

Book Review

Measuring College Learning Responsibly.

Richard J. Shavelson. Stanford, CA:
Stanford University Press, 2009. 207 pp.
ISBN-13:978-1455508389. Hardcover, \$26.99.

REVIEWED BY:

Jeff Kosovich, M.A.; Rory Lazowski, Ed.S.;
Oksana Naumenko, M.A.; & Donna L. Sundre, Ed.D.
James Madison University

Editor's Note: The following review has been co-authored by three graduate students and their instructor. Please note the introduction and conclusion are from the point of view of the faculty member and the body of the review is co-authored by the graduate students.

Richard Shavelson's book, *Measuring College Student Learning Responsibly*, was an answer to a real need for my graduate Psychology 812 course, "Assessment Methods and Instrument Design," a core requirement for Quantitative Psychology concentration Master's and Assessment and Measurement PhD students. A weekly course feature is student written reflections on each assigned reading. The book promised to discuss assessment, accountability and accreditation in the United States and to provide an international perspective. Given Rich Shavelson's prominence as a researcher and instrument developer, the book beckoned. This book promised to fuel our weekly seminar conversations and to provide just the kind of heat and controversy to inspire deep learning and engagement. When offered the opportunity to review this book, the perspectives of my fall 2012 students seemed the ideal ingredient; three of the best students from that cohort were recruited. Throughout the course, their unique perspectives were inspiring and remind us that we are all students together. Enjoy, as I did, the thoughts, reflections, and, yes the rants, of these students as they team to review each chapter.

Assessment and Accountability Policy Context

In the opening chapter, Shavelson wastes little time delving into the impact of the Spellings' Commission (U.S. Department of Education, 2006) recommendations and their impact on higher education assessment activities. One of the most salient issues that Shavelson examined is the continuing "tug of war" among the various cultures and stakeholders involved in assessment: the academic culture, clients (parents, students, government agencies, businesses), and policy-makers. Shavelson notes that the Commission's report took a client-centered stance, noting that universities should be transparent about student learning, the inherent value-added of attending a university, and the outcomes associated with a costly education. Simply put: accountability. It appears as though those in the policy-maker community share this same perspective. This stance

is in contrast to the academic culture, which largely views assessment as leading toward instructional improvement in teaching and learning. This chapter outlined many of the recommendations in the Spellings report and how these recommendations were met by each culture.

Another central concept is the need for institutions to focus on sound assessment methodology. A recurring, albeit previously unheard of theme Shavelson promotes is that institutions may not be ultimately judged on assessment outcomes, but instead by assessment program quality. This is important because many institutions may fear that unsatisfactory results will lead to undesirable repercussions. Keeping quality practice at the forefront can assuage fears and influence more positive engagement in assessment practices. This chapter provides a focused and cogent treatment to some of the most persistent and pervasive issues in higher education assessment, issues that are more fully developed and explained in subsequent chapters.

Framework for Assessing Student Learning

Chapter 2 provides a host of information to consider for best practices in assessment, from how students should be assessed to the range of knowledge, skills, and abilities to be incorporated into an assessment plan. It further guides the reader through proposed measurement and student learning best practices for both accountability and improvement purposes. This treatment provides an accessible framework for designing an assessment plan and ways to improve existing plans.

A recurring, albeit previously unheard of theme Shavelson promotes is that institutions may not be ultimately judged on assessment outcomes, but instead by assessment program quality.

Shavelson outlines several key considerations to student learning assessment. He introduces an essential and critical distinction between learning and achievement: learning requires the observation and measurement of cognitive growth over time, while achievement provides only a single assessment of cognitive ability. Another important assessment best practice was offered through the distinction of the definitions and value of direct vs. indirect measures of student learning. Moreover, Shavelson recommends assessment designs that allow for both comparability across institutions and diagnosis within. This chapter describes a learning outcomes framework and advises us not only on what should be assessed (including soft skills), but also on how to assess efficiently.

