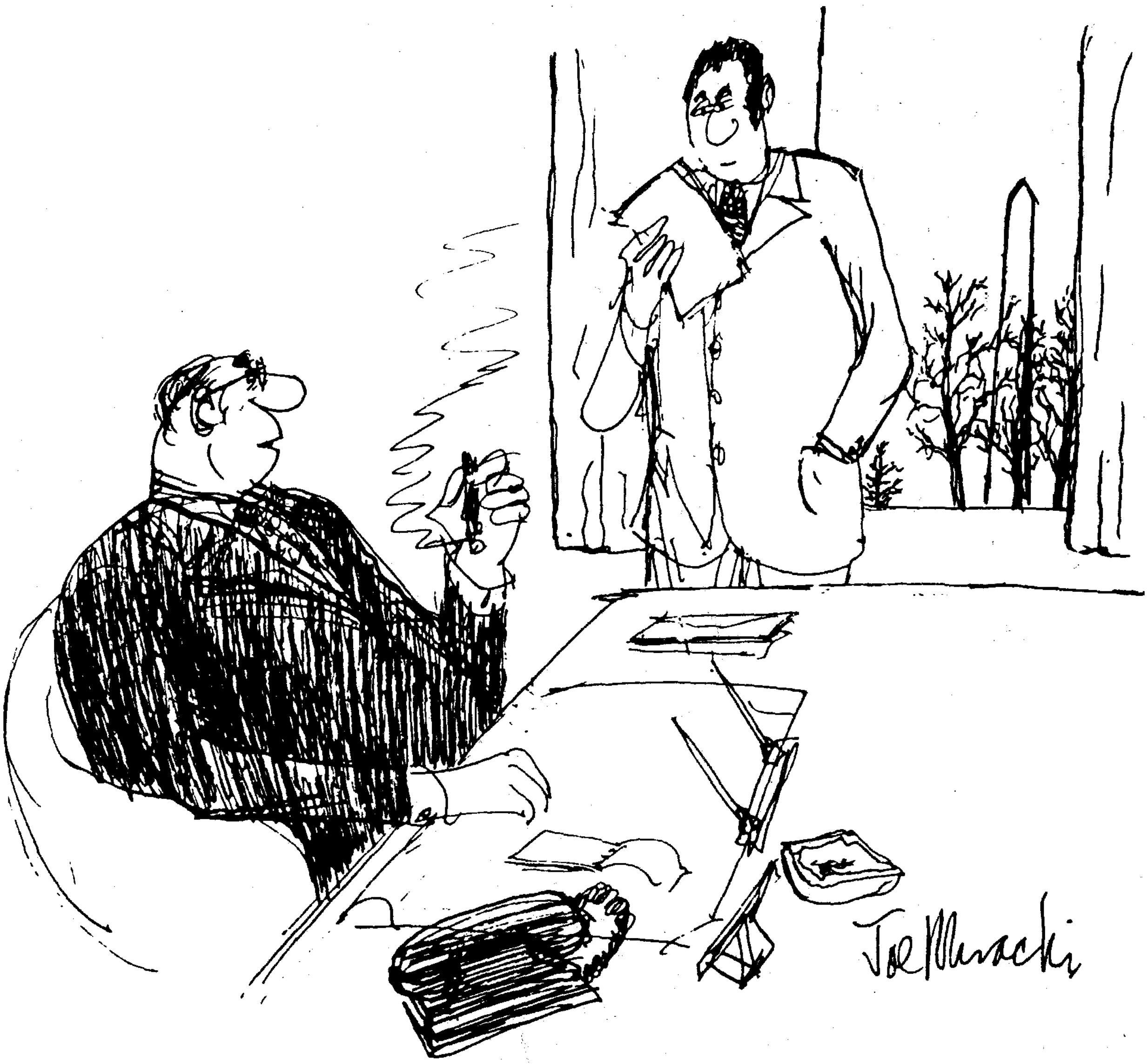


"That's the gist of what I want to say. Now get me some statistics to base it on."

