

# Can Assessment for Accountability Complement Assessment for Improvement?

By **Trudy W. Banta**, professor of higher education and senior adviser to the chancellor for academic planning and evaluation, Indiana University–Purdue University Indianapolis

*All of this focus, of the media, of quality assurance and of institutions, is on assessment as measurement. . . . We should design assessment, first, to support worthwhile learning. . . . Standards will be raised by improving student learning rather than by better measurement of limited learning.*

—Graham Gibbs, Oxford University, UK;  
and Claire Simpson, Open University, UK

These are times of great uncertainty and challenge for those of us who have devoted our energies to encouraging faculty and student affairs colleagues to assess student learning outcomes for the purposes of improving academic progress and student services. We know the academy has been slow to realize the need for assessment to guide improvement, but now some faculty in virtually every institution are at least trying it out. Certainly the regional and disciplinary associations have been emphasizing outcomes assessment, and this is making a difference at most institutions today. Creative work on new tools to assess critical thinking, reflective judgment, and deep learning are being developed by faculty, some individually and some in consortia such as the one Wabash College is leading ([www.liberalarts.wabash.edu/nationalstudy](http://www.liberalarts.wabash.edu/nationalstudy)). At last, colleagues across the country are becoming aware of the potential for positive change offered by the kind of assessment we have championed for all these years.

Unfortunately, if we listen more carefully, we learn that the kind of assessment we believe is beginning to guide improvements in student learning is not what is being discussed by higher education policy makers. It is assessment for accountability, not improvement, that stakeholders outside the academy are proposing.

So those of us in the assessment community are asking each other, “Can assessment for accountability and assessment for improvement coexist? Can the current accountability focus actually strengthen assessment for improvement? Or will an accountability tidal wave roll across the fields, crushing the fragile green sprouts of assessment for improvement that have begun to appear?” In this essay I suggest how assessment for accountability may begin to complement, and even strengthen, assessment for improvement.

## Lessons from Grades K–12

On my campus, I regularly convene a combined group of public school representatives and educators from multiple disciplines who are involved in preparing future teachers. From these colleagues I have heard the following comments: “Some of the most experienced teachers in my school have left the classroom because they feel the public has branded an F on their foreheads that only higher standardized test scores can erase.” “Many who remain in the profession feel pressure to spend much of their time drilling students on the material that will be on the state’s accountability

tests. They say their classrooms have become boring places for both students and teachers as creativity is no longer valued.” “In some schools, the curriculum has been narrowed to focus on the English and math to be tested, so less time is spent on science and social studies, and physical education, art, and music are no longer offered.” “Some students who know they are poor test-takers have given up, exhibit negative behaviors in class, and look forward to dropping out of school at the earliest possible date.”

These educators know the literature and what is needed to create more effective schools—places that delight teachers and students alike, where learning is fun and student achievement is on the rise (Allington 1994). They know that for many students the school day and the school year should be longer, with ninety minutes each for reading, writing, math, and science—interspersed between periods for physical education, art, and music—every day. Personalized instruction should be based on diagnosed needs and learning styles and staff should receive abundant opportunities for professional development that helps them provide such instruction. The tests K–12 teachers value are not the high-stakes state exams, but those that match their teaching objectives and tell them immediately where learning is effective and which students need to improve which skills. But all these improvements cost money, and where will the dollars come from to implement them? To our chagrin, some stakeholders in our community seem more eager to call for spending millions on a second

administration of the statewide tests each year—spring as well as fall testing—than to advocate and find the money for the school improvements we know will enhance student learning. In the meantime, performance of U.S. students on international tests continues a downward spiral.

### **The Press to Make Higher Education More Accountable**

Now we are on the brink of making the press to assess with a test a part of the higher education environment. In September 2006, the Commission on the



Future of Higher Education made a number of recommendations, including the suggestion that “the collection of data from public institutions allowing meaningful interstate comparison of student learning should be encouraged and implemented in all states” (U.S. Department of Education 2006, 24).

