GENERALIZATIONS about longitudinal collegiate assessment are difficult to make, and they are often problematic. For one thing, there is an incredible range of institutional types. Even if we were to stipulate a primary interest in four-year liberal arts institutions, we would find that almost all colleges and universities lay claim—however plausibly—to this designation. And apart from their differing educational structures and missions, institutions also differ in terms of access to resources, geographical location, faculty, and much else. Further, there is no modal student in American higher education. Public discussions frequently imagine student bodies composed of recent high school graduates, mostly native-born, who attend school full time, graduate in roughly four years, live on or near campus, and seek a liberal education. Yet the majority of college students are part-time, and many attend two or even three institutions before gaining a degree. Liberal education is not a personal goal for most of these students.

If we cannot readily generalize about the universe we would like to assess, then how can we even begin to think about the problem of longitudinal assessment? We could take a deep breath and assert that, regardless of the institutional differences, the relevant group in each institution is the students; students are the constant, and student learning is what we need to measure. Or we could throw up our hands.
General Education
and say that, because both students and institutions differ so widely, there is no point in making comparisons that are inevitably artificial or even meaningless. We might still conclude, however, that such objections do not hold for intranstitutional assessment or, perhaps, assessment within peer groups of institutions.

Any attempt at institutional assessment is likely to provoke objections like those raised in reaction to the Spellings Commission, which, in the name of educational accountability, touted mandatory national outcomes assessment through standardized tests as the reform of the future. Although the commission ultimately backed off this extreme and ill-considered recommendation, the specter of a Spellings future has turned many educators into adamant opponents of outcomes assessment per se.

The rationale for these objections needs to be taken seriously. Moreover, there seems to be something inherent in liberal education itself that is hostile to formal assessment. From this point of view, students are works of art in progress; we cannot know whether they have truly been liberally educated until many years after graduation—if ever.

While I strongly believe that we must attempt to assess the effectiveness of liberal education, I recognize that “liberal education” means different things to different people and that different sorts of institutions have quite different formal duties of accountability. I refuse to accept the consumerist assumptions that Chairman Charles Miller of the Spellings Commission and many others apply to higher education, but I also reject the notion that
the higher education community as a whole does not share important values and goals. I am not yet convinced that there is any adequate way to assess across this vast universe, but my mind is open on the subject. I do, however, feel fairly sure that we can assess across comparable groups of institutions and that both the community as a whole and individual institutions would benefit from the effort.

**Institutional self-assessment**

Individual institutions can and should attempt to determine how much and how well their students have learned, but it is not clear how best that could be done. Every institution has a responsibility, both to itself and to the broader educational community, to learn the extent to which it is meeting its educational goals. Therefore, institutional self-accountability is what matters most. As faculty members and administrators, we need to be able to say to ourselves that we are clear about both what our objectives are and whether we are achieving them. From this point of view, the most important use of assessment is in the reevaluation of curriculum, cocurriculum, and anything else relevant to the student learning experience.

Measurable proxies for the student educational experience are clearly needed to provide fixed points of comparison over time, and in order to enable comparisons across institutions, these proxies would have to exist across all forms of liberal education. The more subjective the proxies are—and, thus, more difficult to measure with a high degree of confidence—the less useful they would be. The more objective and easily measured the proxies are, the less subtle and meaningful they may be. The specification of fixed points to measure is the most significant challenge.

Most advocates of liberal education deny that a primarily content-based evaluation of what seniors have learned represents an assessment of the totality of the educational experience. Obviously, the graduating history or biology major should be able to demonstrate considerably greater subject-matter competence than the student just entering the major. Nor do we accept the senior’s comparably broader general knowledge, beyond his or her field of concentration, as a proxy for liberal education. Yet these are the aspects of student learning that are most easily and objectively measured, and they are the ones usually assessed.

The unwillingness to accept these fairly objective proxies is rooted in a commitment to the notion that liberal learning has more to do with the cultivation of qualities of the mind than with the mastery of any quantum of information. What the liberal educator seeks to develop is the capacity to recognize meaningful problems and to identify the information and modes of analysis necessary to address them as well as the instinct to bring these to bear in problem solving. These capacities are much more difficult to measure, although a variety of tests and other assessment exercises have been constructed to try to assess them with some level of precision.

Many right-minded skeptics assert that the very attempt to measure learning outcomes is likely to stifle student creativity, since the forms of assessment create incentives for students to mimic what they assume the assessors seek. This is not a trivial problem, and it is the reason many stress the need for culminating demonstrations of knowledge in the senior year. Having directed senior theses for many years, I am very sympathetic to this view; but senior theses are not enough. Liberal educators should seek more adequate means of both culminating and longitudinal assessment of undergraduate learning. Over the past decade there have been a number of serious attempts to do just that.

**National assessment instruments**

A wide variety of institutional assessment instruments are now available. Some address institution-specific student experiences, others are topic specific, and still others portray national information relating to higher education. The two instruments currently thought to hold the greatest promise are the National Survey of Student Engagement (NSSE) and the Collegiate Learning Assessment (CLA).