White Lies, Damned Lies, and Statistics

Year after year, decade after decade, well-meaning educational researchers and policymakers continue the search for the perfect statistic.

by Lisa C. Heinz

Education is a profligate generator of silly statistics. In particular, attempts to measure the *quality* of education produce data-rich, fervently read reports. (Warning—this article contains a good dose of said silly statistics, and three unanswerable questions.) Year after year, decade after decade, well-meaning educational researchers and policymakers continue the search for the perfect statistic. Over the past year, the U.S. Department of Education spent $78 million on educational research and statistics; the National Science Foundation doled out another $5–8 million for science-education analysis and statistics.

Caltech usually comes out immodestly high in university quality rankings, whether in magazine articles or college guides. Caltech was fourth (behind Yale, Princeton, and Harvard) in *U.S. News & World Report*'s latest ranking of major universities, released in October.

University quality = research quality?

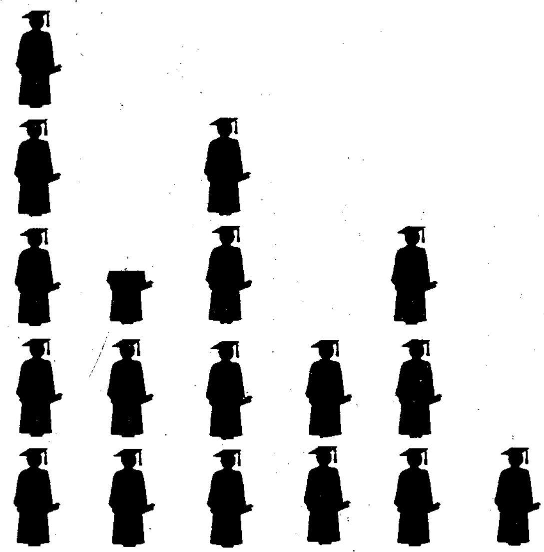
Such rankings usually emphasize Caltech's research preeminence. For its size, Caltech comes out well—number 36—in the favorite Washington statistic for research quality, federal R&D dollars. Federal R&D receipts may be the most obvious, and easiest, metric, but is an unsatisfactory measure of a university's research performance.

The National Science Foundation is sponsoring research on more sophisticated metrics of research quality. Larry Leslie and others at the University of Arizona have compiled a multidimensional research activity index (RAI) for the top 200 research universities of 1980. The RAI combines 14 weighted variables, such as the amount of R&D funding from various sources, total research expenditures, employed scientists and engineers, numbers of full-time graduate students and postdocs, PhDs awarded, and a research library score from the Association of Research Libraries Index. However, the RAI still measures the scale, rather than quality, of research. The Top Ten on the RAI generally are the familiar big-name research universities. The University of Arizona group is currently developing RAIs which adjust for institutional size and for individual fields, and is investigating how to include measures of research outputs (for example, publications and citations) as well as research inputs (dollars and people).

But the amount of federal R&D money a university can attract is fairly far removed from the quality of its education. Unanswerable question #1: Does first-rate research foster first-rate education? Universities have multiple personalities. There is a natural tendency for the strongest persona, whether research or education, humanities or engineering, to dominate. Integrating research and education into a harmonious, yet unique, university takes deliberate effort. Nobel laureates and multimillion-dollar research grants may be all well and good for the *research* university, but the *education* part of the university must be attended to as well.

Trying to measure education

It is education, rather than research, that catches the popular headlines. Unanswerable question #2: Is it possible to measure the qual-



Unanswerable question #2: Is it possible to measure the quality of education?

Table 1
Leading Undergraduate Sources of Science and Engineering PhDs
(s/e PhD productivity)

rank, adjusted for institution size		percent*	rank, EMP **	rank, life sciences	rank, not adjusted for institution size
1	Caltech	44	1	2	32
2	Harvey Mudd	31	2	20	207
3	MIT	21	3	3	2
4	Reed	21	6	3	104
5	Swarthmore	17	11	—	78
6	Cooper Union	14	4	—	146
7	U. Chicago	14	14	10	23
8	Radcliffe	13	—	8	154
9	Rice	12	5	—	51
10	Haverford	12	—	—	179
11	Carleton	11	15	13	111
12	Pomona	10	20	13	108
13	Grinnell	10	—	17	159
14	Oberlin	10	35	32	50
15	UCSD	9	21	5	112
16	Antioch	9	—	39	131
17	Cornell	9	23	8	5
18	Princeton	9	12	—	28
19	Wesleyan	9	—	—	140
20	Wabash	9	23	13	222

* Percent of all baccalaureate graduates from that institution who went on to get s/e PhDs. The study covered graduates between 1950 and 1965, to insure that they would have earned PhDs by 1986.

