February 2008

The Rain Man Cometh - Again

Robert M. Zemsky
University of Pennsylvania, ZEMSKY@thelearningalliance.info

Used with permission. Reprinted from *Academy of Management Perspectives*, Volume 22, Issue 1, February 2008, pages 5-14.
Publisher URL: http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=31217507&site=ehost-live

The Rain Man Cometh - Again

Abstract
For colleges and universities October has traditionally been a tough month — growing darkness, impending rain and cold, the creeping realization that the football team won't win that many games. Not so long ago, October was also the month of reckoning for higher education: On October 16, 1983, *U.S. News and World Report* published its first rankings of institutions. Under the soft-sell headline "America's Best Colleges" and in the breathless prose of a Sunday supplement, *U.S. News* offered up a collegiate roster of who was in, who was out, who was hot, and who was not.

Comments
Used with permission. Reprinted from *Academy of Management Perspectives*, Volume 22, Issue 1, February 2008, pages 5-14.
Publisher URL: http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=31217507&site=ehost-live

EXCHANGE

The Rain Man Cometh—Again

by Robert Zemsky

For colleges and universities October has traditionally been a tough month—growing darkness, impending rain and cold, the creeping realization that the football team won't win that many games. Not so long ago, October was also the month of reckoning for higher education: On October 16, 1983, *U.S. News and World Report* published its first rankings of institutions. Under the soft-sell headline "America's Best Colleges" and in the breathless prose of a Sunday supplement, *U.S. News* offered up a collegiate roster of who was in, who was out, who was hot, and who was not.

A True Phenom

In 2007 the *U.S. News* rankings celebrated their 25th anniversary, having become an American icon. They've also spawned a veritable tribe of rankings telling Americans about the best law schools and graduate schools and medical schools and more.

Not surprisingly, the rankings have become the subject of extended analyses, most of which are designed either to discredit them or to discover what they truly measure. As every right-thinking academic knows, the rankings cannot possibly measure what they propose to measure—that immeasurable quantity: academic quality. In the winter of 1990 Bill Massy and I joined the growing discourse on the rankings with our article "The Rain Man Cometh" (Zemsky & Massy, 1990). Then, as now, there were three basic ways to attack the rankings.

The first, and easiest, was to point out that the rankings were silly: arithmetically precise but largely without meaning. *U.S. News* had started with a clever idea: Ask college and university presidents to list what they thought were the best institutions for an undergraduate education. Even if the results of what became known as the "beauty contest" reflected a bias in the magazine's choice of presidents, most knowledgeable observers were intrigued by the outcome. Despite some notable as well as curious omissions, few readers doubted that those at the top of the presidents' list belonged there. It may have been gossip, but it was good gossip that sold a lot of magazines.

The problem was that the losers in these early polls wouldn't accept the results. With unexpected force—after all, *U.S. News and World Report* was not a particularly important magazine—those slighted by the poll argued that the rankings were too simplistic, too much a product of fading reputations and old-school networks. *U.S. News*' response was to get scientific. Starting in 1989 the annual rankings issue included, in addition to the results of the "beauty contest," a variety of statistics the editors presented as objective measures of institutional quality. Most of the measures reflected educational inputs rather than outputs. How selective was the institution? What was the average SAT/ACT of the freshman class? What was the student/faculty ratio? How much money did the institution spend on undergraduate education? Not surprisingly, many within and a few outside of higher education asked what had happened to that old-fashioned notion that quality in education meant good teaching, engaged faculty, and industrious students.

There were also problems with the statistics themselves. Some numbers, it turned out, counted for more than others, though the reader was never

Robert Zemsky *(zemsky@thelearningalliance.info) is a Professor at the University of Pennsylvania's Graduate School of Education.*

told exactly how much more or why. Some institutional resources were counted twice, first as revenue and then as expense. Some revenues did not count at all in the *U.S. News* formula—tuition, for example, and federal research funds.

The titles of categories did not mean what they seemed. The category "faculty reputation" was not about faculty standing but rather faculty-student ratios. Initially, at least, transcription and data errors abounded. Vassar College complained to the magazine that a simple data-entry error cost it four places in the rankings. Obvious errors originating in an institution's federal Integrated Postsecondary Education Data System (IPEDS) report, which *U.S. News* used to construct the rankings, were allowed to stand without further checking. What's more, in the beginning there was apparently little or no checking of the admissions data submitted by institutions, though in recent years *U.S. News* has both checked and deducted penalty points when its algorithms detected an implausible submission.

