

ROGER BENJAMIN

## **THE PRINCIPLES AND LOGIC OF COMPETENCY TESTING IN HIGHER EDUCATION**

### INTRODUCTION

The aim of this chapter is to argue that more attention should be devoted to student learning assessment in higher education, using multiple types of instrument, some of which link directly to teaching and learning in the future. Examples from the case of the U.S. will be used throughout, as this case is illustrative of the underlying trends faced by all countries, to one degree or another. The essay moves through these steps in order to justify the following assertions. A number of trends suggest that the next few decades will bring substantial restructuring in the higher education sector. This restructuring will require much more evidence-based decision-making because the stakes will be high. In turn, this places the focus on the quality of student learning, which is a critical outcome of higher education institutions. Second, the challenges ahead are sufficiently serious that widespread debate will occur about how to resolve them. Third, the work required to generate appropriate responses in order to assess student learning will be discussed, including the central role of faculty. Fourth, the case for performance assessment, now being widely explored in the U.S., will be presented, followed by a short set of recommendations.

### THE RATIONALE FOR EVIDENCE-BASED DECISION-MAKING

A combination of factors has created an unprecedented crisis in undergraduate education in the U.S.<sup>xvi</sup> Access deficits caused by 47 million high school dropouts (equaling one-sixth of the U.S. population), a college readiness gap signified by 40% of new students who cannot read, write or perform math at college level, and only 57% of students graduating within six years present enormous challenges to higher education in the U.S.<sup>xvii</sup>

The 47 million high school dropouts alone constitute a massive deadweight on the economy. These citizens have been denied the tools they need in order to be productive; they are largely employed in minimum wage service jobs or are not even in the labor force. Their economic and social prospects are bleak, and as they represent such a large percentage of society, America's prospects for both economic growth and reducing inequality are also becoming increasingly problematic.

Rising costs, now combined with declining revenues in higher education, make it much more difficult to reverse this situation. Instead, they exacerbate what has

ROGER BENJAMIN

already become an overall market failure, an example of what political economists such as recent Nobel laureate Elinor Ostrom term a “common pool problem” (CPP). CPPs arise whenever there is confusion or conflict over a public good, e.g., regarding who pays what proportion of the costs and who gets what proportion of the benefits, or where one person's use affects another's ability to use the assets, or when groups, public or private, fail to provide the resources, over-consume them and/or fail to replenish them. American higher education is a public good. As with other public goods, an attempt has been made to shift a great deal of the responsibility for funding public higher education (which is attended by the majority of students) from the state governments to the students through higher tuition fees. However, while tuition has increased significantly, imposing higher financial burdens on students and their families, annual funding increases have not kept pace with cost increases in higher education, which have been, on average, 1% or more higher than the Cost of Living Index (CPI) (see Griswold, 2006). The net effect has been accelerating cuts to academic programs, student aid and infrastructure for colleges. Under these conditions, little headway has been made in providing training opportunities for the large high school dropout group which keeps increasing. When these kinds of conditions apply to a public good like higher education, it warrants the CCP label. When the CPP becomes acute, as in today's higher education sector, either bold action must be taken to solve it or the CPP will become a permanent crisis (see Grant Hardin's tragedy of the commons, 1968).

Efforts made thus far to deal with the CPP have been defeated because the system of incentives that guide the behavior of college administrators, faculty, staff and other supporting stakeholders is not organized to consider it a problem or, in any case, is not focused on solving it. The incentive system of research universities, and the model for faculty and administrators in the rest of the postsecondary education sector, privileges research and scholarship, not teaching and learning. High school dropouts are not seen as coming under the scope of the postsecondary education sector's mission.

All major institutions in society are highly resistant to change from within; the postsecondary education sector is no exception. The imperative for redesign to deal with the CPP, if it arises, is most likely to stem initially from external sources. The elements that make the imperative possible are plain to see. The CPP, framed by a combustible mixture of rising costs and declining revenues, is now viewed through the emerging consensus that human capital (the knowledge, experience and education levels of a nation's citizens) is clearly the principal resource in the U.S.. This consensus should eventually lead to widespread agreement among public and private leaders that as the principal duty of the state is to guarantee the security of its citizens, this now means preserving and enhancing the quality of its human capital.

