

## WHAT DOES COLLEGE TEACH?

*It's time to put an end to "faith-based" acceptance of higher education's quality*

BY RICHARD H. HERSH

"What makes your college worth \$35,000 a year?" It's a hard question for a college president to answer, especially because it's usually raised at gatherings for prospective students and their anxious, checkbook-conscious parents. But it also provides an opportunity to cast one's school in a favorable light—to wax eloquent about admissions selectivity, high graduation rates, small classes, and alumni satisfaction.

The harder question, though, comes when someone interrupts this smooth litany: "But what evidence is there that kids *learn* more at your school?" And as I fumble for a response, the parent presses on: "Are you saying that quality is really mostly a matter of faith?"

The only answer is a regretful yes. Estimates of college quality are essentially "faith-based," insofar as we have little direct evidence of how any given school contributes to students' learning.

This flies in the face of what most people believe about college, and understandably so. After all, if we don't know what makes a school good or bad, then the anxiety-driven college-application process is a terrible waste, the *U.S. News & World Report* rankings are a sham, and all the money lavished on vast library holdings, expensive computer labs, wireless classrooms, and famous faculty members is going for naught. And what about SAT scores, graduation rates, class sizes, faculty salaries, and alumni giving? Surely, a college-obsessed parent might object, such variables make *some* difference.

Perhaps they do—but if so, we haven't found a way to measure it. In *How College Affects Students*, a landmark review of thirty years of research on college learning, Ernest Pascarella and Patrick Terenzini found that simply going to college, any college, makes a major difference in a young person's psychological development: students come away with improved cognitive skills, greater verbal and quantitative competence, and different political, social, and religious attitudes and values. But although the researchers found wide variations in learning within each college or university, they were unable to uncover significant differences between colleges once the quality of the entering students was taken into account.

So it's not just a perverse status-consciousness that makes higher education the only industry in which competitors are rated on the caliber of their customers rather than on their product—or that drives *U.S. News & World Report* to rank colleges on how well they recruit and graduate already successful high schoolers. It's that we have no other discriminating way to measure collegiate quality.

It's possible that this situation reflects a real absence of variation—that there really isn't much difference between, say, an Ivy League education and four years at a middling private or state school. According to this explanation, faculty members across the country tend to graduate from a relatively small number of doctoral programs, use comparable textbooks, construct similar curricula, hold fairly low expectations for student achievement (particularly in an age of grade inflation), and labor under a system that rewards research over teaching. In this homogenized landscape the quality of entering students is the only thing that matters: "Diamonds in, diamonds out; garbage in, garbage out."

A second, more persuasive explanation, however, holds that current assessment measures simply can't pick up the differences in learning from one campus to another. And robust measurements don't exist in part because colleges don't want them—because developing and testing them would be expensive; because faculty members would disagree on what to measure; and because they're wary of anything that calls into question the long-running perception of American higher education as "world class."

But in an era when the importance of a college diploma is increasing while public support for universities is diminishing, such assessment is desperately needed. The real question is who will control it. Legislators are prepared to force the issue: Congress raised the question of quality during its recent hearings on the reauthorization of the Higher Education Act; all regional accrediting agencies and more than forty states now require evidence of student learning from their colleges and universities; and pressure is rising to extend a No Child Left Behind-style testing regime to higher education.

To date academe has offered little in response, apart from resistance in the name of intellectual freedom and faculty autonomy. These are legitimate professional prerogatives; but unless the academy is willing to assess learning in more rigorous ways, the cry for enforced accountability will become louder, and government intervention will become more likely.

Current measures of college quality fall into four major categories, outlined a few years ago by my colleague Marc Chun: actuarial data, expert ratings, student/alumni surveys, and the direct assessment of student performance. While each of these has its uses, none is anywhere close to being a legitimate measure of how much students learn over their college careers. Like the drunk who looks for his keys under the streetlamp because the light is better there, Chun has argued, colleges rely on these measures because they are inexpensive and readily available, not because they actually tell us much.