Like the calculations of the autistic savant Dustin Hoffman played in *Rain Man*, the movie from which Bill and I borrowed the title for our critique, the rankings were invested with arithmetic precision, but the numbers themselves were largely devoid of meaning.

The second way to attack the rankings was to declare them wrongheaded, having confused ends with means. The editors responsible for the rankings issues seemingly did not understand what their statistical staff and consultants were doing—or why. In our critique, Bill and I focused on the essay that ran as the forward to the 1989 issue. It was a sensible, at times eloquent statement of the need for a new sense of financial limits in higher education. To make our point we quoted the editors' declaration that "In the end bidding for limited resources will intensify, and colleges will have to make ever more Draconian decisions about what is truly essential to their mission" (Zemsky & Massy, 1990, p. 22).

It would have been difficult, but not impossible, for the rankings to reflect this new sense of limits, rewarding those institutions that had demonstrated their quality and efficiency by in fact doing more with less. Instead, the numbers embedded in the ranking methodology rewarded profligacy.

The financial information used in the analysis did not reflect efficient expenditures, but rather gross revenues. To make our point we used the case of Carleton College in Minnesota. The college had participated in our study of the cost of an undergraduate education, and we had come away with an appreciation of Carleton's efficiency, the work ethic of its faculty, and the structure and leanness of its curriculum. We were not alone in our admiration.

In the "beauty contest" section of the rankings issue, Carleton placed sixth among national liberal arts colleges. In overall ranking, however, the college was only 14th. Why? Because the college ranked 41st in faculty quality (a slightly higher faculty-student ratio than most of its peers) and 26th in financial resources (a slightly lower endowment income and almost no public appropriation). How, we asked, could "a college judged to have a 'statistically' mediocre faculty in fact have a national reputation that puts it among the top 10 liberal arts colleges year after year?" The answer was simple. The measures were nonsensical.

The third, and by now the most damning, attack on the rankings held that the exercise itself was inherently dangerous. The rankings have become the scorecard of an admissions arms race that has encouraged a spiraling competition for students and faculty that, over the last 25 years, has dramatically escalated the cost of an undergraduate education. I sat in horror one afternoon at a meeting of the board of trustees of a nationally ranked liberal arts college as its members berated the president because the college had once again failed to crack the top 25 national liberal arts colleges—a category that makes up half of the "nifty fifty" that are featured each year in the *U.S. News* annual rankings issue. On this day the president had had enough. Forgetting who and where he was, he snapped. In a voice barely distinguished from a snarl he responded, "You let me run a hundred million-dollar deficit and I will guarantee you a top-10 finish!" He was wrong, of course. It probably would have taken little more than an annual increase of expenditures in the $10 million to $15 million range.

There was, we pointed out in our original "The Rain Man Cometh" article, a secondary danger as well. More than one member of the board of trustees, we suggested, would likely be tempted to take charge of the problem him or herself. Drawing on their business experiences, particularly their successes at having their own companies included among "America's Best Corporations," some trustees and regents would likely argue for an immediate investment in creative public relations. The result would be more business for high-priced consultants who promised to polish the institution's image.

And that is pretty much what has happened. Getting an institution's annual submission to *U.S. News* ready has become "presidential business." With so much at stake the president wants to make sure the numbers submitted allow the institution to put its best foot forward. Presidents are ever watchful for the opportunity to tout the excellence of their institutions—and if those opportunities coincide with the distribution of the institutional rating forms, so much the better. Talk about the rankings and the importance of becoming a top-ranked institution now permeates an ever-increasing number of strategic plans along with promises to be increasingly student-centered and civically engaged.