In short, there will soon be a significant national debate in the U.S. about the undergraduate CCP that will cut across political party lines. We can see this presaged in debates about K-12 education. If human capital is the principal national resource, education should be recognized as the key to success in all other policy areas, such as health, economics, the environment, energy, agriculture and

## PRINCIPLES AND LOGIC OF COMPETENCY TESTING

national security. This means that the quality of education should be the central priority of the national government. The human capital argument will create the basis for new and higher limits for the role of education, because leaders will connect the dots and come to understand the critical importance of dealing with the common problem facing K-16 education, which dwarfs all other issues that America is currently facing.

### THE CHALLENGES AHEAD

Will the combination of rising costs and declining resources, coupled with the growing perception of the centrality of human capital, provide strong enough external forces to compel the resolution of the CPP? This combination is strong enough to provoke a national debate about the need to tackle the CPP, but this does not mean that the CPP will be resolved.

Unless a way is found to create an effective institutional redesign strategy that faculty and administrators in postsecondary education will buy into, the next decade will be a period of turmoil, with continuing cost and resource problems accompanied by growing quality and access deficit issues. For example, because 80% of the potential growth in access to postsecondary education will come from the Hispanic population, many of whom are high school dropouts, improvements to access, retention and graduation rates will be problematic without new approaches to the problem.

The CPP, combined with the effects of the disruptive force of Internet-based education solutions, has created the prospect of substantial restructuring and redesigning of the postsecondary education sector over the coming decade. Examples include the following:

- Mission differentiation in order to address the need to sharpen the focus of colleges instead of pursuing multiple missions simultaneously;
- Identification of the gaps in the quality of student learning between African-American and Hispanic students on the one hand, and non-Hispanic Whites and Asian-Americans on the other hand, with analysis-based recommendations about what to change in order to reduce the gaps;
- Description of the extent and nature of student learning deficits at the national, state and institutional level, as defined by the Carnegie classification. Development and implementation of recommendations to improve student learning;
- Identification and implementation of advances in pedagogy that will improve student learning outcomes;
- Assessment of the benefits and costs of advances in educational technology with regard to student learning growth;
- Description and analysis of the impact of resources on student learning outcomes.

ROGER BENJAMIN

The benchmarking of student learning outcomes, which is only now becoming widespread, is a necessary but not sufficient prerequisite for these and all other prospective attempts to restructure and redesign postsecondary education, in an effort to respond to the CPP. This is because one needs a metric against which to evaluate the benefits over costs against the dependent variables that one is attempting to solve or improve. The outcomes of student learning in undergraduate education are appropriate candidates for this role.<sup>xviii</sup>

Regardless of what actual scenario plays out over the difficult decades while lie ahead for postsecondary education, empirical evidence, including that which is based on the assessment of student learning, will play a much greater role. Without evidence based on credible research, little progress will be made in dealing with the CPP, because it will not be possible to generate accurate descriptions and analyses upon which to base recommendations and solutions.

#### THE RESPONSE NEEDED TO ASSESS STUDENT LEARNING OUTCOMES

Due to the size and importance of the issue, we need all hands on deck. Advocates of portfolios are creating important best practice models for faculty in the classroom to emulate. There are many other efforts to provide ways for individual faculty members to directly assess the results of their teaching and learning. We need also to recognize the contributions of cognitive scientists, who now have much to teach us about how the brain learns (Miller, 2003). Education technologists will be needed in order to scale up the ideas which are created (see, for example, the Open Education Resource (OER) movement). Measurement scientists are also needed because they insist upon measurement instruments with demonstrable validity, meaning they measure what the instrument is intended to measure, are given to students under the same conditions and are based on reliable scoring rubrics. Due to the high stakes involved, direct measures of student learning that meet the highest standards of reliability and validity will be required in order to provide the systematic evidence needed to make the many decisions that will affect resource distribution and the improvement of the quality of student learning over the coming decades.

#### THE CENTRAL ROLE OF THE FACULTY

The discussion about assessment and accountability tends to focus on policy issues or the reliability and validity of assessment instruments. These are, of course, important issues. However, a discussion of the relevance of the assessment instrument to teaching and learning is either completely absent or approached as an afterthought. The threshold question is the instrument's relevance to the faculty in the classroom. In addition, the relevance of the assessment instrument to the faculty in the classroom should take precedence over its technical dimensions and larger policy debates over whether or how assessment or accountability should occur. In other words, the assessment instrument must be known to be reliable and valid, but this should only be a necessary, and not sufficient, condition for its

adoption by a college or university. Do the faculty find the assessment instrument useful? That should be the most important question.

