

THE CHRONICLE OF HIGHER EDUCATION

The Chronicle Review

September 22, 2006

The Legitimacy of Assessment

By *MARGARET A. MILLER*

Many academic leaders have expressed dismay at the work of the Spellings Commission on the Future of Higher Education, especially its calls for "commonly used tests or other assessments" of student learning. They argue that higher education is far too diverse to be measured by standardized tests; that common learning measures will lead inevitably to punitive, costly, and unnecessary federal intervention; and that if assessment is used as a consumer-information tool, it will oversimplify a complex higher-education system and lead to comparisons among unlike institutions.

Those concerns are understandable, and some may even be justified, but they won't make the increasingly insistent demand for accountability disappear. Perhaps the concerns could be quieted, however, if we considered the various forms assessment has taken over the years and the functions it has served — who has assessed whom and why. Such an investigation would probably show that those who pay for higher education have a legitimate interest in knowing about its chief product, learning, and that they can exercise that concern without distorting the educational enterprise.

From its inception, assessment in higher education has had three distinct if often overlapping purposes:

To certify individual student achievement. College-entry and placement exams, rising-junior tests, and licensure and graduate-admissions exams are all examples of common metrics that demonstrate how students perform relative to one another and according to established standards. So, too, are course credits and grades. The Carnegie unit, the original model for the credit hour, began in high schools because colleges found that entering

students were prepared so variously that it was hard to determine on the basis of grades alone who was ready, or not, for college-level work. It, and then the credit hour, marked a certain chunk of academic content that, when combined with a grade, signified what a student knew and to what level of sophistication.

The common currency of credit hours and grades has served higher education well. It has enabled dissimilar institutions to exchange information and allowed students to move across programs and institutions. But it has depended on trust — trust in a common understanding about the content and competencies represented by credits and grades. Over time that trust has eroded within higher education, as evidenced by tussles over transfer credits, worry about grade inflation, and graduate schools' emphasis on entrance-examination scores in the admissions process. It has also eroded between higher education and the professions for which it prepares students. So those professions in which competence is a life-and-death matter, like the health professions, require that practitioners not only complete college programs but also pass separate licensing exams.

Colleges have no trouble with standardized assessments of individual students, but that comfort may be due, at least to some degree, to the fact that the locus of responsibility for learning is the student. He or she is the one being held accountable, the one for whom the stakes are high.

To improve programs. Program and institutional assessment emerged in the mid-1980s, prompted, sometimes vigorously, by the federal and state governments. (The feds acted through accreditors by requiring them to pay attention to outcomes.) In instituting campus-based assessment, policy makers accepted the academy's premise that every campus — indeed, every program — had different learning goals and standards, and hence had to assess success by its own measures. The presumption was that that would give faculty members the necessary information to make their programs better, and students and others would know about the learning that those colleges were generating.

The first premise turned out to be correct, at least on campuses

where the task was taken seriously. But most faculty members did not take it seriously, probably because the locus of responsibility had shifted beyond the students themselves. If the students in a program consistently failed to, say, write well, the professors in charge of the program had to assume some responsibility for that outcome.

To hold higher education accountable to its constituencies. The second premise, that letting a thousand assessment flowers bloom would satisfy accountability demands, turned out to be incorrect. After a while, those who had crafted the original assessment legislation or regulations began to notice that they couldn't get an answer to the famous question from the New York City mayor Ed Koch: "How're we doin'?" The answer to that question, after all, depends on the answer to another: "Compared to what?" Is an 80-percent pass rate on a home-grown exam good or bad news? Who knew? Also, who knew if the goals of the program made sense or how high its standards were?

Such questions about higher education's effectiveness at carrying out its core mission became more urgent as enrollments surged, demands on government coffers multiplied, taxpayer resistance grew, and the diversity of both students and institutions increased. So in the early 1990s, the desire for intelligible results led policy makers in some states to return to standardized testing as a way to compare performance across institutions.

But that testing differs from the kind focused on the performance of individual students. The point is not to determine whether a person knows or is fit to do something, but instead to find out how well an institution is doing at developing the knowledge and skills of its students. How much does the college add to students' intellectual resources (the value-added question), and what level of understanding do its graduates, on average, finally attain? The responsibility is not the individual learner's but the institution's, and it's not necessary to test all students but just a representative sample.

In 2000 the National Center for Public Policy and Higher Education published the first national report card on higher

education, "Measuring Up," which graded states on how well they prepared students for college, made higher education both accessible and affordable, got students through college, and reaped the benefits of a college-educated citizenry. But all states were given an incomplete in the sixth category, learning, since the existing national data on student learning was insufficient to create a good set of indicators.

In response, the Pew Charitable Trusts sponsored a five-state pilot project, the National Forum on College-Level Learning, which I directed, to test a set of indicators of advanced learning. The indicators measured the literacy level of the states' college-educated residents, the readiness for work and graduate education of the graduates of the states' institutions, and the exiting students' general intellectual skills. To develop comparable data across states, the process involved standardized testing: the National Assessment of Adult Literacy for the first category, graduate-admissions and licensing exams for the second, and WorkKeys and the Collegiate Learning Assessment (administered to a random sample of graduating students on a representative set of campuses) for the third.

But the locus of responsibility had shifted once again. The focus was not on the performance of individual students or institutions but on that of the policy makers and the colleges and universities, as a group, in each state. Does the state have economic-development policies that draw highly skilled workers? Does K-12 policy ensure that students are prepared to learn when they go to college? To what degree do the institutions in the state develop their students' intellectual capabilities?

The results helped the states in the pilot project understand what kind of educational capital they had at their disposal to further their civic and economic objectives. The capacity to compare their performance to that of other states helped them determine whether they were doing a stellar, lackluster, or poor job in the national context. The information enabled them to see what their institutions collectively were doing well and not so well — for instance, in some states institutions were doing a good job of preparing students for graduate school but not so well for the

licensed professions, whereas in others the situation was reversed. And the data pinpointed the statewide problems that required further attention, like the disparities among different groups on learning measures and the wide range across states in teacher preparation. In short, it answered the question of how well each state's college-educated citizens were prepared for life and work in the 21st century.

That is what has driven the work of the Spellings commission. Its members want to know if the government's massive investment in higher education is paying off. They want to know if the nation is poised to take on the increasingly fierce economic competition it faces, and whether its citizens are prepared to navigate the complexities of modern life and make good collective decisions. Those are questions we want policy makers to ask.

And we should be ready to do our part in answering them. In recognition of that, the Educational Testing Service and the National Association of State Universities and Land Grant Colleges have proposed that colleges "agree on a meaningful but manageable small set of accountability measures" to "help stakeholders assess undergraduate student learning," and that accreditors take the lead in "integrating a national [NB: not federal] system of assessing student learning into their ongoing reviews of institutions." Both organizations believe that such sample-based testing can be done at a reasonable cost without making inappropriate comparisons among unlike institutions or oversimplifying the remarkably diverse American higher-education system.

That doesn't mean that programs and institutions should stop collecting information about how students are doing — information that is essential for improving performance. But the more-general questions put what we learn about students, programs, and institutions in a larger context. However diverse our colleges are, we are engaged in a collective enterprise and should be prepared to assess our contribution to it.

Margaret A. Miller is director of the Center for the Study of Higher Education at the University of Virginia and executive editor of

Change magazine.

<http://chronicle.com> Section: The Chronicle Review Volume 53,
Issue 5, Page B24