Colin Diver, former dean of the law school at the University of Pennsylvania and now the president of Reed College, summed up this aspect of the problem in 2005 in *The Atlantic*. Diver, an educational insider who by his own testimony now feels liberated from the rankings madness, restated Reed's policy of not cooperating with the rankings and in the process summed up what many believe but only few act upon:

> *A somewhat more important consequence of Reed's rebellious stance is the freedom from temptation to game the ratings formula (or, assuming that we would resist that temptation, from the nagging suspicion that we were competing in a rigged competition). Since the mid-1990s numerous stories in the popular press have documented how various schools distort their standard operating procedures, creatively interpret survey instructions, or boldly misreport information in order to raise their rankings. . . . I was struck . . . in reading a recent* New York Times *article, by how the art of gaming has evolved in my former world of legal education, where ranking pressure is particularly intense. The* Times *reported that some law schools inflate their graduate-employment rates by hiring unemployed graduates for "short-term legal research positions." Some law schools have found that they can raise their "student selectivity" (based in part on LSAT scores and GPAs for entering students) by admitting fewer full-time first-year students and more part-time and transfer students (two categories for which data do not have to be reported). At least one creative law school reportedly inflated its "expenditures per student" by using an imputed "fair market value," rather than the actual rate, to calculate the cost of computerized research services (provided by LexisNexis and Westlaw). The "fair market value" (which a law firm would have paid) differed from what the law school actually paid (at the providers' educational rate) by a factor of eighty!*

He then drives the point home by quoting a presidential colleague who, when asked why he didn't refuse to participate in the *U.S. News* process, responded, "The rankings are merely intolerable; unilateral disarmament is suicide" (Diver, 2005).

Diver's critique reflects the often private anger of presidents who feel helpless to do anything but play the game. More recently, however, a number of presidents, most of whom head schools that are members of the Annapolis Group of 34 private liberal arts colleges, have started taking on the rankings on educational grounds. Often their spokesman is Lloyd Thacker, a former admissions counselor who has seen the light. What alarms Thacker and those for whom he speaks is that the rankings have so corrupted the college choice process that students, particularly the best, brightest, and most affluent, make decisions that are all but devoid of educational content and value. Students don't ask the right questions and don't draw the right conclusions, having bought the implication of the rankings that somewhere out there is the very best college for them. The reality, as most presidents know—though in their public pronouncements they often deny—is that most colleges competing for the same students are pretty much the same—or as Larry Litten, an institutional researcher for some of the U.S.'s best and priciest colleges and now a consultant to Thacker's Educational Conservancy, told a meeting of anti-ranking activists called by Thacker:

> *The message that needs to be delivered by colleges (both by what they say and how they market, recruit, and admit) is that most institutions seek similar goals, at least within broad groupings of institutions with similar resources and programs, and that each institution is sufficiently complex so that a given student can find a situation that is comfortably challenging and rewarding. As long as individual institutions are implying otherwise in their own marketing and competitive behavior, however, students will turn to the mechanism that most efficiently and plausibly differentiates and ranks institutions (Educational Conservancy, 2007).*

The ultimate danger inherent in any ranking scheme is that by enshrining the language of competition and uniqueness the rankings themselves corrupt both student and college. The former arrives for his or her freshman year with the wrong expectations. The latter engages in a process that devalues education. Nobody wins in the rankings game.

The Flip Side

Not everyone and certainly not every president of an American college or university would agree. Competition is not necessarily bad, and, in some cases, it has actually proved beneficial. When Judith Rodin became president of the University of Pennsylvania in 1993 she found an institution that had grown progressively stronger without ever quite losing its inferiority complex. Indeed, that internal sense of not being quite as good as its competition had already led Penn in the 1920s and then again in the 1950s to commission an educational survey designed to help the university understand its place within American higher education. Each exercise yielded roughly the same result: Penn, when all its strengths and weaknesses were taken into account, ranked about 15th among the nation's principal universities. Not coincidentally, 15th is about where Penn consistently placed in the *U.S. News* rankings prior to Rodin's arrival.

Rodin wanted Penn's strengths to be recognized and celebrated both within and without the university. Her message to her deans, including Colin Diver, was simple. She expected them to secure for their schools a top-10 ranking. She did not care which rankings were used as long as they were comprehensive and garnered wide attention. She did not insist that each school use the *U.S. News* ranking, but absent another recommendation, that is the ranking she would pay attention to.

The results were spectacular. Penn rose in the overall rankings to achieve top-10 and occasionally top-five status among national research universities. Its professional schools—Wharton, Medicine, Law, even Nursing and Education—solidified their positions as top-10 institutions. Were the rankings directly responsible for these schools' or the university's achievements writ large? Of course not—but the presence of the rankings and the fact that the president took them seriously certainly helped focus people's attention.