** EMP is engineering. mathematical and computer sciences, and physical sciences (such as astronomy, chemistry, geology, environmental sciences, physics). Life sciences includes agricultural, biological, and health sciences.

ity of education? There is no simple quantitative measure that can be applied to so nebulous a thing as learning. People turn to proxies, such as SAT scores, class size, and student-faculty ratios to size up a university's learning environment.

The university may be considered a black box: students go in, and (for better or worse) future citizens and workers come out. When most analysts look at the black box, they see a university's dollars, books, buildings, or graduation requirements; the "quality" of its faculty and students. Less official but popular measures are the greensward-asphalt ratio, average parties per week, per-capita beer consumption, the ratio of total downhill miles plus annual snowfall to the distance to ski slopes, and the gut-course/killer-course ratio.

The more pragmatic analyst, rather than looking at the black box, might consider the *output* of a university. After all, parents and students are interested not only in the college experience, but in how an $80,000-plus college education will advance an eventual career. Today's students seek power careers and high salaries, while social activism and life enrichment have waned in value; according to a recent UCLA survey, being "very well off financially" is the top goal of incoming freshmen, a goal that has risen steadily in popularity since the mid-1970s.

Ideally we might like to measure how well a college grooms students for successful, accomplished, rewarding, and satisfying lives and careers. However, the desire to measure some-

Table 2
Leading Undergraduate Sources of s/e PhDs
(rank, not corrected for institutional size)

1	UC Berkeley
2	MIT
3	University of Illinois
4	University of Michigan
5	Cornell
6	University of Wisconsin
7	CUNY — City College
8	UCLA
9	University of Texas — Austin
10	Harvard
⋮	
32	Caltech

Covers bachelor's degrees awarded between 1950 and 1975. UC Berkeley spawned nearly four times as many s/e PhDs as Caltech.

SOURCE: Betty D. Maxfield, "Institutional Productivity: The Undergraduate Origins of Science and Engineering PhDs," U.S. Office of Technology Assessment Contractor Report, July 1987, Appendix A.

However, the desire to measure something interesting, such as learning, must always be compromised by the unfortunate necessity to count what is countable.

thing interesting, such as learning, must always be compromised by the unfortunate necessity to count what is countable, such as degrees or test scores.

PhD productivity

One attractive measure for the science-minded is a college's output of students who go on to become quality researchers. Now, what we can count fairly easily is a college's baccalaureate graduates who go on to get PhDs in science or engineering. Quite a few studies over the years have attempted to calculate this sort of "PhD productivity." Although PhD productivity is a fairly coarse measure, it is one of the best proxies available for a college's "output." The rest of this article discusses a recent study of science and engineering (s/e) PhD productivity, undertaken by the U.S. Office of Technology Assessment and what it does and does not tell us. This study calculated the number of BS or BA graduates in all fields between 1950 and 1976, who received PhDs in science or engineering from any U.S. institution between the 1950s and 1986. Science/engineering includes the social sciences, and the study includes only colleges that sent more than 50 students on for PhDs during the study period. (More information on methodology and results are in the original report, available from the Office of Technology Assessment, U.S. Congress, Washington, DC 20510.)

Table 1 shows some of the results of the study of s/e PhD productivity of American colleges and universities. This study developed and evaluated universities' s/e PhD productivity ratio: the percent of all graduates from that college who went on to earn a PhD in science or engineering. The results show Caltech and Harvey Mudd as clear leaders, with MIT and Reed not far behind. Over the study period, 44 percent of Caltech baccalaureates went on to earn s/e PhDs.

This productivity ratio adjusts for the size of the institution. Certainly, it is nice to know which universities send the largest numbers of warm baccalaureate bodies on for s/e PhDs. However, as Table 2 and the last column of Table 1 show, high absolute numbers of eventual s/e PhDs do not necessarily mean that the university has a high productivity. The university that sent the greatest number on for s/e PhDs is UC Berkeley, but it ranked 26th when size is taken into account. Conversely, Caltech was first in productivity, but ranked 32nd in absolute numbers of eventual s/e PhDs. The appealing thing about PhD productivity (besides making Caltech look good) is that highly productive institutions should provide lessons about the type of college environment that fosters students' interest in s/e graduate study, and their ability to earn a PhD.

