assessments of undergraduates' learning, which are primarily multiple-choice tests. This paper describes the CLA and discusses its role in the larger context of assessment and accountability.

Collegiate Learning Assessment

The CLA was developed to measure undergraduates' learning—in particular their ability to think critically, reason analytically, solve problems, and communicate clearly. The assessment focuses on the institution or on programs within an institution. Institution or program-level scores are reported both in terms of observed performance and as value added beyond what would be expected from entering students' SAT scores. The CLA also provides students their scores on a confidential basis so that they can gauge their own performance.

The assessment consists of two major components: a set of performance tasks and a set of two different kinds of analytic writing prompts (see Figure 1). The performance tasks component presents students with problems and related

Figure 1. Collegiate Learning Assessment Structure
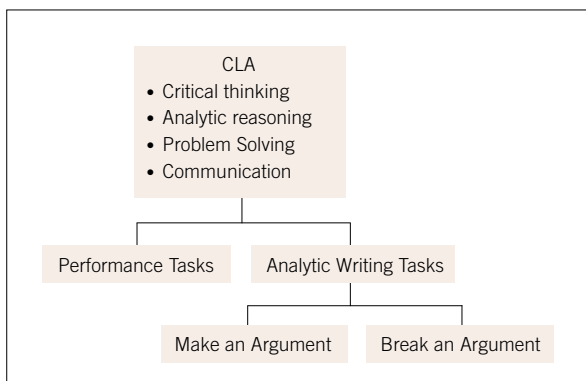
information and asks them either to solve the problems or recommend a course of action based on the evidence provided. The analytic writing prompts ask students either to take a position on a topic or to critique an argument.

The Collegiate Learning Assessment's Criterion-Sampling Approach

The CLA differs substantially—in terms of both its philosophical and theoretical underpinnings—from most learning assessments, such as the Measure of Academic Proficiency and Progress (MAPP) and the Collegiate Assessment of Academic Progress (CAAP). Most learning assessments grow out of an empiricist philosophy and a psychometric/behavioral tradition. From this stance, everyday complex tasks are divided into components, and each component is analyzed to identify the abilities required for successful performance. For example, suppose that components such as critical thinking, problem

solving, analytic reasoning, and written communication are identified. A separate measure of each ability would then be constructed and students would take each test. At the end of testing, students' scores on the tests would be added up to construct a total score to describe their performance—not only on the assessment at hand, but also generalizing to a universe of complex tasks similar to those the tests were intended to measure.

In contrast, the CLA is based on a combination of rationalist and sociohistorical philosophies in the cognitive-constructivist and situated-in-context traditions (e.g., Case, 1996). The CLA's conceptual underpinnings are embodied in what has been called a *criterion sampling* approach to measurement (see Table 1). This approach assumes that the whole is greater than the sum of its parts and that complex tasks require an integration of abilities that cannot be captured when divided into and measured as individual components.

The criterion-sampling notion is straightforward: If you want to know what a person knows and can do, sample tasks from the domain in which that person is to act, observe her performance, and infer competence and learning. For example, if you want to know whether a person not only knows the laws that govern driving a car but also if she can actually drive a car, don't just give her a multiple-choice test. Rather, also administer a driving test with a sample of tasks from the general driving domain such as starting the car, pulling into traffic, turning right and left in traffic, backing up, and parking. Based on this sample of performance, it is possible to draw valid inferences about her driving performance more generally.

The CLA follows the criterion-sampling approach by defining a domain of real-world tasks that are holistic and drawn from life situations. It samples tasks and collects students' *operant responses*. Operant responses are student-generated responses that are modified with feedback as the task is carried out. These responses parallel those

Table 1. CLA's Criterion-Sampling Approach to Measurement

Criterion-Sampling Approach		Collegiate Learning Assessment (CLA)
Samples tasks from "real-world" domains	→	Samples holistic, real-world tasks drawn from life experiences
Samples operant as well as respondent responses	→	Samples constructed responses (not multiple choice)
Elicits complex abstract thinking ("operant thought patterns")	→	Elicits critical thinking, analytic reasoning, problem solving, and communication
Provides information on how to improve on tasks ("cheating" is not possible if can do criterion task!)	→	Provides tasks for teaching as well as assessment

expected in the real world. There are no multiple choice items in the assessment; indeed, life does not present itself as a set of alternatives with only one correct course of action. Finally, the CLA provides CLA-like tasks to college instructors so that they can "teach to the test." With the criterion-sampling approach, "cheating" by teaching to the test is not a bad thing. If a person "cheats" by learning and practicing to solve complex, holistic, real-world problems, he has demonstrated the knowledge and skills we seek as educators to develop in students. That is, he has learned to think critically, reason analytically, solve problems, and communicate clearly. Note the contrast with traditional learning assessments, for which practicing isolated skills and learning strategies to improve performance may lead to higher scores but is unlikely to generalize to a broad, complex domain.