The faculty should be the focus of assessment, because individual instructors are at the center of matters relating to teaching, learning and the curriculum. The implication of this point is that faculty buy-in is critical to the future of assessment and accountability in the academy. Until it is clear that testing organizations have developed assessment instruments that are accepted by the faculty as valuable aids to their instruction, it is unlikely that we will move forward in the policy debates on assessment and accountability in higher education. Thus, our focus should be on encouraging the faculty to use assessment instruments that are in line with their teaching and learning goals.

If the faculty buy in to using assessment instruments as central tools to monitor and improve teaching and learning, this will increase the probability of positive developments on other fronts, such as accountability and the use of assessment-based evidence for internal governance and diagnostic purposes, because it will be possible to base these other activities on assessments which faculty members perceive to be authentic.<sup>xix</sup> However, this process must begin with the faculty recognizing the inherent value of assessment to their own work as teachers. This will occur only if the assessment tools themselves are proven to be effective for the cycle of teaching, learning and assessing for continuous improvement. Of course, additional significant changes are needed in order to make this equation work. The faculty must have incentives to encourage them to focus on student learning rather than research alone, and students need support and encouragement to learn.

#### THE RATIONALE FOR PERFORMANCE ASSESSMENT<sup>xx</sup>

The current assessment regime, dominated by multiple-choice tests, is no longer sufficient in the knowledge economy. For a century, multiple-choice tests have been the principal assessment method in education. This probably made sense in the industrial era of development, as this method mirrors the focus on the mastery of content demonstrated by students' ability to recall facts. Today, we live in an economy dominated by information and services rather than physical goods. In the knowledge economy, it is more important to be able to access, structure and use information than merely recall facts. This places a premium on the ability of students to reason, assess the relevance of information and make arguments; in short, to think critically. This effort to focus on critical thinking skills is being implemented in classrooms across the country, in which faculty are arming their students to navigate a constantly changing world defined by an ever-increasing volume of information. The manner in which we assess students must reflect these interests. Multiple-choice tests may present examples of correlations and causation and ask students to identify whether each is correctly or incorrectly applied. However, responding passively to such choices is very different from asking students in performance assessments to actively critique a case study that presents an argument about data in which the concepts of correlation and causation are misused. It is also important to underline the requirement in the knowledge

economy for citizens to actively shape the information at their disposal, rather than simply to respond passively to choices put before them.

Assessment must therefore catch up with an emerging reform agenda in higher education, resulting from our new understanding of student learning. At the most basic level, this involves understanding that the meaning of knowledge itself is undergoing a significant shift. New theories from the field of cognitive science stress the importance of improving students' ability to structure learning experiences that help them to use what they have learned in new settings. Simon argues that the meaning of "knowing" has changed from being able to recall information to being able to find and use it (Simon, 1996, p. 43). Under these conditions, the proponents of the new learning theories argue that active learning is critical, because students must learn to recognize when they understand a subject and when they need more information (Pellegrino, Cudowsky, & Glaser, 2001). The implications for higher education are profound. If we consider the assumptions which structured higher education in the industrial era, the lecture format was the norm, with students seen as passive receptacles receiving the content provided by lecturers. The role of higher education was to transmit knowledge. Faculty and administrators were comfortable with these assumptions, because even though it was understood that knowledge was progressing in multiple fields, most shared the view that there was a stable, enduring stock of knowledge that graduating seniors should know. Under these circumstances, content was emphasized and multiple-choice tests were the preferred assessment tool. However, although I am arguing for a greater emphasis on performance assessment in the 21<sup>st</sup> century knowledge economy, this does not mean that multiple-choice tests are inferior. Good performance assessments, just like good multiple-choice tests, must be constructed on clear definitions of what students should know and be able to perform. They then must measure that domain. There are no silver bullets in assessment instruments, just as there are no universal solutions to the improvement of student learning in higher education. Now that the Internet and computer-assisted scoring have made it feasible to scale up performance assessment, it is time to explore the practical possibilities of this testing paradigm more fully.

Over the past two decades, it has become clear that a new vision of undergraduate education is developing in response to the changing definition of knowledge. It is comprised of three parts:

- A shift from the lecture format to a student-centered approach that emphasizes analytical writing. Faculty are much more interested in active student participation in the learning process, and students appear to be equally interested in doing so. Although evidence is still in the formative stage, it appears that colleges that emphasize analytical writing produce students who do well in assessments that benchmark higher-order skills;
- There has been a change in emphasis from the pre-existing focus in curricula and texts on content to case- and problem-based materials that ask students to apply what they know to new situations. This is reflected in curriculum reform and is also resulting in textbook publishers producing solely content-filled

- volumes. The graduate business school emphasis on the case approach to learning may be an early example of this strategy;
- There has also been a change in assessment from multiple-choice and short answer formats to open-ended essays that are better aligned with the first two parts of the reform.

