

As the NRC prepares for another survey of U.S. graduate schools, educators find themselves in a heated debate about how to measure—and define—a quality education

# Graduate Educators Struggle To Grade Themselves

David Schmidly of Texas Tech University in Lubbock can't wait for his next report card. Five years ago his school's 17 graduate programs earned a dismal composite score—92nd out of 104 comparable institutions—in a National Research Council (NRC) assessment of U.S. academic programs. But Schmidly, vice president for research and graduate studies, parlayed that low ranking into a successful bid for more state money for faculty and help in boosting the school's research budget by 50%. He expects that the NRC's next survey, in 2003–04, will validate his efforts, boosting the school's reputation and making it easier to attract more money and better students and faculty.

Top-tier schools rely on rankings, too. Just last week, Yale University President Richard Levin cited the university's #30 ranking in engineering as proof of the need for a half-billion-dollar science and engineering construction binge (see p. 579). And holding the top ranking in physiology and pharmacology in the 1995 NRC study, he noted, "makes it imperative that we invest enough to stay at the forefront in those fields." But not everyone is so enamored of grades for graduate programs. Patricia Bell, chair of Oklahoma State University's sociology department—ranked dead last out of 95 participating programs in the NRC's last survey—scoffs that the reputational rankings, which are based entirely on scholars' opinions of their peers, are a "popularity contest" that "dismisses the value of teaching."

Such conflicting opinions are part of the debate over how to rank graduate research programs, a debate that has sharpened in recent months as the NRC gears up for its third attempt since 1982 to plumb the world's best academic research system. Almost everyone agrees that assessing graduate research programs is a useful tool for administrators who manage the programs and provides an important window into the system for students, faculty, funding agencies, and legislators. But that's where the agreement ends. All previous U.S. assessments—there have been 10 major attempts since 1925, most of them heavily dependent on reputational rankings—have been controversial, as is an ongoing exercise in the United

Kingdom (see sidebar). And now, just as college football fans argue endlessly about the relative importance of such factors as won-lost record, coaches' ratings, and strength of schedule in choosing the #1 team, those who follow graduate education are weighing the merits of reputation vs. quantitative measures—such as numbers of published papers and amount of research funding—in assessing research and de-



bating how to measure the caliber of teaching and the fate of graduates.

The stakes are high. Top-ranked programs attract more funding as well as high-quality faculty and students, while "low rankings can shrink or even kill off a program," notes Vanderbilt University historian Hugh Graham, author of a well-regarded 1997 book on the history of U.S. research universities. And it extends far beyond the campus. "It's a spiraling effect," says a spokesperson for Arizona State University (ASU) about its well-regarded business school. "High-achieving alumni are valuable to their companies, who see ASU as a good place to invest their money. Corporate donations allow us to offer talented faculty the salaries that attract and retain them, which contributes to higher rankings."

NRC officials hope to begin exploring

these issues later this year with a series of pilot studies that will culminate in a full-blown survey in 2003–04. An added factor is a highly visible rating system that already exists: The news magazine *U.S. News & World Report* publishes best-selling annual issues that tout the country's best graduate and professional programs and the best undergraduate institutions via a reputational ranking. Most university officials say the magazine doesn't capture the opinions of real peers, and they accuse it of deliberately shaking up the ratings to retain reader interest—a charge that editors hotly deny. But universities are also quick to cite flattering results in press releases and recruitment ads.

The magazine's popularity makes it imperative for academics to stay in the game, says John Vaughn of the Association of American Universities. The AAU hopes this spring to begin its own 5-year effort to collect graduate education data from its 59 members, which include most of the country's research powerhouses. "We shouldn't cede our capacity for thoughtful analysis to a commercial operation that must put business first," says Vaughn.

And although most administrators prefer the more sober NRC effort, many worry about how it will turn out. "There should be a study," says graduate school dean Lawrence Martin of the State University of New York, Stony Brook, who is also head of a panel of land-grant colleges that has drafted a position paper urging coverage of more fields, greater use of objective research criteria, exploration of some measures of program outcome, and ranking institutions by cluster rather than individually. "The fundamental issue is whether they will do it right."