CLA Performance Tasks

Recall that the CLA is composed of performance tasks and analytic writing tasks. "DynaTech" is an example of a performance task (see Figure 2). DynaTech is a company that makes instruments for aircraft. The company's president was about to approve the acquisition of a SwiftAir 235 for the sales force when the aircraft was involved in an accident. As the president's assistant, you (the student) have been asked to evaluate the contention that the SwiftAir is accident prone. Students are provided an "in-basket" of information that might be useful in advising the president. They must weigh the evidence—some relevant, some not; some reliable, some not—and use this evidence to support a recommendation to the president. (Incidentally, it might be that the SwiftAir uses DynaTech's altimeter!) DynaTech exemplifies the type of performance tasks found on the CLA and their complex, real-world nature.

To get a better understanding of what might be contained in a performance task's in-basket, consider the "Crime" performance task (see Figure 3). You are now a consultant to the incumbent mayor, who is up for re-election. At issue is a rising number of crimes in the city and their association with drug trafficking. The mayor has proposed increasing the number of police to address crime. His opponent is a city council member, and she has proposed an alternative to police—increased drug education. Her proposal, she argues, addresses the cause and is based on research studies. You are given an in-basket of information regarding crime rates, drug usage, the relationship between the number of police and robberies, research studies, and newspaper articles—some relevant, some not; some reliable, some not. Your task is to advise the mayor, based on the evidence, as to whether his opponent is right about both drug education and her interpretation of the positive relationship between the number of police and number of crimes.

CLA Analytic Writing Tasks

The CLA contains two types of analytic writing tasks. The first type of task asks students to build and defend an argument. For example, students might be asked to agree or

Figure 2. CLA's DynaTech Performance Task

You are the assistant to Pat Williams, the president of DynaTech, a company that makes precision electronic instruments and navigational equipment. Sally Evans, a member of DynaTech's sales force, recommended that DynaTech buy a small private plane (a SwiftAir 235) that she and other members of the sales force could use to visit customers. Pat was about to approve the purchase when there was an accident involving a SwiftAir 235.
You are provided with the following documentation:

1. Newspaper articles about the accident
2. Federal Accident Report on in-flight breakups in single engine planes
3. Pat's e-mail to you and Sally's e-mail to Pat
4. Charts on SwiftAir's performance characteristics
5. Amateur Pilot article comparing SwiftAir 235 to similar planes
6. Pictures and description of SwiftAir Models 180 and 235