Actuarial data and expert ratings are familiar to anyone who has spent an afternoon leafing through the *U.S. News* rankings. The former consist of quantifiable information such as graduation rates; data on racial diversity, admissions selectivity, and research funding; student-teacher ratios; and SAT and ACT scores. These statistics are easy to gather, and have long been assumed to reflect institutional quality. But there is little evidence that the attributes they measure have a decisive impact on student learning.

Equally easy to compile are surveys of institutional quality, in which faculty members and administrators across the country are asked to rate their competitors, typically on a five-point scale. These surveys are interesting if not taken too seriously, but the participants may not know enough about other institutions to make such judgments, and the variables they find most noteworthy may not be the ones that are actually important.

More promising are the surveys that ask students and recent graduates to assess their experiences. One of the most prominent and useful is the National Survey of Student Engagement (NSSE, pronounced *Nessy*), launched in 1999 and currently administered by 573 colleges and universities (see "[What Makes a College Good?](#)" by Nicholas Confessore, November 2003 *Atlantic*). NSSE asks students to rate their educational experience by reporting, for instance, on the quantity and quality of contact with the faculty and on how much homework they receive. Given that previous research shows a strong correlation between such educational "engagement" and learning, NSSE scores may be a better measurement of how well schools teach than many of the statistics that find their way into college rankings. But correlation isn't causation, and surveys like this one offer at best an indirect assessment of educational quality. Their findings rarely touch on either what was learned or, more important, what ought to have been learned.

Finally there is the direct assessment of student learning that takes place constantly on college campuses, usually symbolized by grades and grade point averages. For our purposes these are nearly useless as indicators of overall educational quality—and not only because grade inflation has rendered GPAs so suspect that some corporate recruiters ask interviewees for their SAT scores instead. Grades are a matter of individual judgment, which varies wildly from class to class and school to school; they tend to reduce learning to what can be scored in short-answer form; and an A on a final exam or a term paper tells us nothing about how well a student will retain the knowledge and tools gained in coursework or apply them in novel situations.

Nor can grades capture the cumulative effects of taking dozens of courses over a single four-year stretch. Sometimes the whole of a college education is less than the sum of its parts. Sometimes it's far greater. And in neither case does a student's GPA, whether 2.2 or 4.0, really tell us how much he or she has learned.

Just as challenging as the absence of reliable measures is the resistance to developing them within the academy itself. Even as the initiative for comprehensive educational assessment builds outside the university, academics continue to shy away from the issue.

Their reasons are various. What is worth learning cannot be measured, some say, or becomes evident only long after the undergraduate years are over. Others claim that any kind of assessment is a threat to academic freedom and a power grab by administrators and legislators seeking to micro-manage instruction, impose a partisan agenda, or curry favor with voters by claiming to have brought "accountability" to higher education. And the academy has observed with alarm the problems states are having with K-12 assessment.

But perhaps myopia is operating here as well. No one doubts that professors care deeply about whether students learn what is taught in *their* courses. One suspects, however, that academic turf wars have a lot to do with why *cumulative* learning is rarely measured. Academics have trouble agreeing with their colleagues in the same field on what students ought to be taught, let alone with colleagues in other disciplines. As a result, to borrow from G. K. Chesterton, measuring cumulative learning hasn't been tried and found wanting; it has been found difficult and left untried.

Or again, the skeptics are right that what often passes for assessment, both in higher education and in grades K-12, too often trivializes learning. But that tells us only what is, not what can or ought to be. And it's ironic that academics so disdain the pursuit of data on the subject, given that the academy's culture of evidence is the enviable foundation for the greatest research universities in the world.

Finally, faculty members are perfectly correct to point out that a well-conceived assessment program would take considerable time, energy, and money. They are also correct that it would require a difficult rebalancing of research and teaching priorities. But perhaps such a rebalancing, with a renewed focus on undergraduate assessment and an end to the suffocating power of the research ethic, is exactly what universities need.