Despite the clashing opinions about what it means to "do it right," educators agree that an assessment is no trivial matter. The last NRC survey covered 3634 programs in 41 fields at 274 institutions; this time, it will have to do all that and more, says Charlotte Kuh, head of the NRC's Office of Science Education Programs. She hopes to raise at least \$5 million, four times the cost of the 1995 survey and 25 times the 1982 price tag, for the two-phase study.

ILLUSTRATION BY JOE SUTCLIFFE



## Support Grows for British Exercise to Allocate University Funds

Next year British universities go under the microscope for the country's fifth Research Assessment Exercise (RAE), an attempt to rank departments' research output and help the government invest wisely in academic infrastructure. And even some of its most persistent critics acknowledge that the RAE, which has drawn international attention, has overcome a rocky start and is working increasingly well. "Each successive RAE is moving closer to the consensus [of what constitutes high-class research]," says Paul Cottrell, the assistant general secretary of the Association of University Teachers, which opposed RAE's introduction in 1986.

The RAE began as a way to funnel dwindling resources into the best research programs at a time when severe cuts in public spending raised fears of a major brain drain. Unlike the U.S. National Research Council (NRC) ratings (see main text), a school's RAE ranking has a direct impact on government funding. "The better you do [in the RAE], the more money you get," says John Rogers, RAE manager at the Higher Education Funding Council for England, which oversees the exercise. The process, which covers 68 fields, puts every university department or program under scrutiny by an independent panel of peers. Each department receives a score, from 1 to 5\*, that is supposed to be based on four pieces of work submitted by every participating researcher and such information as prizes, outside funding, and research plans. "The gold standard is always international excellence," says Rogers. That score, adjusted for the number of participating researchers, determines funding levels.

Despite its narrower focus and deliberate elitism—last year

about 75% of the council's \$1.3 billion budget went to the top 13% of Britain's 192 institutions of higher education—the British assessment exercise has raised some of the same concerns as the NRC surveys. "Teaching is not esteemed as highly as research and always gets a back seat," Cottrell argues. Although teaching skills are evaluated in a separate exercise, the Teaching Quality Assessment, the outcome is not linked directly to funding. The question of how much panels are affected by a researcher's reputation also remains an issue, although Rogers says attitudes will play a smaller role next year than in earlier exercises.

Rogers calls RAE "the longest standing assessment process on this scale worldwide," and its influence may soon be spreading beyond its borders. In Japan, where university funding has been based on precedent and enrollment and there is little oversight of performance, the government is moving slowly toward greater accountability. This year the Ministry of Education, Science, Sports, and Culture (Monbusho) hopes to create an evaluation organization that will serve initially as an accreditation board to review curricula and to prod universities to raise education standards. But an advisory panel has also recommended that Monbusho begin evaluating university research efforts on a departmental level, with the results somehow tied to funding for new buildings and large-scale equipment.

The RAE approach has also found a home in Eastern Europe, where this spring the Czech Republic hopes to begin a long-awaited review of academic research at its 27 universities. The reports from the visiting panels, which will include foreign scientists, are expected to lead to a two-tier university system that favors a handful of elite schools.

—MICHAEL HAGMANN

With reporting by Dennis Normile and Richard Stone.

## Not by reputation alone

By tackling these thorny issues early, Kuh hopes to avoid the blizzard of criticism directed at the previous survey for flaws ranging from factual errors to a disregard for applied fields. First up is the charge that the NRC relied too heavily on research reputation, one of many categories of data but the sole source for the numerical rankings of programs. For the reputational rankings, NRC asked more than 16,000 scientists to assess the quality of the faculty and the relative change in program strength over the past 5 years for as many as 52 programs. Each rater was provided a list, supplied by the university, of faculty members in each program.