Just as we know what works to improve learning in grades K–12, we also have good evidence of what it takes to improve student growth and development in college. Decades of research have demonstrated

that students learn more if they engage actively in learning, spend more time studying, interact frequently with faculty and with student peers concerning intellectual matters, experience high expectations for their persistence and achievement, and encounter and interact with diverse people and ideas (Pascarella and Terenzini 2005). Many students need academic support programs and lots of time with peer and faculty mentors. Most current faculty are not trained as teachers, so extensive faculty development is needed to raise awareness of good practice in enhancing learning.

Just as weighing a pig will not make it fatter, spending millions to test college students is not likely to help them learn more. Equally important, faculty who are just beginning to use assessment aimed at improvement may ask why they should continue to do so if the quality of their institution is going to be judged on the basis of standardized test scores achieved by a small sample of students.

Without question, we will see more emphasis on assessment of learning in college using standardized tests of general intellectual abilities. The commercially available tests of these abilities that I have studied recently are surprisingly lacking in vital information about their reliability and validity—the very characteristics that we expect to give standardized tests the edge over ones faculty develop to test what they are teaching. Moreover, we in colleges and universities have not yet conducted the studies that are needed to test the validity of these exams in our own contexts. We don’t even know how students’ scores on these tests



compare with those of students who have matured over the same four to six years, but who have not gone to college. The impact of what goes on in the classroom—only 15 percent of a student's time—and elsewhere on campus is very hard to tease out of change that is the result of simple maturation, as well as learning that occurs on the job, in the family setting, in the community, and in interactions with peers outside the campus setting. Before we commit millions of dollars to a national testing program, we need to see the results of such validation studies.

The fact that so much is left undone in developing the standardized measures of general intellectual abilities available to us today suggests that there is yet time to develop more meaningful measures of student learning in college. And the impetus provided by the current accountability wave could give such efforts a boost.

### So What Can We Do?

In some states, the mandate to administer one or more standardized tests has already been issued. Elsewhere we have the opportunity to try out some of the instruments being suggested. If scores are going to be used to compare institutions, we have an obligation to learn all we can about these tests. An excellent guide for conducting this investigation is *Standards for Educational and Psychological Testing* (AERA, APA, and NCME 1999). We can ask faculty groups to study the instruments and answer the following questions, among others.

- Do the tests have scales that match our goals for student learning? That is, if we aim to develop good writers

and critical thinkers, do the tests give us scores for these skills?

- Do students in the norm group come from institutions like ours?
- Have valid techniques been used to draw samples of test-takers on these campuses?

- Are convincing studies available that demonstrate test–retest reliability, construct, and content validity?
- Have items been studied to see if they function differently for different groups?
- Are we ready to undertake studies to demonstrate the validity of the tests in our own contexts?
- Will faculty embrace the tests and encourage their students to take them seriously?
- Can students be persuaded to do their best work on the tests?

All of these questions must be answered affirmatively if institutional reputations are to be judged on the basis of scores on standardized tests.

Other options offer advantages over standardized tests for addressing the press for accountability. One possibility is a report card that uses research-based indicators of good practice in higher education. Good

practice in promoting student learning, for instance, can be measured using the National Survey of Student Engagement. We can develop standardized methods for reporting retention and graduation statistics; the portion of financial aid that is need-based; aspects of alumni satisfaction; job

**Most current faculty are not trained  
as teachers, so extensive faculty  
development is needed to raise awareness  
of good practice in enhancing learning.**

placement rates by field; and proportions of students engaged in undergraduate research, service learning, study abroad, and other engaging pedagogies.

Just as some standardized test providers have developed rubrics for scoring students' written work, faculty are capable of developing rubrics that can be applied to virtually every behavior that can be observed. Rubrics make seemingly immeasurable things measurable, and therefore comparable. We can share rubrics across campuses and conduct blind scoring of senior projects, capstone papers, and products of undergraduate research. Students can use rubrics to self-assess their performance in an internship or service-learning setting and compare their perceptions with those of faculty and field supervisors using the same rubrics.