Northeastern University presents an even more dramatic example of how paying attention to the rankings was essential for institutions seeking to reposition themselves in an increasingly competitive market for students, faculty, and research dollars. In the 1980s Northeastern had floundered. Once a blue-collar university whose co-op programs had particular appeal for Boston's working-class students, Northeastern was losing prospective full- and part-time students to a newly invigorated and very public University of Massachusetts system. The university's first response was to expand its offerings, making up in volume what it was necessarily losing in price as it tried to at least come close to matching UMass-Boston's lower tuition. When that strategy failed, Northeastern underwent a dramatic downsizing, shedding programs, faculty, and staff and betting on its ability to reposition itself as a selective institution worth the higher price it found itself charging out of necessity.

The president who steered Northeastern through the shoals of repositioning was Richard Freeland, something of a surprise choice who nonetheless had a remarkably sure-footed instinct for what had to be done. The challenge facing Northeastern, he said not just once but over and over again, was to become a top-100 university. At that moment Northeastern was ranked 162nd. Freeland believed that the rankings methodology actually told him and his colleagues what they had to do: improve selectivity, improve retention, increase the proportion of students graduating within six years, garner more sponsored research, and increase Northeastern's media visibility. At the same time, Freeland believed that focusing on the rankings would animate the Northeastern community. To track progress he established an annual scorecard with specific targets to be met each year in each of these categories. And year by year and inch by inch Northeastern improved its metrics—and each year that progress culminated in a higher *U.S. News* ranking. In the year Freeland stepped down as president, Northeastern ranked 98th among national research universities.

Was Northeastern really that much better a university? Most would agree that it was and, more important, that it was worth the increased tuition it now charged. Were the rankings really that important an element in Freeland's march to reposition Northeastern? Skeptics point out that what really transformed Northeastern was the trustees' willingness to authorize more than $400 million in construction debt that allowed the university to become a more residential as well as a more attractive campus. But as in Penn's case, Freeland's scorecard focused the campus's attention on what the rankings measured. The naysayers—and there were some who decried Northeastern's abandoning its traditional mission to become a BU look-alike—never had a chance as long as the rankings improved and Northeastern looked both more solid and more successful.

In the Eye of the Beholder

The differences in perspective between Thacker, Litten, and the liberal arts presidents who champion the Educational Conservancy on the one hand and presidents like Rodin and Freeland who have used the rankings to move their institutions up the rankings ladder on the other are important for what they reveal about the kind of strategies each group wants higher education to pursue. The former see in the rankings a world in which competition and conflict have replaced collaboration and cooperation. Because rankings reward the wrong behavior they diminish the importance as well as the uniqueness of the American educational experience. The latter group is much more accepting of competition as a way of life in a market economy. If handled right, competition brings out the best in people and makes them reach inside themselves to do better. In other words, only by asking "how is the other guy doing?" can an institution know for sure that it is making the most of its assets and opportunities.

These differences in perspective also highlight the importance of determining just exactly what the rankings measure. It is a question we—and here I include my colleagues at Penn's Institute for Research on Higher Education, Susan Shaman and Dan Shapiro—stumbled into when working as part of the National Center for Postsecondary Improvement (NCPI) based at Stanford. As before, we were interested in the structure of the market for undergraduate education, only this time we wanted to build a model that explained the various prices charged by institutions. The result was a market taxonomy—many remember its initial publication that used a paper-airplane–like sketch—that distributed colleges and universities across a continuum that began at one end with the high-priced, highly selective institutions and ended at the other with institutions designed for students more likely to earn their higher educations by taking one course at a time—often from several institutions. The former we dubbed "medallion" institutions; the latter we labeled "user-friendly or convenience" institutions; in between were "name-brand," "good-opportunity," and "good-buy" institutions (National Center for Postsecondary Improvement, 1997).

What we liked about the taxonomy was on the one hand, its ability to actually predict the prices institutions charged, and on the other, its link to commonsense definitions of market attractiveness. One of the parlor games we liked to play was to ask a group of educators to list, preferably in rank order, the two dozen or so "medallion" institutions that resided at the top of the market. It was a test almost no one failed.