Please prepare a memo that addresses several questions, including what data support or refute the claim that the type of wing on the SwiftAir 235 leads to more in-flight breakups, what other factors might have contributed to the accident and should be taken into account, and your overall recommendation about whether or not DynaTech should purchase the plane.
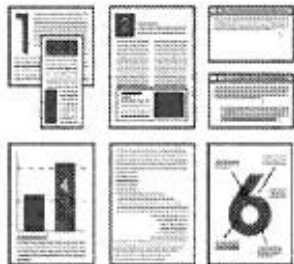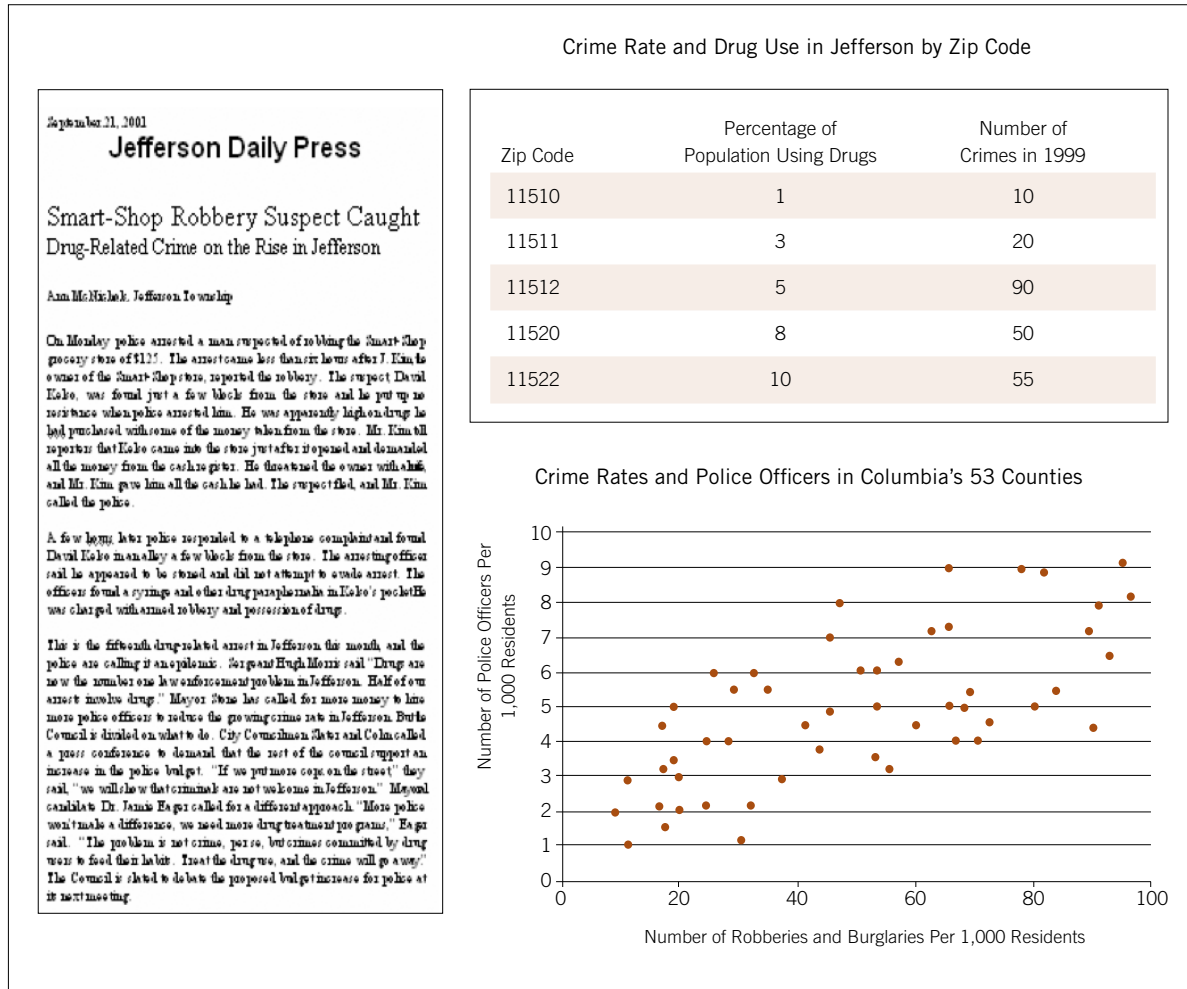
Figure 3. CLA In-Basket Items from the Crime Performance Task

Crime Rate and Drug Use in Jefferson by Zip Code

Zip Code	Percentage of Population Using Drugs	Number of Crimes in 1999
11510	1	10
11511	3	20
11512	5	90
11520	8	50
11522	10	55

Crime Rates and Police Officers in Columbia's 53 Counties

(Scatter plot: y-axis "Number of Police Officers Per 1,000 Residents" ranging 0 to 10; x-axis "Number of Robberies and Burglaries Per 1,000 Residents" ranging 0 to 100.)

---

disagree with the following premise, justify their position with evidence, and show weaknesses in the other side of the argument: "College students waste a lot of time and money taking a required broad range of courses. A college education should instead prepare students for a career."

The second type of task is one in which a student is asked to critique an argument such as the following:

A well-respected professional journal with a readership that includes elementary school principals recently published the results of a two-year study on childhood obesity. (Obese individuals are usually considered to be those who are 20 percent above their recommended weight for height and age.) This study sampled 50 schoolchildren, ages 5–11, from Smith Elementary School. A fast food restaurant opened near the school just before the study began. After two years, students who remained in the sample group were more likely to be overweight relative to the national average. Based on this study, the principal of Jones Elementary School decided to confront her school's obesity problem by opposing any fast food restaurant openings near her school.