NSSE and CLA are, however, very different sorts of assessment instruments. NSSE is designed to obtain from large numbers of colleges and universities, on an annual basis, information about participation in programs and activities provided for student learning.
and personal development. NSSE inquires about institutional actions and behavior, student behavior inside and outside the classroom, and student reactions to their own collegiate experiences. CLA, on the other hand, is an approach to assessing an institution’s contribution to student learning by measuring the outcomes of simulations of complex, ambiguous situations that students may face after graduation. CLA attempts to measure critical thinking, analytical reasoning, written communication, and problem solving through the use of “performance tasks” and “analytical writing tasks.”

Both of these assessment projects result from substantial financial investment and research, and both are serious efforts to create quantifiable measures of the outcomes of liberal education for samples of college students. The proxies each uses for learning outcomes are very different, as are their assessment strategies. Nonetheless, the general project to collect meaningful institutional data represents a good-faith effort to respond to national calls for accountability, and NSSE and CLA provide such data. We are in the early days of this movement, however, and are not yet capable of confidently assessing the assessments.

The more interesting question is, what can and should be done with these data? Some clearly hope that cross-institutional comparisons based on them will provide for a quality ranking of institutions that is more reliable than the faux-scientific rankings currently offered in popular magazines. If one takes a consumerist view of higher education, this makes sense; it would give prospective college students (and their parents) more objective information about the respective educational merits of different colleges and universities. Yet many of the institutional participants in these new assessment exercises will not release the findings, and even if they did, it is not altogether clear that what is being measured would provide satisfactory purchasing signals to prospective students. To the extent that making the findings available can genuinely facilitate informed college selection, however, all responsible educators should favor it. But we are not there yet, and so we will just have to live with U.S. News & World Report for the foreseeable future—doing our best to be noncooperative and noncomplicit in a deeply flawed and poorly motivated process.

More consequentially, we will also have to live with continuing Spellings-like demands for accountability in higher education. This, I suspect, is why the Association of American Colleges and Universities (AAC&U) has collaborated with the Council for Higher Education Accreditation to issue the recent report New Leadership for Student Learning and Accountability (2008). Most striking about this report is what, in contrast to the Spellings Commission report, it does not include—namely, a demand for standardized metrics of outcomes assessment that would permit systematic comparison across institutions of higher education. Instead, the report emphasizes the responsibility of colleges and universities to “develop ambitious, specific, and clearly
stated goals for student learning” appropriate to their “mission, resources, tradition, student body, and community setting” in order to “achieve excellence.”

Each college and university should gather evidence about how well students in various programs are achieving learning goals across the curriculum and about the ability of its graduates to succeed in a challenging and rapidly changing world. The evidence gathered . . . should be used by each institution and its faculty to develop coherent, effective strategies for educational improvement. (2)

I agree with almost all of this—and why not, since the statement so carefully avoids endorsing cross-institutional comparisons? Yet I am concerned by the double bottom line: (1) the assessment of the achievement of “learning goals across the curriculum” and (2) the assessment of “the ability of graduates to succeed.” Are we really responsible for the future success of our graduates in the same way that we are responsible for their learning while in college? Can we really assess future success in ways that do not privilege income and social status as measures of accomplishment? I doubt it, and I firmly oppose such a commitment.

Individual colleges and universities themselves might be able to make creative use of data from institutional assessment instruments in order to determine what works in their own learning programs, and what does not. The one certainty that emerges from the rather tedious ongoing debate about accountability is that each institution has a duty to itself rigorously to evaluate the effectiveness of the student learning it facilitates. And it follows that we owe this duty to our students, to their parents, to the faculty, and to our (public and private) donors.

Public institutions have additional responsibilities to external stakeholders, not least their state legislatures. In response to the special political pressures for accountability on public institutions, the National Association of State Universities and Land-Grant Colleges
and the American Association of State Colleges and Universities have created the Voluntary System of Accountability (VSA), a highly objective and potentially comparable cross-institutional database. Setting aside such a project as unlikely for all of higher education, and acknowledging that the non-compliance of the University of California system does not bode well for the VSA, it seems to me that American educational pluralism dictates that each institution must think through its own standards of accountability. In doing so, an institution should involve all of its many stakeholders.

It seems quite reasonable for colleges and universities to avail themselves of national assessment instruments that at least provide significant benchmarks for evaluating the success of student learning. Depending upon the open availability of data, it may also be possible for institutions to compare themselves to one another. Yet many institutions will not consider these instruments to be sufficient or, perhaps, adequate. But in that case, the burden should fall on that college or university to develop its own internal modes of assessment.

Lessons from history

Conscientious institutions are careful in constructing curricula. We devote time and human resources to designing both general education and field of concentration curricula. We take evaluation of student performance in courses seriously. We are increasingly concerned to support new types of learning experiences, from freshman seminars and undergraduate research to service learning and study abroad. We are experimenting with the potential of information technology and new media to enhance student learning. We are steadily making new fields of knowledge, most of them interdisciplinary, available to our students. These are exciting days in American higher education. But do we really know whether and why we are successful in promoting student learning, if we are?