The statistical model underlying our market taxonomy had the additional advantage of being remarkably simple. Just five variables were needed to predict price: the percentage of an institution's freshman class that graduated with a baccalaureate degree within six years of initial enrollment; the proportion of applicants an institution admitted and the proportion of those admitted applicants who actually enrolled; the geographic region of the country in which the institution was located; and whether it was a public or private institution. The most powerful variable was the six-year graduation rate. Probably the best explanation of what made that variable so important in predicting what price an institution could charge was offered by a graduate student at the University of Pennsylvania who, in a previous life, had been a Wall Street trader. "Think how the best fund traders hedge their bets," he told us. The extra tuition a medallion can charge was a premium against the risk of not graduating and thus losing the value of the money the student had already invested. The fact that the country's most selective students were attracted to as well as sought out by these medallions made enrollment there that much more of a bargain.

Inevitably, we were asked how well our market taxonomy tracked with the *U.S. News* rankings. The answer, we quickly learned, was unnervingly well. Both the market taxonomy and the *U.S. News* rankings produced roughly the same list in roughly the same order. But then again, we should not have been surprised, since both analyses use the six-year graduation rate as a key variable. What did surprise us, however, was the fact that we could largely replicate the *U.S. News* list using just two variables from the magazine: six-year graduation rate and score on the reputation survey. Among the 50 top liberal arts colleges the fit was almost perfect. Research universities presented a slightly more complex result, since the *U.S. News* list commingled public and private institutions. Once a research university's control—either public or private—was taken into account, the result nearly matched that of the liberal arts colleges.

We now had our answer. The *U.S. News* rankings measure not quality—a conceit only *U.S. News* itself still promotes—but *market position*. Once that proposition is understood it becomes perfectly clear why presidents like Rodin and Freeland are attracted to the rankings. Both made improving their institutions' market position the focus of their presidencies. At the same time, the fact that the *U.S. News* rankings measure market position and not much else helps explain the growing discomfort of educational reformers in general—and liberal arts college presidents in particular—who want to de-escalate institutional competition by instead stressing the importance of educational values and institutional collaborations.

Customer Satisfaction

U.S. News is not alone in seeking to rank American colleges and universities in order to determine who is best. The University of Illinois-Chicago maintains a rankings Web site that contains thumbnail sketches of two dozen separate ranking schemes. In addition to the *U.S. News* rankings, the Web site lists entries that vary from "Young America's Foundation Top Ten Conservative Colleges" to "New Mobility Disability-Friendly Colleges" to "John Templeton Foundation: Colleges That Encourage Character Development."

Most of these other ranking schemes use methodologies roughly similar to *U.S. News*'s collection of statistical variables, which are indexed to produce a rank ordering. The most important exception to this pattern is the rankings of MBA business schools conducted by *BusinessWeek*. Starting in 1988 *BusinessWeek* began an every-other-year look at what it called the "Best B-schools." Instead of compiling a statistical portrait of the schools, however, the magazine chose instead to "ask the customer"—or, in this case, the two sets of customers most important to business schools: their students and the corporate recruiters who do the interviewing and the hiring. What the *BusinessWeek* rankings measure, the magazine proudly tells its readers, is nothing less than "customer satisfaction."

For its 2006 ranking of B-schools *BusinessWeek* contacted 16,565 recent graduates of the top 100 business schools. In all 9,298 responded, for a more than respectable 56% response rate. Only two schools refused to supply the names and addresses of their recent graduates: the Harvard Business School and the University of Pennsylvania's Wharton School. Using public records *BusinessWeek* tracked down and got responses back from 39% of the recalcitrant institutions' recent graduates. In the final result, Wharton ranked second and Harvard fourth; the first and third spots went to the University of Chicago Graduate School of Management and Northwestern University's Kellogg School, respectively.