Performance assessments (constructed responses that require students to demonstrate their ability to perform tasks) appear to be better aligned with the focus of this education reform movement. Performance assessments are congruent with recent theories of learning and knowledge that focus on applying what one knows to new situations, typically including the ability to think critically, solve problems and write effectively. The Internet and computer-assisted scoring have enabled performance assessments to be administered, scored, analyzed and reported to students and their instructors in increasingly cost-effective and accurate ways. Faculty do not like multiple-choice tests, but perceive performance assessments as being authentic. This means that performance assessment, uniquely, can be used in both the standardized and formative assessment space. This is encouraging, because the only way to create a sustainable assessment system in postsecondary education is to create a more systematic, continuous system of teaching and learning improvement based on assessment instruments that all parties (the faculty, administrators, governmental authorities) can use in order to achieve their objectives.<sup>xxi</sup>

## CONCLUSION

In today's knowledge economy, it is more important to be able to access, structure and use information than merely to accumulate facts. Performance assessments are appropriate for benchmarking and stimulating the development of the necessary critical thinking skills. As there are no "correct" answers to performance tasks, they are worth teaching to. Moreover, in the case of the Collegiate Learning Assessment (CLA, a performance assessment used in the U.S., 2010), because it is not focused on discipline- and content-based knowledge, its use does not narrow the curriculum.

Recommendation One: Consider adopting performance assessments for higher education, because they can be used in both formative and standardized assessment applications. They are unique in this sense. This does not mean that performance assessment of critical thinking skills should be the sole approach to assessment in higher education. Performance assessment should be combined with multiple other forms of assessment.

Recommendation Two: Technology is a key enabler for possible advances in assessment. Performance assessment has been around for a long time, but the Internet unlocked its large-scale potential because complex tasks can be placed on an Internet website and administered, scored, analyzed and reported back to the student and their college with fewer errors and in a cost-effective way. The advent of computer-assisted scoring has created the opportunity for on-demand testing in

ROGER BENJAMIN

the classroom, because the rapid turnaround of test results is now possible (see the description of computer-assisted scoring in Appendix 1). In addition, consider the potential use of sharing best practices for teaching, learning and assessing on the new OER Internet platforms now being developed.<sup>xxii</sup>

Recommendation Three: Consider collaboration between cognitive scientists, educational technologists, formative assessment experts and measurement scientists in order to create more unified approaches to teaching, learning and assessment.<sup>xxiii</sup> A much greater effort than is currently being made anywhere in higher education will be required to improve teaching and learning, which is the fundamental goal of assessment.

#### APPENDIX<sup>xxiv</sup>

##### *Computer-assisted scoring*

Computer-assisted scoring employs computer models to score open-ended assessment responses. These models are created from the scores assigned by trained and calibrated graders. The computer uses these grades to operationally infer the rubric and scoring scale. The computer-assisted scoring process does not deal with the content of a complex performance assessment; instead, it is dependent on the scoring of human experts. Several hundred expert-scored student responses are used to train the computer-assisted scoring engine. The computer-assisted scoring engine “learns” the features and characteristics of the scoring rubric and each score from the expert-scored responses, which it uses to evaluate student responses. The engine relies on the collective wisdom of the expert scorers, reflected in the scores they assigned to a representative set of actual student responses. Much like the training of human scorers, the engine “learns” how to score student responses through repeated exposure to expert-scored examples.

Once approximately 500 student responses have been double scored by experts and the quality of the task has been verified, the results of the experts’ scoring are used to generate the computer-scoring model. The computer-assisted scoring engine is presented with the complete text of approximately 500 student responses, along with the experts’ scores. The engine examines the content and structure of each response and associates the information with the score assigned in order to create a model of what each point “looks like.”

In sum, the computer-assisted scoring approach has been shown to be as accurate as expert scorers, and in some cases, more accurate than expert scorers. The use of computer-assisted scoring allows testing organizations to offer accurate, fast and cost-effective value-added assessment services to institutions of higher education.<sup>xxv</sup>

#### NOTES

<sup>xvi</sup> This chapter is based on examples from the American postsecondary education sector, set in the American social, economic and political context. However, most political leaders now understand