Brief History of Student Learning Assessment

In Chapter 3, Shavelson describes historical roots of assessment and notes significant trends for student learning

outcomes. Of particular note, Shavelson provides a very useful discussion concerning the transition from internal institutional initiation of assessment to more external demands. This transition is coupled with a review of the paradigm shifts that accompanied student learning assessment purposes. Shavelson provides an excellent summary of landmark tests from the past century (e.g., the Pennsylvania Study, 1928-32; GRE) and notes that past perspectives of learning, thinking, and the purpose of higher education are still present in the way we measure learning today. However, Shavelson contends that standardized tests are incapable of providing granular data that can impact instruction and actual learning. He believes internal instruments are necessary to supplement the blunt standardized tools used for accountability purposes. This represents a challenge that few assessment practitioners will be able to address efficiently or effectively. Curiously, Shavelson neglects discussion on the topic of performance assessment until late in the chapter despite its long history (Johnson, Penny, & Gordon, 2008). However, he notes an important paradigm shift in what is considered a critical outcome of higher education: fact-based and procedural knowledge was of past value; broad skills and abilities such as critical thinking and verbal reasoning currently dominate today's standardized testing contexts. It is here that he introduces the College Learning Assessment (CLA). The CLA represents one example of large scale assessment that avoids the multiple choice format; however, the level of consideration it receives in this chapter and the rest of the book cumulatively becomes more of a very thinly-veiled attempt at product placement than an objective treatment of student learning assessment.

The Collegiate Learning Assessment

Chapter 4 focuses entirely on the development, philosophy, and strengths of the CLA, often citing promising results that require further investigation. Shavelson wittingly showcases the “prominence” of the CLA, framing it within the context of reliability and validity. First, although some of the presented reliability values are acceptable, those at the lower end of the range are quite frightening for a potentially high-stakes test. Given that test results may be used to provide information about school accreditation, funding, and diagnostic decision-making, it is necessary to ensure that scores are of high quality (i.e., good reliability and validity). Second, the use of “holistic” scores reported by the CLA seems to ignore the benefits and recommendations of the field of psychometrics, as well as the earlier plea for more granular, diagnostic data to drive learning improvement. Even if the test operates on the assumption that the test is ‘greater than the sum of its parts,’ the parts are not to be discarded. What does a total score of 1150 mean? This is the heart of validity. Third, evidence of high student motivation to perform well on this instrument appears to be assumed. Shavelson defends the CLA by arguing that test motivation is a concern with any standardized test; however, poor motivation, particularly on

arduous performance tasks, brings the validity of test scores into question.

Finally, the CLA is said to be useable as an index of value-added, which is problematic. In terms of psychometrics, difference scores (i.e., the differences between test scores at two time points) tend to be even less reliable than test scores. This lack of reliability is problematic because the desire to compare schools on value-added is a major driving force in higher education accountability. If the test scores are inconsistent—and the value added scores are even less consistent—should we tempt policy makers to use these scores for accountability purposes? The problem with

Pitching the CLA as a measure of value added is essentially implying that the CLA measures the whole of student learning in some meaningful way. Shavelson repeatedly acknowledges that the CLA requires better validity evidence, yet he presents the CLA as a panacea to the woes of higher education assessment.

the value-added scores is further compounded by the way in which value-added is defined. Pitching the CLA as a measure of value added is essentially implying that the CLA measures the whole of student learning in some meaningful way. Shavelson repeatedly acknowledges that the CLA requires better validity evidence, yet he presents the CLA as a panacea to the woes of higher education assessment. Given Shavelson's expertise in psychometric matters (e.g., Shavelson & Webb, 1991), it is disappointing to see gross misrepresentation of the CLA's quality.

Exemplary Campus Learning Assessment Programs

In Chapter 5, Shavelson discusses the internal (assessment for improvement) and external (accountability for comparison) foci of assessment, displaying the diversity of assessment practice through the discussion of two named and four unnamed universities (one of which was clearly ours) that serve as exemplars of particular assessment practices. These profiles were useful in identifying distinctive characteristics and attributes among existing assessment programs and provide a convenient summary table (pp. 81-82). After describing each model in detail Shavelson came to several conclusions: 1) it is campus leadership that assures and inspires quality assessment programs, not accreditation agencies; 2) although all programs were outcomes-based, they significantly differ in the focus of their programs; 3) faculty engagement is critical; 4) explicit, measureable learning outcomes are key to appropriate use of data; 5) successful assessment programs require champions from diverse perspectives; 6) data must provide relevant information to guide faculty; 7)

incentives are required for policy, hiring, and rewarding assessment practice; and 8) practice must be balanced and sustainable to thwart morale problems. The value of the chapter is in the elucidation of several characteristics that differ broadly across existing model programs. This chapter was a highlight of the book and should have great utility for practitioners and those planning assessment centers.