I have argued elsewhere that standardized testing in major fields will pay far richer dividends than standardized tests of

general intellectual abilities (Banta 2007a, 2007b). Many professional fields already have such tests, and disciplinary associations in other fields can develop their own if we must test and compare.

The most authentic assessment will be achieved through electronic portfolios for which students themselves develop the content. On my campus, as at many others, we have developed expected learning outcomes in general education and the major field. Students select graded written, spoken, and artistic works from courses throughout their college careers, as well as photographs and videotapes of speeches, work-related events, and other leadership experiences on and off campus to illustrate their achievement of the expected outcomes. They write reflective essays to demonstrate the connection between portfolio artifacts and the expected outcomes. Again faculty use rubrics to grade students' achievement of each outcome. Just as with written work on standardized tests, faculty can use the same rubric to grade student work on multiple campuses if we must compare institutions.

Finally, we can use the "assessment for accountability is coming!" warning to mobilize colleagues to do their own pioneering work in developing measures of critical thinking, reflective judgment, and deep learning. Instead of throwing in the towel, let's roll up our sleeves and show our critics how creative we can be in developing our own instruments to assess and report on the knowledge, skills, and dispositions our mission statements say we value. And let's argue for the use of multiple measures—questionnaires, interviews, and focus groups as well

as various direct measures of learning—since no single measure is perfectly reliable or valid.

### Conclusion

Should we prepare ourselves to imagine a time when our students' scores on a standardized test become an important component of judging our own effectiveness for promotion, tenure, and raises? Will the college curriculum then be narrowed, as it has been in grades K–12, to focus students' attention on attaining the knowledge and skills defined by the content of these tests? Instead of reading the Great Books, will students have workbooks that help them drill on the concepts on which they will be tested? If so, the United States will fall ever farther behind in the global economy. It is knowledge creation, not knowledge reproduction, that creates competitive advantage. The strength of American higher education has been in the diversity of opportunities we provide for students and faculty with diverse interests and talents. Will standardized testing across all institutions make higher education more homogeneous? And if assessment becomes synonymous with standardized testing, what will happen to assessment undertaken for the purpose of guiding improvement in instruction, curricula, and student services?

As suggested in the examples above, we must work together with our stakeholders to make assessment for improvement and assessment for accountability complement, even strengthen, one another. Ralph Wolff, executive director of the Senior College Commission of the Western Association of

Schools and Colleges, noted recently that accreditors ask each institution "to define its learning outcomes, and to assess the achievement of those outcomes . . . to determine whether improvement is needed. We believe we should keep that locus of responsibility at the institutional level" (quoted in Lederman 2007). If the recommendations of the Commission on the Future of Higher Education and subsequent actions of the U.S. Department of Education related to accountability can provide the impetus for more of the activity Wolff describes, assessment undertaken to guide improvement will be broadened and strengthened. ■

### References

- Allington, R. L. 1994. The schools we have. The schools we need. *The Reading Teacher* 48 (1): 14–29.
- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). 1999. *Standards for educational and psychological testing*. Washington, DC: AERA.
- Banta, T. W. 2007a. A warning on measuring learning outcomes. *Inside Higher Education* (January 26). [www.insidehighered.com/views/2007/01/26/banta](http://www.insidehighered.com/views/2007/01/26/banta).
- \_\_\_\_\_. 2007b. If we must compare. . . . *Assessment Update* 19 (2): 3–4.
- Gibbs, G. and Simpson, C. 2004. *Learning and Teaching in Higher Education*, 1: 3–31.
- Lederman, D. 2007. When is student learning 'good enough'? *Inside Higher Education* (February 23). [www.insidehighered.com/news/2007/02/23/accredit](http://www.insidehighered.com/news/2007/02/23/accredit).
- Pascarella, E. T., and P. T. Terenzini. 2005. *How college affects students. Volume 2: A third decade of research*. San Francisco: Jossey-Bass.
- U.S. Department of Education. 2006. *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: U.S. Department of Education.