In this case, the student must evaluate the claims made in the argument and either agree or disagree, wholly or in part, and provide evidence for the position taken.

CLA Technology

Many of the ideas underlying the CLA are not new. The history of learning assessment (e.g., Shavelson, 2007a, 2007b; Shavelson & Huang, 2003) as far back as the late 1930s shows that assessments similar to the CLA have been being built for decades. In the late 1970s, John Warren at the Educational Testing Service (ETS) was experimenting with constructed-response tasks, American College Testing (ACT) created the College Outcomes

Measurement Project (COMP), and the state of New Jersey created Tasks in Critical Thinking to assess undergraduates' learning. These assessments had marvelous performance tasks but in all cases the attempts to build these assessments failed. They were costly, logistically challenging, and time consuming to score.

The CLA solves past problems of time, cost, and scoring by capitalizing on Internet, computer, and statistical-sampling technologies. The advent of these technologies has made it possible to follow in the tradition of the criterion-sampling approach. Students' complex performance still is scored by human judges, but their performance on the analytic writing prompts can be scored by natural language processing software without compromising reliability or validity (e.g., Klein et al., 2005, 2007). Moreover, the CLA uses matrix sampling so that not all students answer all questions, which reduces testing time. (Nevertheless, even with this technology, it takes a fair amount of time—90 minutes—to answer subsets of questions.) Finally, reports can be produced rather quickly because of the technology used.

Table 2 presents a summary description of CLA tasks, technology used, and reporting of results.
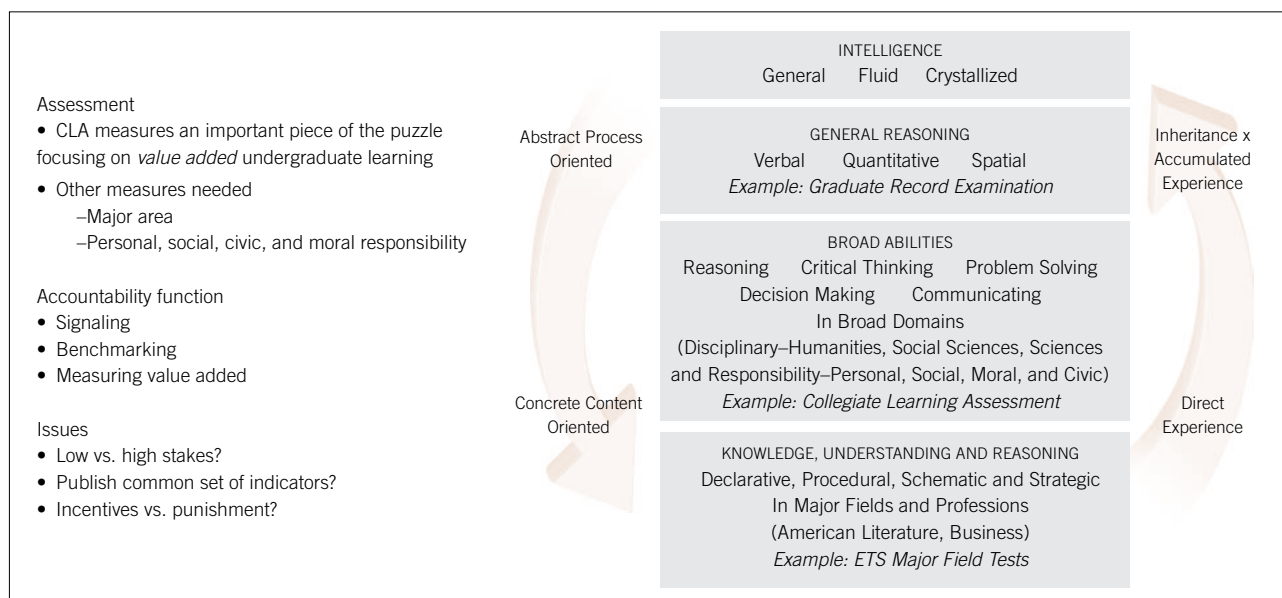
Assessment and Accountability