In a new report, the Education Testing Service advocates a seven-step approach to creating “an evidence-based accountability system for student learning outcomes.” The penultimate step is to determine “What Institutional Changes Need to Be Made to Address Learning Shortfalls and Ensure Continued Success.” The report recommends communicating the results of data analysis, “using the internal decision-making processes of the institution to determine the meaning of the successes and the shortfalls,” and then making educational policy decisions based on this analysis (Millett et al. 2008, 17). The double assumption here is that the data analysis will lead to clear policy decisions and that the institution will have the will and capacity to shift educational policies accordingly. Is it too cynical to suggest that few institutions are fully capable of carrying out step six?

The history of higher education in this regard is not encouraging. We have been worrying about just these sorts of problems for more than a century. In 1933, for example, the American Association of University Professors (AAUP) addressed many of the same questions that vex us today. That year, the AAUP Committee on College and University Teaching issued a report that begins by noting the very problem that remains at the core of our assessment challenge: “The college classroom is the professor's castle. He does not object to the invasion of it by his own colleagues who understand his problems and difficulties, but he reacts against the intrusion of anyone outside that circle who undertakes to scrutinize and appraise his work” (7). The report then picks up on what is today one of the leading criticisms: “Does the college teacher, as such, have a clear conception of what he is trying to do, or indeed of what his institution is seeking to do?” (36). And its assumption of what constitutes good teaching is one that still commands respect. Good teaching “is the kind of teaching which inspires the student to take an active part in the educational process, in other words inspires him to educate himself rather than to expect that someone else will do it for him” (36). The report goes on to assert that “any teacher who gains the desired end, who induces self-education on the part of his students, is an effective teacher no matter what his methods or personal attributes may be” (36). It then asserts that “the main function of the teacher is to stimulate critical thinking, to train his students in methods of reasoning and to carry them back to the sources of the

Is it possible that we will need to assess something more precise than “liberal education” and broader than student performance in courses?
facts, as well as to encourage them to form their own conclusions” (95).

Reading the 1933 AAUP report is an unsettling experience; it might well have been written last year. It notes that many of our difficulties “are connected with the great expansion in college enrollment which has taken place during the past twenty years” (39). And the bottom line is devastatingly reminiscent of our current assessment debates:

There is reason to feel that the general standards of college teaching in the United States have been, on the whole commendably high. Unfortunately, when any one takes issue with this assertion, there is no convincing way of substantiating it. For college teachers have as yet devised no systematic means of having the results of their own work fairly evaluated. They have worked out no objective way of determining whether their work is good or bad.

The college teacher plans his own course and gives his own instruction; at the end of the term he prepares his own examinations, tests his own students, and renders his own verdict upon what he has accomplished. He looks on his handiwork and says that it is good. This self-appraisal of results is not checked by anyone else. (41)

Taking away the current student evaluation system, a very partial reform at best, can we claim much more today in most colleges and universities? The AAUP concluded that “if even a small portion of the ingenuity and persistence which are now being expended on research of the usual type in American colleges and universities could be deflected . . . toward research into the results of their own teaching, the improvement in the general standards of collegiate instruction might be considerable” (68). These words were written eighty-five years ago!

The AAUP report is disturbing for another, more important, reason. Its authors were propagandists for “general education” in the same way that AAC&U currently advocates for “liberal education.” But the history of higher education in the last century suggests that, in fact, “general education” was a weak idea with little unifying or binding power either within or across institutions. If that is correct, then how likely is it that switching from “general” to “liberal” is going to solve the problem? If we are to have meaningful assessment, is it possible that we will need to assess something more precise than “liberal education” and broader than student performance in courses? The courage to assess, and the capacity to assess, is dependent upon the courage of an institution to do something other than to put the pea under a different shell. This is the real issue, and resolving it will take a massive effort of educational reimagination.

Conclusion

The present utility of institutional assessment lies in its potential capacity to enable us to begin to understand what we are doing, and to plan for educational change. If we truly can measure student learning at the levels of sophistication necessary to know whether we are achieving the outcomes we strive for, then assessment data should enable us to begin to make informed judgments about what we are doing wrong, and what we are doing right. And then we can adjust our learning strategies accordingly. Or at least we can have a more meaningful debate about our goals and strategies. Right now, we too often fail to see beyond tactics.

Many thoughtful critics of outcomes assessment warn that a Spellings-like mandate would violate the educational self-determination of institutions and move higher education in the direction of robotic self-imitation. That is indeed a terrible prospect. Secretary Spellings herself has moved away from such an approach, however, and I do not think we ought to behave like Chicken Little. It is possible for us to assess ourselves in ways that will not only help institutions improve student learning, but might also create the norms and benchmarks that will enable us to move ahead nationally in our quest to improve the quality of undergraduate education.

To respond to this article, e-mail liberalled@aacu.org, with the author’s name on the subject line.

REFERENCES