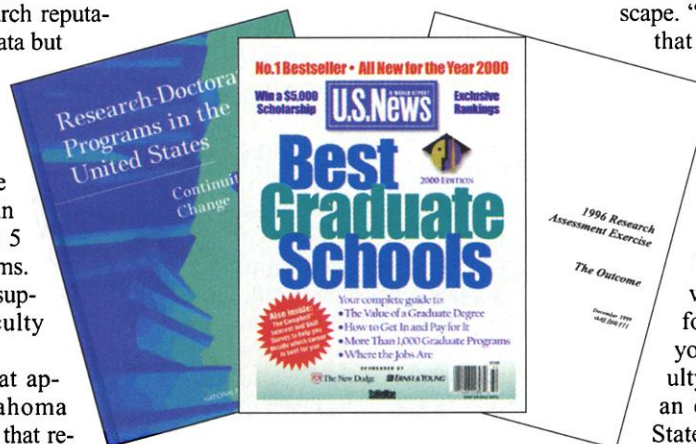
Many academics believe that approach is badly flawed. Oklahoma State's Bell, for example, argues that relying on reputations penalizes what novelist Tom Wolfe has called "flyover universities" like hers that don't have national reputations but emphasize teaching. And it's not just those on the bottom who complain. "Most people think that it was a mistake," says Jules LaPides, outgoing president of the 400-member Council of Graduate Schools (CGS), about the NRC's deci-

sion to gather lots of kinds of data, but to rank programs simply by reputation. "It legitimizes a flawed concept, that there is a single 'best' graduate program for all students. But graduate education is not a golf

productivity that do not rely on the memories of beleaguered reviewers. Such measures as citation impact, levels of funding, and awards, when applied on a per capita basis, he argues, would provide a more accurate picture of the current research landscape. "There are a lot of rising institutions that are being ignored," he says. "The next NRC study should help to reveal this layer of excellence that is waiting to be tapped."

Critics also note that larger departments have an unfair advantage in reputational rankings because of the bigger shadow cast by their graduates, as do those with a handful of standout performers. "The best way to improve yourself quickly is to hire a few faculty superstars," says David Webster, an education professor at Oklahoma State who has written about both NRC studies. But superstars don't necessarily enhance the educational experiences for grad students, he says.

Yet few administrators are willing to jettison reputation. The reputational ratings "don't capture the whole picture, but they capture people's perceptions, and that's important," says Yale's Levin. And even Webster believes that "reputational rankings, for



**Pulling rank.** U.S. News offers a much splashier wrapping for its survey than does the NRC, left, or the RAE.

tournament, with only one winner."

Vanderbilt's Graham and others argue that reputational rankings have become obsolete, as fields expand too rapidly for anyone to remain familiar with all the players. He favors quantitative measures of research



all their faults, provide a type of subtlety that you don't get in more objective measures." Paraphrasing Winston Churchill's views on democracy, he says that reputational rankings of academic quality "are the worst method for assessing the comparative quality of U.S. research universities—except for all the others."

#### Don't forget the students

Focusing on reputation, however, ignores the question of how to calibrate many of the other complex elements that make up a graduate education. "The previous [NRC] survey was misnamed," says the AAU's Vaughn, echoing the views of many. "It was an assessment of the quality of research faculty, not of graduate programs. And we don't really know how to measure the quality of graduate education." Kuh plays down the distinction. "I don't think you can separate the two," she insists, adding that she thinks previous surveys got it right in emphasizing research.

Still, many administrators feel that the next NRC survey must do a better job in exploring the quality of education. That includes such factors as the time to degree, dropout rate, and starting salary of graduates, as well as such intangibles as the quality of mentoring, opportunities to attend meetings, and the extent of career advice offered students. "It's not easy to do, but without it the community support [for the next survey] will vanish," says Debra Stewart, vice chancellor and graduate dean at North Carolina State University in Raleigh, who served on the advisory panel for the 1995 study and who in July becomes CGS president.