If assessment is to take hold in the university, however, it's crucial that the impetus for reform come from within. It's a terrible idea to have people outside the academy—whether consultants, politicians, or businessmen—telling professors how, what, and what not to teach.

Nonetheless, there are outside examples worth considering. For instance, a hardheaded assessment ethic makes a big difference in medicine, where survival rates for conditions such as colon cancer and cystic fibrosis can vary dramatically from hospital to hospital. The most successful hospitals are those that measure outcomes and give patients access to the information—which is exactly the model that higher education ought to follow.

A number of promising approaches are already moving academic "doctors" in this direction. At Carleton College, in Northfield, Minnesota, for example, faculty panels assess students' writing using samples from different courses. The portfolios are turned in at the end of every student's sophomore year—an ideal point for remediation. Preliminary reports suggest that the system has helped clarify the school's expectations and standards for faculty and students, and has improved students' writing. This kind of evaluation is of course time-consuming, and Carleton has the advantage of being a small school (1,800 students) with a low student-teacher ratio. Portfolio assessment is not limited to small colleges, however. For example, Washington State University, with 18,700 students on its Pullman campus, has developed a similar system that also incorporates a faculty-graded two-hour writing exam.

Another innovative way to assess overall student performance can be found at Alverno College, in Milwaukee, Wisconsin. Alverno's faculty has created an integrated liberal arts and professional-studies curriculum focused on abilities ranging from analysis and problem solving to effective citizenship and engagement with the arts. Students do not receive grades in the usual sense; instead entering students learn to assess their own course work, and also receive feedback from faculty members and from assessors in the local business and professional

communities. Students keep their assignments and feedback, along with their self-evaluations, in electronic portfolios, to track their progress over time, and a faculty council monitors the quality of the assessment across majors.

Also promising is the movement toward "value-added" assessment, which attempts to measure what a particular college or university contributes to its students' knowledge and capabilities during their four or five years.

One interesting project recently launched by the Education Trust, a nonprofit group that advocates for education reform, compares the graduation rates of close to 1,500 colleges and universities. The comparison assumes that a school is adding considerable value if it graduates more of its students than would be expected given their high school records and socioeconomic background, and adding little if it admits a bumper crop of high-achieving kids and then graduates them at a below-average clip. It's not a perfect metric: accumulating credit hours and earning passing grades isn't equivalent to actual learning, and a school can easily improve its graduation rate by grading more generously. But comparing graduation rates with actual learning measures should prove interesting.

Two other value-added initiatives may soon be able to provide such measures. In the fall of 2006 the Center of Inquiry in the Liberal Arts, at Wabash College, plans to initiate a longitudinal study of 5,000 students at sixteen institutions. Researchers will use existing standardized tests along with in-depth interviews to examine the students' development of problem-solving abilities and their inclination to learn, cultural sensitivity, leadership, and moral character. They hope their findings will help reveal which teaching conditions are most conducive to learning and whether initiatives such as study abroad, service learning, and diversity programs are effective.

At the same time, the Collegiate Learning Assessment Project (of which I am a co-director) has created two types of tests that evaluate students' ability to articulate complex ideas, examine claims and evidence, support ideas with relevant reasons and examples, sustain a coherent discussion, and use standard written English.