The survey itself was a 50-item questionnaire that asked about everything from what the graduates thought of the teaching at their schools to the kinds of career services the school provided to their initial salary upon graduation with an MBA. *BusinessWeek* went to considerable lengths to ferret out the possibility that the schools were trying to game the survey by hiring two CUNY educational psychologists to test the probability that individual institutions were coaching their students to ensure the best possible responses. In the final rankings, a school's score on the student survey received a 45% weight in the final rankings. Also receiving a 45% weight was each school's score on the corporate recruiter survey. Recruiters were asked to judge the quality of the students they interviewed for employment in terms of their communication skills, aptitude for teamwork, and analytic skills. Beginning in 2006 *BusinessWeek* began using a weighted average of the recruiter survey that included the current plus the previous two survey administrations in order to provide a more robust and stable total score.

The final 10% of a school's score reflected what *BusinessWeek* called "intellectual capital." For this index the magazine focused on the school's faculty, noting their publications in the field's top 20 publications, from the *Journal of Accounting Research* to the *Harvard Business Review*. *BusinessWeek* also searched the *New York Times*, the *Wall Street Journal*, and its own archives, adding points if a professor's book was reviewed there. The scores were then adjusted for faculty size.

What the *BusinessWeek* rankings promise and deliver is an ordering of business schools based on the outcomes that matter most to students (average salary after graduation, career support, engaged teaching) and to the companies that hire them (capacity to communicate, analyze, and be an effective team member).

What the *BusinessWeek* rankings also provide first and foremost is a diagnostic tool that tells the aspiring institution what it has to do to rise in the rankings. The folklore that surrounds the competitive world of B-school rankings abounds in stories of institutions caught short and their subsequent scramble to set things right. My favorite centers on a top-10 business school that suddenly plummeted in the rankings in large part because it had angered its student body. In the midst of the ensuing furor the school went about the business of recruiting a new dean. One candidate arrived with a detailed analysis of the *BusinessWeek* rankings, telling the interview committee what the school needed to do to right itself. When the candidate sensed no one was really interested, he withdrew his candidacy. The school has yet to return to the top 10.

Purposes

Actually, the *BusinessWeek* rankings serve both of the principal purposes of educational rankings. On the one hand, the rankings provide, as the magazine proclaims, an important measure of "customer satisfaction." In the hands of an adroit dean, the rankings are a guide to quality improvement because they highlight both the successes and the failures that matter most to students and employers. At the same time, the rankings clearly signal market position—those at the top have the same market power among B-schools that the colleges and universities at the top of the *U.S. News* rankings of undergraduate institutions have in that domain. And no one seems too worried about the very public nature of the exercise itself. Given the institutions' educational commitment to competition and market forces, one should expect that judgments of quality and market position should be both public and entangled.

For undergraduate education this commingling of quality improvement and public measurement presents a host of difficulties—just ask the academics and institutional leaders responsible for the National Survey of Student Engagement. The NSSE, as it is fondly known, was the brainchild of Russ Edgerton, who was first the president of the American Association of Higher Education (AAHE) and later the educational director of the Pew Charitable Trust, which provided NSSE's start-up funds. His collaborator was Peter Ewell, who over the last two decades has solidified his position as the researcher who best understands learning outcomes and how they are achieved. Together, Edgerton and Ewell argued that while it was not really possible to test for learning outcomes, it was possible to test for the conditions that promoted the best learning—what they came to summarize as the conditions promoting student engagement.

The NSSE itself is a four-page survey given to a sampling of college and university juniors and seniors that asks questions about what they typically do both in and out of class, how much they study, how often they write papers, how often they engage in collaborative learning projects, and how many hours per week they work for pay. Toward the end of the instrument the student is asked to evaluate the "quality of your relationships with people at your institution" (i.e., other students, faculty, and administrative personnel). The student is given a seven-point scale for each group, ranging from "unfriendly, unsupportive, sense of alienation" to "friendly, supportive, sense of belonging" for fellow students; from "unavailable, unhelpful, unsympathetic" to "available, helpful, sympathetic" for faculty; and from "unhelpful, inconsiderate, rigid" to "helpful, considerate, flexible" for administrative personnel.

The NSSE has been wildly successful, particularly among public institutions and lesser-ranked private colleges and universities. To date, more than 1,200 baccalaureate institutions have administered the NSSE to a sample of their students—and many repeat the survey on an every-other or every-third-year cycle. What the NSSE has provided these institutions is a clear sense of what they need to improve specific to student engagement and, by implication at least, learning outcomes. Faculties have been put on notice when the NSSE shows they have not been attentive enough. Curriculum committees have taken note when the NSSE reports that, compared to similar institutions, students on the committee's campus do not write enough or do not speak up in class regularly or do not engage in collaborative learning projects on a regular basis.