The CLA, with its focus on broad cognitive abilities of analytic reasoning, critical thinking, problem solving, and communication is but one piece of the assessment and accountability puzzle (see Figure 4). Other outcomes

Table 2. CLA Technology and Reporting

Characteristic	Attributes
Open-ended tasks	Tap critical thinking, analytic reasoning, problem solving, and written communication.
	Provide realistic work samples.
	Engage students, as suggested by alluring titles such as "Brain Boost," "Catfish," and "Lakes to Rivers."
Computer technology	Interactive Internet platform.
	Paperless administration.
	Natural language processing software for scoring students' written communication.
	Online rater scoring and calibration of performance tasks.
	Reports institution's (and subdivision's) performance (and individual student performance confidentially to student).
Focus	Institution itself or school, department, or program within it.
Sampling	Samples students so that not all students perform all tasks.
	Samples tasks for random subsets of students.
	Creates scores at institution or subdivision level as desired (depending on sample sizes).
Reporting	Controls for students' ability so that "similarly situated" benchmark campuses can be compared.
	Provides value-added estimates—from freshman to senior year or with measures on a sample of freshmen and seniors.
	Provides percentiles.
	Provides benchmark institutions.

Figure 4. Summative Function of Accountability

Assessment
• CLA measures an important piece of the puzzle focusing on *value added* undergraduate learning
• Other measures needed
  –Major area
  –Personal, social, civic, and moral responsibility

Accountability function
• Signaling
• Benchmarking
• Measuring value added

Issues
• Low vs. high stakes?
• Publish common set of indicators?
• Incentives vs. punishment?

Abstract Process Oriented

Concrete Content Oriented

INTELLIGENCE
General   Fluid   Crystallized

GENERAL REASONING
Verbal   Quantitative   Spatial
Example: Graduate Record Examination

BROAD ABILITIES
Reasoning   Critical Thinking   Problem Solving
Decision Making   Communicating
In Broad Domains
(Disciplinary–Humanities, Social Sciences, Sciences
and Responsibility–Personal, Social, Moral, and Civic)
Example: Collegiate Learning Assessment

KNOWLEDGE, UNDERSTANDING AND REASONING
Declarative, Procedural, Schematic and Strategic
In Major Fields and Professions
(American Literature, Business)
Example: ETS Major Field Tests

Inheritance x Accumulated Experience

Direct Experience

need to be measured as well; for example, we have begun the process of adapting the criterion-sampling approach and CLA technology to assess measures of performance in specific academic disciplines. Measures of personal, social, civic, and moral responsibility are needed as well; several projects currently are engaged in experimenting with such outcomes (e.g., AAC&U, Wabash College Project).

Summative Function of Accountability

The CLA is a summative instrument that focuses on outcomes rather than on the processes that gave rise to those outcomes. Hence, summative accountability asks the question of how well, compared to other colleges or some standard, this college is performing. It sends a *signal* of where a campus is successful and where more work is needed to improve student outcomes. By estimating value added or by benchmarking with peer institutions, it addresses the question, "How good is good enough?" Without these measures, institutions cannot answer that question.

A number of contentious issues associated with summative accountability should be considered. The political pressure being placed on assessment results is one such issue. When politics enter into the discussion (and it inevitably does), the assessment switches from low to high stakes, and yet the consequences may very well outweigh the capacity of the assessment instrument. The CLA position is quite clear: high-stakes use of learning

assessments corrupts the very thing it is intended to improve—teaching and learning. Assessments are delicate instruments and cannot, alone, support the weight of high-stakes accountability. Moreover, such uses of as-

> ASSESSMENTS ARE DELICATE INSTRUMENTS AND CANNOT, ALONE, SUPPORT THE WEIGHT OF HIGH-STAKES ACCOUNTABILITY. MOREOVER, SUCH USES OF ASSESSMENTS LEAD TO BIZARRE BEHAVIOR (E.G., CHEATING, NARROWING THE CURRICULUM), AS IS QUITE EVIDENT FROM EXPERIENCE WITH NO CHILD LEFT BEHIND, THE CURRENT FEDERAL EDUCATION POLICY.

sessments lead to bizarre behavior (e.g., cheating, narrowing the curriculum), as is quite evident from experience with No Child Left Behind, the current federal education policy.