Joseph Cerny, vice chancellor for research and dean of the graduate division at the University of California, Berkeley, and his Berkeley colleague, Maresi Nerad, took a first crack at the issue by surveying some 6000 graduates a decade after they received their Ph.D. The study, carried out in 1995 and still being analyzed (*Science*, 3 September 1999, p. 1533), surveyed graduates on the quality of the training they received and whether they would do it again, among other questions. The results were quite different from when the NRC asked peers to rate the quality of both faculty and programs.

"The [NRC] found an almost perfect correlation," says Cerny, who as a member of the advisory panel lobbied unsuccessfully for outcome data to be collected in the 1995 survey. "But when we asked graduates to rank such things as the quality of the teaching, the graduate curriculum, and the help they received in selecting and completing their dissertation, we got dramatic differences. Instead of a slope of 45 degrees, indicating a perfect fit, we got a

20% fit. The graph looked like it had come out of a shotgun."

The data on whether students would repeat their training are also eye-opening. "Computer science ranked the highest, at 85%, and biochemistry was the lowest, at 69%," he says. And the performance of individual programs varied wildly, including two biochemistry programs that scored 100% and one that received only 15%.

Kuh argues that such ratings from graduates have limited value because the information quickly becomes dated and doesn't take into account the variation among students. She adds that a stressful graduate experience could still lead to a successful career. Cerny argues, however, that even stale information on student outcomes would be extremely valuable to the university administrators who run the programs—and to the federal agencies that fund graduate training. "I'd certainly want to know if I was the dean at a school [where only 15% of students would redo their training]," he says. "Even if you consider 60% to be a passing grade, we

found that only one-third of the programs scored at or above that level."

Ultimately, say Kuh and others, the key to a successful assessment is giving customers something they need. Bell and Webster of Oklahoma State say that the NRC and *U.S. News* surveys have had little impact on their university's research policies not because they fared badly but because the yardstick—the research reputation of its faculty—was seen as tangential to the university's main mission of educating students. "It's like MIT's [the Massachusetts Institute of Technology's] reaction to the weekly Associated Press football polls," says Webster. "We're just not a big player in that sport." Kuh hopes that the next NRC survey is good enough to generate as much interest at Oklahoma State as it does at MIT or Yale, setting the standard for anybody interested in assessing graduate education. "The U.S. is at the top of the world in higher education," says Graham, "and it's too important a topic to produce reports that aren't used."

—JEFFREY MERVIS

#### MEETING

## An Integrative Science Finds a Home

**ATLANTA**—This year's annual meeting of the Society for Integrative and Comparative Biology (SICB), held here earlier this month, marked a milestone for the fledgling discipline known as evo-devo biology. Beginning about a decade ago, modern biologists realized that they might glean clues to how organisms evolved by studying the genes that control development (*Science*, 4 July 1997, p. 34). Now, the discipline is so strong that last year it gained its own division within SICB and was invited to present its inaugural symposia in Atlanta. Prominently featured were new findings on the genes needed for butterfly wing development and on the homeobox genes, a key group of developmental genes involved in organizing animal body plans.

#### Hox and the Simple Hydra

Evolution, like development, entails a transformation of shape and form, but on a much longer time scale. The discovery by developmental biologists that the *Hox* (for homeobox) genes play a major role in guiding this transformation in fruit fly and vertebrate embryos posed an intriguing question. Researchers wondered whether the appearance and proliferation of these genes 550 million years ago made possible the transformation of early life into the wide range of complex shapes and forms seen today. Results presented at the meeting supported this idea.

The work, which comes from the labs of



**Hox history.** Hydrazes seem to need *Hox* genes to define their body axis.

two independent investigators, John Finnerty of Boston University and Hans Bode of the University of California, Irvine, deals with the homeobox genes of cnidarians. These animals—a group that includes corals, hydra, sea anemones, and jellyfish—resemble the fossils of some relatively simple pre-Cambrian creatures that existed before the large-scale diversification took place. Both groups of organisms lack a key

feature of most modern animals—bilateral symmetry, a balanced shape that distributes the body in mirror-image halves along a line

CREDIT: G. OLIVER/IMAGES UNLIMITED