The first, called a "performance task," provides students with a mini-library of diverse documents, such as letters, memos, summaries of research reports, newspaper articles, photographs, diagrams, tables, charts, and interview notes or transcripts. Students are asked to identify the strengths and limitations of alternative hypotheses, points of view, and courses of action. An example:

A catfish with a grotesque mutation is caught in Paradise Lake, the source for the local water supply. Local press coverage has the village buzzing. Mayor Carp has asked you and some others in your community to serve on a panel to investigate this matter. You are provided with the following documents:

- a newspaper article that contains a picture and description of the fish and the opinion of a recognized expert as to its source
- an editorial by an environmental activist
- a radio interview with a biologist who teaches at a nearby college
- a State report with the results of water testing and other investigations regarding Paradise Lake
- a map of the area
- an article about similar fish "catches" from ECO, a journal focusing on issues of clean air and safe water

Using these data sources, please prepare a memo to the chair of the panel regarding (1) your analysis of the strength and limitations of various explanations for finding such a fish in Paradise Lake and (2) your recommendations regarding what should now be done about this situation and your reasons for these recommendations

All the tasks demand similar skills: students must weigh, organize, and synthesize evidence from different sources; distinguish rational from emotional arguments and fact from opinion; analyze data; deal with inadequate, ambiguous, or conflicting information; spot deception and holes in the arguments of others; recognize what information is or is not relevant to the task at hand; and identify additional information that might help to resolve issues.

Each performance task is set in the context of a broad academic field, such as science and engineering, business, the social sciences, or the arts and humanities. But a student should be able to respond adequately to a task without having specialized knowledge in the particular field. Indeed, students do as well on CLA performance tasks drawn from other fields as they do on those related to their own majors.

The second test, of analytic writing, requires two essays: a forty-five-minute "Make an argument," in which students either support or reject a position on some issue; and a thirty-minute "Break an argument," in which they consider the validity of someone else's reasoning.

"Make an argument" asks students to react to an opinion—for example, "There is no such thing as 'truth' in the media. The one true thing about information media is that they exist only to entertain." They can address the issue from any perspective they wish, as long as they provide support for their views. "Break an argument" might present students with a passage like this:

A well-respected professional journal whose readership includes elementary school principals recently published the results of a two-year study on childhood obesity. (Obese individuals are usually considered to be those who are 20 percent above their recommended weight for height and age.) This study sampled 50 schoolchildren, ages 5-11, from Smith Elementary School. A fast food restaurant opened near the school just before the study began. After two years, students who remained in the sample group were more likely to be overweight relative to the national average. Based on this study, the principal of Jones Elementary School decided to confront her school's obesity problem by opposing any fast food restaurant openings near her school.

Rather than agree or disagree with the position, students must discuss how well reasoned they find the argument by considering the soundness of its logic.

Approximately 19,000 students from 134 colleges and universities participated in the CLA through May of 2005. In this academic year an additional 100 institutions and more than 30,000 students are participating. They come from a national sample of colleges, universities, and community colleges—private and public, large and small, more selective and less selective. Results are being aggregated at the institutional level, to permit comparisons across institutions and to determine how well individual schools are doing.

The findings to date are illuminating. After controlling for admissions selectivity, the CLA shows that which school a student attends *does* make a difference. Assuming that these initial results hold up, and that learning differences can be attributed to variations in campus culture, curricula, and pedagogy, the next step will be to create case studies of the schools that achieve the best results, and conduct follow-up studies of those that make changes to improve learning outcomes. The ultimate goal is not to create a new college ranking (though some will be tempted to use our findings for that end) but to let colleges and universities share their successes the way doctors and hospitals do.

Some cautionary notes are warranted. Value-added assessment tells us only how schools are doing in relation to their competitors, not what absolute standards of excellence they should be setting. Nor should it be allowed to crowd out other measures—particularly affordability and equity—that bear on how we judge a school's quality. Finally, there are many things we cannot yet measure accurately—and some aspects of "quality" will always remain elusive.

Nonetheless, value-added assessment offers an excellent place to start, and a chance for higher education to demonstrate that "faith-based" answers about quality are no longer acceptable. This country has always looked to higher education to take the lead in innovation, and to define, seek, and demand excellence from its students. Today's academy should be satisfied with nothing less.

The URL for this page is <http://www.theatlantic.com/doc/200511/measuring-college-quality>.