



“Going Naked”

By Richard H. Hersh, codirector of the Collegiate Learning Assessment project and former president of William and Hobart Colleges and Trinity College (Hartford, Connecticut)

Education—from preschool through college—is the primary means of improving human capital and is therefore understood to be the single most important factor in the ability of America to compete in the global economy. But there is a growing unease about what now passes for higher education—a vocal concern led not by angry students, as in the sixties, but by parents and business, political, and academic leaders who sense a dangerous hollowing of an increasingly precarious ivory tower.

Virtually every study within and outside the academy acknowledges that that we need to significantly improve our undergraduate colleges, not only to compete globally, but also to enrich an active democracy here at home, a public life marked by liberty, dissent, and robust civic engagement. The critics, in essence, have declared, “The academy has no clothes!”

The Spellings Commission

Joining the critics and jumping into the vacuum created by higher education leaders perceived to be unwilling to take on the necessary reform agenda to substantially improve quality, Secretary Spellings’s Commission on the Future of Higher Education identified accountability as *the* fundamental issue—an issue that can only be resolved through the assessment of value-added learning. The commission’s logic is as follows: (1) undergraduate educational quality is inadequate, given the challenges we face in the twenty-first century; (2) quality improvement requires a more *transparent* accountability; (3) assessment, especially value-added learning assessment, is fundamental to the improvement of quality and accountability. The commission’s report states that

We believe that improved accountability is vital to ensuring the success of all the other reforms we propose. Colleges and universities must become more transparent about cost, price, and student success outcomes, and must willingly share this information with students and families. Student achievement, which is inextricably connected to institutional success, must be measured by institutions on a “value-added” basis that takes into account students’ academic baseline when assessing their results. (U.S. Department of Education 2006, 4)

Assessment as a Force for Accountability and Excellence

The Spellings Commission got it right—quality needs to improve, accountability must become far more transparent, and assessing learning is crucial to both. This is not to say, however, that one single test must be imposed on all institutions or that we know how to measure all that is worth learning. But it is to say that transparent, systematic learning assessment can be a powerful force for improvement and that such assessment is necessary for regaining public trust in the *public* good served by higher education.

There is, of course, the apparent conflict between assessment for improvement and assessment for accountability. I say “apparent” because I do not think this is an either/or situation; assessment for improvement and accountability are inextricably related. The public has every right to expect that it is higher education’s educational and professional duty to systematically assess its

impact on student learning as an essential condition for improvement *and* transparent accountability.

From an improvement perspective, student learning is higher education's *raison d'être*, and we know that appropriate and timely feedback to students and faculty increases student learning and informs institutional change. From an accountability perspective, rigorous, specialized professional training and the status it confers obligates the academy to be transparent in its endeavors, something expected of all professions. Moreover, colleges and universities are subsidized by the public, directly through tax revenues and/or through tax exemptions, and thus do have responsibility for rigorous student and institutional assessment and public accountability. The challenge is to make sure *appropriate* assessments are being used for each function and that the "stakes" attached to each are fair.

In light of the commission's recommendations, the academy is rightly worried about the imposition of federal and state mandates and the resultant loss of institutional autonomy. In terms of learning outcomes, we do not have—and it is not possible to have—one measure that does sufficient justice to the outcomes promised by colleges and universities. And certainly we know better than to defend *U.S. News and World Report* criteria as being worthy of anything other than our contempt as measures of quality—their variables of reputa-

tion, retention and graduation rates, and alumni giving, for example, are predicted mostly by admissions selectivity and endowment per student.

So how might the conversation about learning assessment and institutional accountability be reconciled in the name of institutional and student learning improvement without becoming politicized, as happened in the K–12 sector? The best answer from my perspective is for higher education,



both institutionally and via its accreditation agencies, to take the professional lead on issues of learning assessment and public accountability.

"Going Naked"

There is a useful analog in medicine, summarized in the December 12, 2004, *New Yorker* article "The Bell Curve" by Atul Gawande, which centers on the treatment of cystic fibrosis. The outcomes of various

treatments across the very best hospitals, Gawande notes, are distributed on a bell curve. For example, in 1997, patients at an average center lived to be just over thirty years old; similarly situated patients at the most effective center typically lived to be forty-six. Clearly this is a difference that matters! But what causes that difference? As it turns out, perceived reputation and rankings of hospitals and clinics do not predict excellence in this case. What matters is a caring and demanding institutional culture that also requires rigorous and transparent measurement of outcomes. Shared assessment data in the best clinics informs prescriptive compliance by patients and aids doctors constantly trying to improve treatment.