## PRINCIPLES AND LOGIC OF COMPETENCY TESTING

- the vital need to ensure and enhance the skills of their workforce and the human capital of their citizens. Many countries are also facing major CPPs as a function of immigration, population growth, a lack of resources or the absence of postsecondary education institutions of a sufficient size and quality to improve the human capital in their country. See Benjamin (forthcoming, 2012) for a more extensive version of the points made in this chapter.
- <sup>xvii</sup> Moreover, a recent Social Science Research Council study found that student learning in colleges was not as high as previously thought (Arum & Rotska, 2011).
- <sup>xviii</sup> For research, the other principal public good produced by colleges and universities, there are a number of empirical-based metrics that permit serious examination of the factors that produce stronger research programs.
- <sup>xix</sup> For example, faculty may be able to reclaim governance over the undergraduate curriculum (Benjamin, 2007).
- <sup>xx</sup> Performance assessments are designed to evaluate the ability of students to apply what they know to new situations. For an example of a performance assessment, see <http://starttest.com/7.0.0.1/programs/clacross/Practice%20Test%20Page.htm>, which displays a collegiate learning performance task used in many colleges in the U.S.
- <sup>xxi</sup> The performance assessment paradigm has recently been embraced by the U.S. Department of Education and the Gates Foundation. Due to a number of grants, performance assessments are being developed for college readiness tests (see website of the SMARTER Balanced Assessment Consortium <http://www.k12.wa.us/smarter/>). One performance assessment, the CLA, with which the author is associated, is used extensively by U.S. higher education institutions. For a description of the assessment instrument see the paper of R. Shavelson in this volume (Part 1). See also *The Architecture of the CLA* (Benjamin et al., 2009).
- <sup>xxii</sup> See the description of the OER on the Hewlett Foundation website (<http://www.hewlett.org/programs/education-program/open-educational-resources>). The purpose of the OER is to place intellectual property such as textbooks, curriculum materials and disclosed assessment instruments on the Internet as open resources which are available to use for free by faculty, teachers or any interested parties.
- <sup>xxiii</sup> A best practice example of such an approach that combines cognitive science, subject matter specialists, assessment specialists and education technologists is the Open Learning Initiative at Carnegie Mellon University (see <http://oli.web.cmu.edu/openlearning/>).
- <sup>xxiv</sup> This description is an excerpt from an essay on computer-assisted scoring by S. Elliot (2011).
- <sup>xxv</sup> I would like to thank the conference organizers Sigrid Blömeke and Olga Zlatkin-Troitschanskaia for organizing such a stimulating conference. I would also like to thank the two referees for their helpful comments. Finally, I would also like to thank Christiane Kuhn for her gracious hospitality before, during and after the Berlin conference.

## REFERENCES

- Arum, R., & Roska, J. (2011). *Academy adrift: Limited learning on our campuses*. Chicago: University of Chicago Press.
- Benjamin, R. W. (2007). Recreating the faculty role in governance in research universities. In J. Burke (Ed.), *Fixing the fragmented university* (pp. 70–98). Bolton: Anker Publishing.
- Benjamin, R. W. (forthcoming, 2012). The new limits of education policy: Avoiding a tragedy of the commons. London: Edward Elgar.
- Benjamin, R., Chun, M., Kugelmass, H., Nemeth, A., & Shavelson, R. (2009) The Architecture of the CLA. New York: CAE. Retrieved from [collegiatelearningassessment.org/files](http://collegiatelearningassessment.org/files).
- Carnegie Mellon University. *Open Learning Initiative*. Retrieved from <http://oli.web.cmu.edu/openlearning/>.
- Council for Aid to Education. (2010). *CLA performance task*. Retrieved from <http://starttest.com/7.0.0.1/programs/clacross/Practice%20Test%20Page.htm>.
- Elliot, S. (2011). *Computer-assisted scoring of performance tasks for the CLA and CWRA*. Retrieved from <http://www.collegiatelearningassessment.org/files/ComputerAssistedScoringofCLA.pdf>
- Hardin, G. (1968). The tragedy of the commons. *Science*, *162*, 1243–1248.
- Griswold, J. (2006). *Higher Education Price Index (HEPI) report*. Retrieved from <http://www.commonfund.org> (Common Fund Institute).
- Miller, G. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144.
- Pellegrino, J., Chudowsky, N., & Glaser, R.. (2001). *Knowing what students know: The science and design of educational assessment*. Washington DC: The National Academy Press.
- Simon, H. (1996). *The sciences of the artificial*. Boston: MIT Press.
- The William and Flora Hewlett Foundation. *Open educational resources*. Program of Grant Making. Retrieved from <http://www.hewlett.org/programs/education-program/open-educational-resources>.

*Roger Benjamin*  
*Council for Aid to Education (CAE), USA*