A more sophisticated measure of a university's output counts only those s/e PhDs who go on to do active research. This might be called a university's "researcher productivity." Limitations in data collection and coding make this analysis a difficult proposition. Preliminary work, done by type of institution rather than individual college, has revealed that s/e PhDs who had done their undergraduate work at technical institutions, such as Caltech, MIT, IIT, and Carnegie-Mellon, were by far the most likely to go on to careers in research. On the other hand, s/e PhDs who had come from women's or black colleges were much less likely to go into research.

One flaw in this and all similar studies so far (due to the difficulty of extracting field-specific baccalaureate data from the paper-ridden seventh circle of data hell in the Department of Education) is that the basis of all calculations—bachelor's degrees—aggregates all fields. In these studies a college's s/e PhD productivity is based on the percentage of baccalaureate graduates in all fields who went on to get s/e PhDs. In reality, schools differ strikingly in the percentage of their baccalaureates who take science or engineering degrees. Common sense would argue that colleges with a high proportion of undergraduates who major in science are much more likely to send a higher proportion of their baccalaureates on to s/e PhDs. Aggregating all

Academic reputation is the single most important consideration in students' choice among colleges.

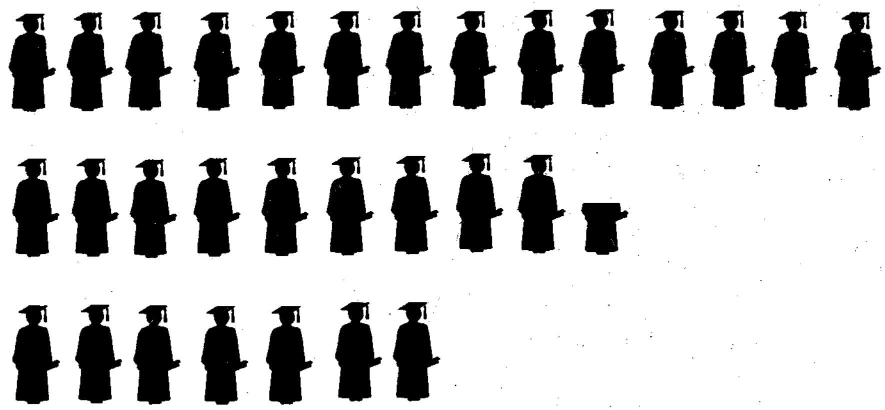


Table 3
s/e Baccalaureate to s/e PhD Productivity of Universities

rank	productivity ratio (percent of s/e baccalaureates who earned s/e PhDs)
1 **Caltech**	43
2 **Harvey Mudd**	34
3 **Reed**	26
4 **MIT**	24
5 **U. Chicago**	21
6 **Haverford**	17
7 **Swarthmore**	17
8 **Pomona**	16
9 **Oberlin**	16
10 **Carleton**	16
11 **Cooper Union**	16
12 **Rice**	14
13 **Bryn Mawr**	14
14 **Amherst**	13
15 **UCSD**	13
16 **Princeton**	13
17 **Cornell**	13
18 **Yale**	12
19 **Johns Hopkins**	12
20 **Brown**	12

SOURCE: Betty D. Maxfield, "Persistence in Higher S/E Education: S/E Baccalaureate to S/E Doctorate Productivity of U.S. Baccalaureate-Granting Institutions," U.S. Office of Technology Assessment Contractor Report, September 1987.

fields tends to inflate the s/e PhD productivity of nerd-packed schools like Caltech. Other, more diverse schools, particularly the liberal arts colleges, are surprisingly productive of s/e PhDs, given the high percentage of their undergraduates that take degrees outside the sciences.

A follow-on study corrected this flaw, calculating an s/e PhD productivity which looked only at baccalaureates who majored in s/e. When this correction was done, however, results changed surprisingly little (see Table 3). (This constancy may in part be due to counting social sciences as part of s/e.) Some schools, such as Oberlin, Pomona, and the University of Chicago, did have much higher PhD productivities when the college's emphasis on science was taken into account.