Making data about their outcomes public leaves centers with no alternative but to do everything possible to help patients survive. Significantly, the ability to *compare* results across similarly situated institutions lays bare (pun intended) the advantage of being candid and the opportunity to be challenged; there is no place to hide. And with it comes the ability to benchmark excellence and establish a culture of continuous improvement. As one doctor said, this is like "going naked."

The Collegiate Learning Assessment Project

The academy is populated with "doctors," and while we are not literally brain surgeons, the quality of life of the mind and



heart is very much in our hands. Assessing outcomes to inform improvement should be just as important to colleges and universities as it is to the medical profession. Yet higher education has neither developed adequate metrics nor demonstrated a willingness to make such results public; instead, it is content to rely on, even while condemning, college guides and reputation rankings. And it is not uncommon to hear faculty and administrators across the country protest that most of what we teach is too complex and cannot be measured, that the diversity of college and university missions precludes one-size-

fits-all assessment, and that the marketplace is the only required arbiter of quality. This implicit “trust us” attitude is now confronted by stakeholders who are questioning quality and no longer willing to accept higher education’s sense of “faith-based” entitlement.

Seven years before the Spellings Commission, the Collegiate Learning Assessment project (CLA) began as an approach to assessing core outcomes espoused by all of higher education—critical thinking, analytical reasoning, problem solving, and writing. (Fig. 1 provides a small sample of questions used in developing our

scoring rubrics.) These outcomes cannot be taught sufficiently in any one course or major but rather are the *collective and cumulative* result of what takes place or does not place over the four to six years of undergraduate education in and out of the classroom.

The CLA is an institutional measure of value-added rather than an assessment of an individual student or course. It has now been used by more than two-hundred institutions and over 80,000 students in cross-sectional and longitudinal studies to signal where an institution stands with regard to its own standards *and to other similar institutions*:

One of the most important features of the CLA program is its policy of reporting results in terms of whether an institution’s students are doing better, worse or about the same as would be expected given the level of their entering competencies. . . . [It] also examines whether the improvement in average student performance between entry and graduation at a school is in line with the gains of comparable students at other colleges. The program is therefore able to inform schools about whether the progress their students are making is consistent with the gains at other institutions. Thus, the CLA program adheres to the principle that post-secondary assessment programs should focus on measuring and contributing to improvement in student learning. (Klein et al. forthcoming)

Figure 1. Sample questions used in developing CLA scoring rubrics.

The CLA measures critical thinking, analytic reasoning, problem solving, and writing skills. These skills include the ability to evaluate and analyze source information, draw conclusions, and present an argument based upon that analysis. Below are some of the many factors that may be included in a task’s scoring guide.

How well does the student

- determine what information is or is not pertinent;
- distinguish between rational claims and emotional ones;
- separate fact from opinion;
- recognize the ways in which evidence might be limited or compromised;
- spot deception and holes in the arguments of others;
- present his/her own analysis of the data or information;
- recognize logical flaws in arguments;
- draw connections between discrete sources of data and information;
- attend to contradictory, inadequate, or ambiguous information;
- construct cogent arguments rooted in data rather than opinion;
- select the strongest set of supporting data;
- avoid overstated conclusions;
- identify holes in the evidence and suggest additional information to collect;
- recognize that a problem may have no clear answer or single solution;
- propose other options and weigh them in the decision;
- consider all stakeholders or affected parties in suggesting a course of action;
- articulate the argument and the context for that argument;
- correctly and precisely use evidence to defend the argument;
- logically and cohesively organize the argument;
- avoid extraneous elements in an argument’s development;
- present evidence in an order that contributes to a persuasive argument?



The purpose of comparison is to stimulate benchmarking and standard-setting discussions that can inform changes in institutional culture, pedagogy, and curriculum to improve student learning. And, as in the medical example above, CLA institutional comparisons result in a bell curve and bear no correlation with rankings such as those reported in *U.S. News and World Report*.

Does It Matter Where One Goes to College?

While the CLA's institutional comparison feature is important, measuring value-added is a necessary but not sufficient condition for improvement; defining standards of excellence must also be part of the improvement process that comparable learning assessment data afford. For example, over the past five years we have found that simply going to college makes a difference—no matter where they go to college, students do show statistically significant gains in the learning of critical thinking, analytical reasoning, problem solving, and writing. Yet virtually all colleges and universities claim that “coming here” versus going elsewhere makes a difference.