A second, related, issue is that of incentives. In the United States, at least currently, the prevailing view is that if an organization does not perform as expected, sanctions should be applied. Again, No Child Left Behind is a case in point. Yet from all we know of the psychology of reward and punishment, such a tack is unlikely to improve education in the long run. Rather, punishment suppresses some behaviors and makes others more prevalent—but the new behavior is largely symbolic, and when the sanctions go away, as inevitably they do, little if anything is changed. That said, the few examples of the application of rewards in higher education raise doubts as well; campuses respond symbolically without real change. The question to be addressed is, "How should incentives be used for the improvement of teaching and learning?"

Formative Function of Accountability

Accountability also serves a formative function—improvement of teaching and learning (see Table 3). This function serves to monitor, feed back, and act on information for improvement. While the CLA is largely summative, it can be used for formative purposes as well (Benjamin, Chun, & Shavelson, 2006). CLA performance and analytic writing prompts make good teaching tools. By using CLA-like tasks in class, instructors can, through students' writing and discussion, come to understand the strengths and weaknesses in their students' critical thinking, analytic reasoning, problem solving, and communication. Armed

Table 3. Formative Function of Accountability

Assessment
- CLA-like tasks used for teaching and diagnosing students' needs and providing feedback on how to improve
- Campus assessment program including portfolios and capstone courses/projects
  - Identify areas for improvement
  - Experiment with alternative "solutions"
  - Act on findings

Accountability function
- Monitor change
- Act on findings to improve teaching and learning
- Feedback to students, faculty, department chairs, deans, provost, and president for improvement

with this information, instructors are better positioned to close the gap between what students know and are able to do and desired outcomes.

CLA tasks are not the only sources of information for the formative function of accountability. Assessment needs to be seen as part of the teaching-learning processes of the institution, and these processes need to be supported by faculty and administration and institutionalized so that immediate feedback is available up and down the system from students to president. Capstone assessments in the form of courses and projects, portfolios of progress over the undergraduate years, and other campus-specific assessments offer important ways of augmenting external assessments and serving the formative function.

Conclusion

The CLA was built for the purpose of improving teaching and learning at the program or institution level, such as when CLA-type tasks are used for instruction. It was built to conform to the kinds of outcomes colleges highlight in their mission statements, and to signal how well a campus is performing relative to its intake of students or their benchmark peers. However, the CLA is limited in that it focuses on broad cognitive abilities and needs to be supplemented with measures of outcomes in specific majors, as well as with measures of social, moral, and civic outcomes. These are the arenas for the next evolution of the criterion-sampling approach with CLA technology.

RICHARD SHAVELSON is the Margaret Jack Professor of Education and professor of psychology at Stanford University, and director of the Stanford Education Assessment Laboratory. He served as dean of the Stanford School of Education from 1995 to 2000. He is a primary architect of the Collegiate Learning Assessment. Shavelson can be reached at richs@stanford.edu.

References

Case, R. (1996). "Changing views of knowledge and their impact on educational research and practice." In D. R. Olson & N. Torrance (Eds.), *Handbook of Human Development in Education: New Models of Learning, Teaching, and Schooling.* Oxford: Blackwell.

Benjamin, R., Chun, M., & Shavelson, R. (2006). "Holistic tests in a sub-score world: The diagnostic logic of the Collegiate Learning Assessment." Accessed Nov. 24, 2007, at www.cae.org/content/pdf/WhitePaperHolisticTests.pdf.

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). "The Collegiate Learning Assessment: Facts and fantasies." *Evaluation Review* 31(5), 415–439.

Klein, S. P., Kuh, G. D., Chun, M., Hamilton, L., & Shavelson, R. J. (2005). "An approach to measuring cognitive outcomes across higher-education institutions." *Journal of Higher Education* 46(3), 251–276.

Shavelson, R. (2007a). "Assessing student learning responsibly: From history to an audacious proposal." *Change* (January/February), 26–33.

Shavelson, R. J. (2007b). *A Brief History of Student Learning: How We Got Where We Are and a Proposal for Where to Go Next.* Washington, DC: Association of American Colleges and Universities' The Academy in Transition Series.

Shavelson, R.J., & Huang, Leta (2003). Responding responsibly to the frenzy to assess learning in higher education, *Change,* 35(1), 10-19.