True, NSSE institutions have been known to festoon their Web sites with NSSE factoids like "we rank in the 90th percentile" on particular survey items. For the most part, however, the NSSE is advertised as and used for purposes of improvement, though from the outset Edgerton, in particular, believed that in time colleges and universities would grow comfortable with having their NSSE results in the public domain as a kind of natural antidote to the *U.S. News* rankings. Most institutions, however, have demurred, wanting to contemplate their results in private.

Last year the NSSE organization joined with *USA Today* to launch an "Initiative to Focus on Meaningful Indicators of Collegiate Quality," an effort that promised to publish better measures of institutional quality. But for the moment, only the good results will be presented in *USA Today*, those that highlight what the partners described as "different types of colleges and universities that involve their students at high levels in effective educational practices."

Critics and observers who want higher education to be more accountable want the NSSE results released—and there's the rub. Will institutions still use the NSSE if the results are made public and used for the purpose of holding them accountable to their accreditors and, in the case of public institutions, their state regulatory agencies? Won't the temptation be, as Colin Diver has suggested is already the case with the *U.S. News* rankings, to "game" the system by lessening the probability that unhappy or disengaged students make it into the test's sample? Won't institutions, subtly or not, encourage those students selected to fill out the NSSE to "win one for the Gipper" by focusing on what is right as opposed to what is wrong with the institution?

Even tougher to gauge is the reaction of the faculty. Already there is a tendency for the very best or at least the top-rated and most expensive undergraduate institutions not to use the NSSE, sensing that they have little to learn and a fair amount to lose. If the past is any guide, faculty will find it in their interest to attack the instrument itself, bringing all their considerable verbal and mathematical firepower to bear to demonstrate that NSSE results are at best inconclusive. What had hitherto been seen as an interesting diagnostic tool, no doubt with some methodological problems, may become in the eyes of many a political tool of uncertain value that will likely be used to bludgeon otherwise right-thinking faculty into adopting cookie-cutter approaches to undergraduate education.

That is already the fate befalling another instrument that seeks to measure learning outcomes. The Collegiate Learning Assessment (CLA) was originally developed by researchers at the RAND Corporation and then later spun off to something called the Council for Aid to Education. The CLA is a Web-based instrument that tests samples of an institution's students in four broad areas: critical thinking, analytic reasoning, written communication, and problem solving. While the test itself is automated, its grading is not and depends on the ability of the test givers to train a sufficient number of graders.

Presumptive regulators, in particular, have been much impressed with the CLA. Charles Miller, the chair of the Spellings Commission,[1] early on urged his commission to adopt the CLA as exactly the kind of test that would tell the nation which colleges were up to the mark and which were not. From Miller's perspective, the CLA offered precisely the objective measure that would allow everyone—government funders and regulators, accreditors, and potential students and their parents—to at last hold higher education accountable. The CLA would be to higher education what the battery of No Child Left Behind tests are to primary and secondary education.

The problem is that the CLA won't do. It is expensive to administer. Its results are difficult to interpret and verify. It largely tests general education rather than the kind of advanced and technical knowledge that forms the basis for most collegiate majors. At least in the pilot tests made public so far, there has been an uncomfortable correlation between SAT/ACT results and the CLA—hence, the institutions with the most selective admissions practices are very likely to be the ones that score highest on the CLA. The Council for Aid to Education counters that the CLA really allows the measurement of value added—that is, how much more the student knows upon leaving college—using the students' high school SAT/ACT results as the benchmark. But here too the results have been confusing, leading institutions to ask which is the better measure, the SAT/ACT or the CLA?

Actually, there is a prior technical difficulty that renders the CLA less than ideal. While the CLA results are to be used to evaluate an institution and its learning effectiveness, students take the test. Since they have no personal interest in the outcome—their scores, for example, will not be reported to prospective graduate schools or employers—they are not likely to prepare for the test or necessarily do their very best. Faculty are quick to point out that they have little intention of being judged by student performances on a test that does not matter to the student. My guess is that given the choice, faculty would prefer an old-fashioned beauty contest survey or even the NSSE to the CLA.