Does it matter, then, where one goes to college? In our sample of colleges and universities, we have found that twenty percent of colleges and universities provide substantially greater value-added than other similarly situated schools. We are currently looking at these one-in-five schools to begin to identify what in their cultures, curricula, and pedagogy might explain such significantly better learning gains.

Questioning the CLA

As the CLA has captured greater public attention, a number of fundamental issues have been raised. Trudy Banta, for example, has raised questions about the appropriateness of the value-added approach to learning assessment:

ment focused on the individual student level and did not effectively control for student entry characteristics. This problem is remedied by the CLA, which aggregates student-level data to the institutional level. And while Banta asserts that higher education's mission and student diversity makes valid

Over the past five years we have found that simply going to college makes a difference—no matter where they go to college, students do show statistically significant gains in the learning of critical thinking, analytical reasoning, problem solving, and writing.

For nearly 50 years measurement scholars have warned against pursuing the blind alley of value added assessment. . . . Moreover, we see no virtue in attempting to compare institutions, since by design they are pursuing diverse missions and thus attracting students with different interests, abilities, levels of motivation, and career aspirations. (Banta 2007)

Steve Klein and his colleagues rebut that conclusion by pointing out that prior to the CLA, attempts at value-added assess-

comparisons across institutions difficult, it is precisely for this reason that the CLA assesses core outcomes transcending diverse missions and is designed to permit comparisons between similarly situated students and institutions.

George Kuh has been critical of aggregating individual student scores up to the institution level. Specifically, he suggests that when this is done, “the amount of error in student scores compounds and introduces additional error into the results, which makes meaningful interpretation difficult” (NSSE 2006, 9). Actually, measure-

ment theory predicts just the opposite—results should become much more rather than less reliable when results are aggregated to the school level, especially if there is reasonable variability in scores among campuses, as there is in the CLA. Our further analysis confirms this prediction.

Some believe comparing campuses is invalid because the amount of measurable value-added would be especially limited in

so little in four years that the value-added would be negligible. What she is suggesting, however, is that measures like the CLA cannot detect such learning gains at highly selective schools. Yet no such “ceiling effect” has been found in the CLA national data sample, which includes schools as selective as Penn.

A major concern also has been raised about the potentially brutish purposes for which the CLA or any single assessment might be used. The CLA is not meant to be used as a new ranking tool or as a tool for state or federal agencies to use when deciding how to distribute funding, and this is why CLA data are not made public. If useful learning assessment is the goal, multiple kinds of assessment are required, such as portfolios, comprehensive exams covering both general education and majors, thesis requirements (with and without oral examinations), and capstone courses, although in combination they are rarely utilized in a comprehensive, coherent, or cumulative way within any single institution.

Conclusion

The CLA's purpose is improvement of teaching and learning. The assessment measures core outcomes shared by all institutions and complements more local and

specific assessment techniques with important comparative and value-added data. It communicates that specific higher-order learning is valued, enables institutional improvement by utilizing institutional comparisons to benchmark quality, and emphasizes that such outcomes are accomplished collectively across the entire curriculum.

Higher education has been reticent to measure and share what students are learning, although institutions using the CLA and working in consortia are more willing to take on this transparent task of comparison knowing that others are engaging in the same self-critical analysis. Improvement requires far more substantial and transparent learning assessment, a process that requires going institutionally naked. ■

References

- Banta, T. W. 2007. A warning on measuring learning outcomes. *Inside Higher Education* (January 26). www.insidehighered.com/2007/01/26/banta.
- Gawande, A. 2006. The Bell Curve. *The New Yorker* (December 12): 82–91.
- Klein, S., R. Shavelson, R. Bolus, and R. Benjamin. Forthcoming. The Collegiate Learning Assessment: Facts and fantasies. *Evaluation Review*. www.CAE.org.
- Lifshin, I. 2006. Gutmann: Standard tests a waste at Penn. *The Daily Pennsylvanian*. www.dailypennsylvanian.com/news/2006/08/31/.
- National Survey of Student Engagement. 2006. *Engaged learning: Fostering success of all students*. Bloomington, IN: Indiana University Center for Postsecondary Research.
- 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.



highly selective institutions. President Amy Gutmann, for example, said that if such tests were implemented at the University of Pennsylvania, “students would do superbly when they came in, and superbly when they left, and it would be no measure of what they learned at Penn” (Lifshin 2006). Surely President Gutmann does not mean to suggest that Penn’s students learn