A more significant flaw is the difficulty of controlling for differences in the quality of incoming students among various universities. Certainly, the quality of the student body is at least as important as the faculty and offerings of the university itself. Unanswerable question #3: Is the high productivity of a university like Caltech due to the superior quality of its undergraduates, or to Caltech's providing a superior education?

Another recent Office of Technology Assessment study concluded: ". . . active researchers come from graduate study at a small number of top research universities. These elite research universities, however, draw on a broader base— the successful graduates of highly productive undergraduate institutions. The career decisions made by PhD recipients are influenced as much by their college experiences as by their graduate school."

Different methodologies and different simplifications result in slightly different results, but studies tend to converge on the same set of productive colleges. Taken together, these studies argue convincingly that some institutions are more likely to send their undergraduates on for s/e PhDs. What can we learn from this?

The link between PhD productivity and educational environment

Even a partly qualitative, partly quantitative ranking allows a vaguely systematic analysis of mushy things like education.

What is it that makes a college highly productive? One answer comes from a report issued by a group of liberal arts colleges, known as the Oberlin 50. In touting the reasons why small liberal arts colleges produce more than their share of scientists, the Oberlin report claimed, "personalized instruction by senior scientists and widespread student involvement in research are the primary distinguishing features of these institutions, and account for their record in both attracting and producing young scientists."

The interesting thing is that the message of the liberal arts colleges—reknowned for their emphasis on teaching rather than research—echoes a core characteristic of Caltech: close interaction between student and mentor in an intellectual apprenticeship, in the laboratory as well as in the classroom. (Another shared characteristic is carefully selected, high-quality students.)

The PhD productivity studies and the Oberlin report share some basic lessons: small is good, research is good, interaction is good (and admitting superior students doesn't hurt either). This suggests that other institutions might encourage the intimate, interactive, research-immersed approach to education with more and better instructional labs, more chances for undergraduates to get involved in research, and more contact between senior faculty and undergraduates. To encourage faculty to dawdle with undergraduates, colleges might promote lighter teaching loads, starter grants, teaching sabbaticals for research faculty, and research sabbaticals for teaching faculty.

Are rankings worth their shortcomings?

Are rankings worth their unavoidable shortcomings? Academic reputation is the single most important consideration in students' choice among colleges, according to a 1987 UCLA study. Part of this is that many students and parents believe that a good reputation promises a good education. But these wise students and parents also realize that a good reputation also has long coattails: a degree from a prestigious school, the more ivy-covered the better, is a lifelong advantage.

Rankings are misleading, but people will insist. Even a seemingly simple, purely quantitative ranking hides biases in the choice of variables, weightings, or manipulations. At worst, statistics can be manipulated to produce almost any desired result. Yet despite their unavoidable subjectivity, university rankings are useful. Even a partly qualitative, partly quantitative ranking allows a vaguely systematic analysis of mushy things like education. And colleges should be able to tout their strong points. The above-mentioned Oberlin report carefully but appropriately crafted a credible argument for the undersung role of liberal arts colleges in educating scientists. The thoughtful statisticians selected specific fields—the basic sciences of biology, geology, physics, and chemistry—and specific colleges that played up the strengths of the liberal arts colleges. The obvious conclusion is that such carefully crafted rankings demand well-informed, skeptical consumers.

Rankings are important. They figure importantly into students' enrollment decisions. They also figure into the opinions of bureaucrats, businessmen, politicians, and other well-pocketed fund-givers, who want their names to be associated with a prestigious college. In science, federal and state patrons are always trying to rationalize their R&D and fellowship decisions. It behooves colleges being ranked to invest some effort into the art, and engineering, of rankings. □

Lisa (Cox) Heinz graduated from Caltech in 1978 with an option in biology. This article arose out of a study she recently completed at the U.S. Office of Technology Assessment, where she's employed in the communication and technologies section. Much of the data analysis she found too "entertaining" to fit in the confines of a government report, so E&S *was the beneficiary instead. She's also the Washington, D.C., chapter representative on the Alumni Association's board of directors. Although she doesn't intend to get a PhD, she thinks the quality of her Caltech education was terrific.*

The views expressed in this article are entirely those of the author and not necessarily those of OTA.