More of the Same

Despite the muddle caused by the rankings, two important aspects of the problem are now clearer. First, the *U.S. News* rankings are here to stay, though October now comes in late August with the prerelease of the list of "Best Colleges." Autistic savant or not, *U.S. News* will continue to go with the numbers—and the story those numbers tell will remain largely the same.

At the same time, calls for an alternative to the *U.S. News* rankings will increase in both frequency and fervor. Policy wonks and would-be regulators intent on remaking higher education in their own image will demand more transparency and more testing in the name of making higher education more accountable. Advocates for alternative learning paradigms will insist that more robust measures of learning be included in the rankings' calculations. The most coherent and perhaps most successful calls for developing cogent alternatives to the *U.S. News* rankings will come from those like Lloyd Thacker of the Educational Conservancy and Doug Bennett, Earlham College president and chair of the NSSE board,

[1] In the fall of 2005, U.S. Secretary of Education Margaret Spellings appointed a national Commission on the Future of Higher Education. Charles Miller of Texas was its chair, and I joined 17 others drawn from both within and without higher education as members. Despite the Commission's early promise, it delivered a report that failed to lead to meaningful change within higher education.

who want to make the conversation about what constitutes a good college or university better reflect the educational values that ought to shape the enterprise in the first place.

There ought to be calls as well to make rankings and institutional evaluations more an exercise in gauging customer satisfaction—more like *BusinessWeek*'s rankings of business schools. Students actually know better what they want and need than the reformers and critics give them credit for. Ask recent graduates what they thought of the institutions they attended; ask them if they are happy in the jobs they took following graduation; ask them how much money they are making and to what extent they are actually using on the job what they learned inside and outside the classroom.

But moving from statistical measures of market position to customer satisfaction will not be so easy. *BusinessWeek* learned this lesson in 2006 when it sought to match its rankings of MBA programs with a similarly constructed ranking of undergraduate schools and business programs. *BusinessWeek* wanted to deal with only a hundred or so top undergraduate business programs and actually settled for just 61 institutions for its inaugural effort. In all, the magazine got valid e-mail addresses from the colleges and universities themselves for just over 81,000 current business program students. The problem was that only 19,550 students completed the online survey, a 24% response rate. Most analysts think a 35% to 45% response rate is necessary to draw the kind of conclusions *BusinessWeek* wanted to draw. For its MBA survey the response rate was 56%.

By comparison, developing a valid customer satisfaction survey for colleges and universities conferring baccalaureate degrees will prove a gargantuan task involving some 1,800 institutions and a sample of 900,000 potential respondents (at 500 students per institution). Best-guess estimates run in the neighborhood of $8 million to $9 million per administration. It is unlikely that any commercial venture—or even a foundation—would risk that kind of funding, leaving the federal government and its Department of Education as the only likely sponsors. Whether institutions would trust the department not to misuse the data is an open question, particularly given that organization's recent attempts to federalize the accreditation process and its pushing of a national student registry that would allow it to track individual students regardless of how many different institutions they attend.

If these security and political concerns can be assuaged, however, the payoff would be considerable—particularly if the resulting survey contains some NSSE-like items along with the bread-and-butter satisfaction questions *BusinessWeek* asks of the students in its survey. Initially, I suspect, most consumers will stick with the market position indicator championed by *U.S. News*, though in time they too may begin to look at measures that tell them something about the learning environments of the different institutions. The big winners would be the faculty, once they got over the notion that the survey's results were just another scorecard or ranking and began treating them as a set of comparative diagnostic tools. The result just might be a giant step toward creating the culture of evidence the rankings were supposed to engender in the first place.

References

Diver, C. (2005). Is there life after rankings? *The Atlantic*, November 2005, online: www.theatlantic.com/doc/200511/shunning-college-rankings.

Educational Conservancy (2007). College admissions: What are students learning? Conference materials, *Beyond ranking: Responding to the call for useful information*, Yale University, September 25, 2007.

National Center for Postsecondary Improvement (1997). The landscape, in search of strategic perspective: A tool for mapping the market in postsecondary education. *Change*, November/December, 23–38.

Zemsky, R., & Massy, W. (1990). The rain man cometh. *AGB Reports*, January/February, 21–23.