The Centrality of Information in the Demand for Accountability

In his discussion of the centrality of information in Chapter 6, Shavelson briefly lays out the nature of higher-education accountability. Overall he does an excellent job introducing readers to the purpose of accreditation as well as the role of accrediting bodies. More importantly, Shavelson gives due treatment to the summative versus formative debate that has haunted assessment since the very beginning. The business-driven philosophies that have contributed to continued socio-political conflict within assessment are also examined. The bulk of the conflict appears to stem from the different types of information demands from various stakeholders. A particularly interesting contrast is made between politicians and consumers (e.g., students, parents). Shavelson also gives a nod to the Voluntary System of Accountability (VSA). His discussion of the VSA is one of the weaker points of the chapter because it is only portrayed in a positive light. Readers wooed by the VSA coverage should consider its success and shortcomings more deeply before endorsing it outright. For example, the participation rate in the VSA has plummeted since its inception. Ironically, much of the disillusionment with the VSA stems from its very reliance on standardized tests (Jankowski et al., 2012) that Shavelson himself earlier laments (while, even more ironically, steadfastly endorsing the CLA). In general, Chapter 6 serves as a primer for the later portions of the book dedicated to the topic of accountability.

Accountability: A Delicate Instrument

This chapter addresses complications that arise in higher education accountability efforts. At face value, accountability may appear reasonable and practical; however, these complications can lead to unintended consequences if applied inappropriately. Shavelson discusses six complications that underscore the notion that accountability is both a powerful tool and a delicate instrument. The complications include: 1) accounting for assessment outcomes vs. processes; 2) consideration of what processes and outcomes to measure; 3) problems associated with the use of sanctions (and/or incentives) as a vehicle for change or improvement; 4) the functions of formative vs. summative assessment purposes to inform multiple audiences; 5) appropriate and ethical interpretation of results; and 6) the balance between complying with external expectations and maintaining flexibility and innovation in assessment

efforts. These complications highlight more global issues related to accountability that are important considerations. Although we persevere on development of more specific measurement and instrument design (with good reason), this chapter helps to identify the “forest” issues that may hide the accountability system “trees.” The reader can take a step back and reflect on how delicate these considerations may be and how important it is to think intelligently about these more overarching accountability issues. This chapter is highly recommended for policy-makers.

State Higher-Education Accountability and Learning Assessment

Similar to the previous chapter, chapter eight provides another global view of the issues in higher-education learning assessment and accountability with a focus on states’ influences. Shavelson delves further into accountability policy by examining how learning is assessed in US higher education. One of the central points of the chapter is the catch-22 facing higher education institutions.

The mad dash by various organizations to compare and rate the effectiveness of institutions often leaves stakeholders with piles of uninterpretable numbers.

The mad dash by various organizations to compare and rate the effectiveness of institutions often leaves stakeholders with piles of uninterpretable numbers. For example, the sheer volume of learning indicators can result in different institutions measuring different constructs with different instruments, making any meaningful comparison impossible. To make matters worse, the indicators reported (e.g., graduation/retention, enrollment, tuition) essentially define what is and is not important in assessment in the eyes of some stakeholders. This cycle of confusion and disorganization leads to wave after wave of expensive, loosely-aimed assessment initiatives. These initiatives provide minimally-useful information while simultaneously contributing heavily to institution-level decisions. The focus on measures needs to be shifted from indirect measures, such as graduation rates and retention, to more direct measures of learning. Overall the chapter provides an even-handed treatment of the issues. Unfortunately, the topic of direct measures opens up Shavelson’s obligatory CLA pitch for the chapter.

Higher Education Accountability Outside the United States

This chapter outlines international assessment of student learning. Countries other than the U.S. that are covered in this chapter approach accountability in qualitatively different ways by employing a four-stage quality assurance process rather than isolated assessment programs. Instead of looking at several different input and

output indicators, the countries that address accountability focus more on “quality assurance processes.” Shavelson describes these as more focused on the processes that ensure quality education. This diagnostic, monitoring approach has many advantages over the U.S. tradition including continuous access to evidence-based practices, by advancing and reinforcing continuous learning improvement. Additionally, Europe and Australia do not appear to wage wars on colleges and universities within this system, though their faculty may disagree. Shavelson is correct in pointing out that these quality assurance practices seem sustainable and should be noted by American policy-makers. He explains why the academic audit stage is the one stage from which the U.S. can learn. In the context of other nations’ federal burden of funding higher education outside of the U.S., it appears justified that universities conform to external statutes. Shavelson highlights a stark contrast between other nations’ explicit responsibility in funding higher education and the invasive nature of the U.S. government policies, which are responsible for only small portions of higher education funding. Overall, this chapter represents an excellent example of “compare and contrast” between higher education learning measurement across nations from which policy makers can and should learn.

Learning Assessment and Accountability for American Higher Education

This final chapter nicely ties together information discussed in previous chapters to form a vision of assessment in higher education. In doing so, Shavelson integrates the existing body of knowledge presented in the book and attempts to conceive a better system for United States learning assessment and accountability. He effectively addresses many of the extant tensions, with potential solutions to these tensions, both in terms of substance (what should be measured) and structure (how to measure it). Whether or not it can be achieved is another question. However, credit is due for putting forth pragmatic strategies for resolving the reconciliation of assessment and accountability issues.

It would have been helpful to provide alternative assessment options for those who are not interested in using or cannot afford the CLA.

In yet another attempt to advertise, Shavelson recommends the CLA to achieve this vision of assessment in higher education. It would have been helpful to provide alternative assessment options for those who are not interested in using or cannot afford the CLA. Nevertheless, there are some clear guidelines set forth regarding what should be measured and how to measure it. This can be particularly valuable to assist institutions struggling to improve in their assessment and accountability efforts.

Summary

You may agree or disagree with the views provided by these graduate students; however, it cannot be ignored that Shavelson’s book provided a powerful accelerant to provide not just a fire—perhaps a bonfire—of graduate level discussion. On several occasions we basked in the heat and glow of these conversations discussing all of the topics so well covered. These are important topics for all interested in assessment and accountability; we invite you to join the conversation and make positive change. For those less desirous of potential conflict, there are at least two other viable choices for textbooks or learning more about higher education assessment practice. Linda Suskie’s (2009) 2nd edition, fully lives up to its title, *Assessing Student Learning: A Common Sense Guide*, and provides what many would consider a desktop reference. Another excellent choice would be Astin and Antonio’s (2012) 2nd edition *Assessment for Excellence*. For a recent review of the latter by Linda Sax (2012), please look no further than *Research & Practice in Assessment*. Both books are excellent new additions to our assessment bookshelves.

References

- Astin, A.W., & Antonio, A. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education* (2nd Edition). Lanham, MD: Rowman & Littlefield and the American Council on Education.
- Jankowski, N. A., Ikenberry, S. O., Kinzie, J., Kuh, G. D., Shenoy, G. F., & Baker, G. R. (2012). *Transparency & accountability: An evaluation of the VSA College Portrait Pilot*. (A Special Report from NILOA for the Voluntary System of Accountability). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. Retrieved from http://www.learningoutcomeassessment.org/documents/VSA_000.pdf
- Johnson, R. L., Penny, J. A., & Gordon, B. (2008). *Assessing performance: Designing, scoring, and validating performance tasks*. New York, NY: Guilford Press.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. New York, NY: Sage Publications, Incorporated.
- Suskie, L. (2009). *Assessing student learning: A common sense guide*. (2nd Edition). San Francisco, CA: Jossey Bass.
- U. S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: Author.