NRC Ratings of Research Doctoral Programs

Saunders Mac Lane

This note is to examine the present popular practice of providing numerical ratings of the relative standing of the various Ph.D.-granting departments in American universities. In order to be specific, I will consider the ratings of departments of mathematics, as presented in the NRC comprehensive 1995 ratings of graduate programs in mathematics, as described in an article in the *Notices of the American Mathematical Society* (vol. 42, no. 12, for December 1995, 1535–1542). The reputational ratings presented there, in math as in other fields, are averages of reports from many mathematicians, each of whom gives ratings of 50 departments (other than the rater's own), using faculty lists for those departments. The result is a listing of departments in linear order, with occasional ties.

There have been several such ratings in the past (e.g., Cartter, "An Assessment of Quality in Graduate Education", American Council of Education, 1966). Even earlier, everyone in 1900 knew that Chicago had the best math department and in 1912 that Harvard was best. In those days the American mathematical community was much smaller; as a result, active mathematicians then had a more varied acquaintance with different fields (for example, before 1940 I customarily sent reprints of my published papers to many mathematicians not active in my own field of research). Today, specialization is pervasive and necessary; many mathematicians's firsthand knowledge is limited to others working in their own field. Thus the published ratings now conflate the opinions of many individuals, variously well or poorly informed.

Saunders Mac Lane is Max Mason Distinguished Service Professor, Emeritus, The University of Chicago.

The National Research Council (NRC) is the working arm of the National Academy of Sciences. In 1980 the Conference Board of Associated Research Councils decided to sponsor new doctoral ratings. The NRC undertook to prepare these ratings—those subsequently published in 1982; I was a member of the committee responsible for that report. As vice-president of the NAS at that time, I did hope that the NRC could develop and use some new and really "objective" ratings. The AMS had been accustomed to selecting the invited hour speakers at its meetings from careful analysis of annual "Lists of Published Papers". I then thought that some index of such invitations, or otherwise based on the recognition of department members by professional societies, would be practical and objective. But it did not work out that way in fields other than mathematics, and even in mathematics any such objective index would have been quite erratic for smaller or lower-ranked departments. So the 1982 NRC report used the accustomed reputational index then called "Measure 08" as well as some other new indices, hopefully more objective. Page 24 of the 1982 volume on mathematics and physical science says:

> Measure 08…[is] subject to many of the same criticisms that have been directed at previous reputational surveys. Although care has been taken to improve the sampling design…the survey results merely reflect a consensus of faculty opinions. …
>
> These opinions may well be based on out-of-date information or be influenced by a variety of factors unrelated to the quality of the program.

> As might be expected, the smaller and less prestigious programs were not as well known, and for this reason one might have less confidence in the average ratings of these programs.

The current report states (p. 23): "Reputational ratings are influenced by a number of other factors that limit their usefulness in judging quality."

It then lists, complete with those fashionable "bullets", several such factors.

The current NRC report has indeed tried out and presented several different alternative ratings, such as the number of citations per faculty member. Such an index is of uncertain worth, since it reflects such nonquality items as the length of papers. Just for example, the number of future citations to Andrew Wiles just now doubled when his justly famous proof of Fermat's Last Theorem was published in two separate papers rather than one!

Each NRC rater assesses the "scholarly quality of program faculty" under one of six labels, from "distinguished" to "not sufficient". These words are turned into numbers and then averaged. This average is the reputational "rating" of the program; it is the measure 08 of the 1982 report. But since the numbers come in order, they produce a ranking: a complete linear order of Ph.D. granting departments of mathematics: the first, the second,… eightieth… and beyond, with occasional ties. The current NRC report, unlike the previous one, emphasizes this linear order by tabulating all the alternative indices (in table P) for universities listed in the first (reputational) order. Thus that linear order is the principal apparent product and the one most often cited in the secondary literature (e.g., in the *Notices*, in "The Chronicle of Higher Education", and in the *U.S. News & World Report*. (Yes, the secondary literature comes in order of quality.)

In 1982 the NRC report, as a matter of principle, did not give the ranked list; this was produced (with occasional error) only in the secondary literature.

Appendix Q of the current report gives "confidence intervals" for the reputational ratings. These help little; they refer only to confidence which might be engendered by getting infinitely many raters; they do not relate to confidence which might come about from the use of other means of judgment.

I submit that this linear order is meaningless. There is no such order in the real world. What there is is something much less—a partial order. Given two departments, A and B, sometimes A is better than B, sometimes B is better; but often there is no objective comparison between A and B. This partial order is not even a lattice, and the ambiguous comparisons are especially noticeable in the bottom part of the present list.

(Here is an overly simple example of a partial order. Given two measures, x and y, locate each program by a point in the plane with coordinates x and y. This order is actually a lattice; the real partial order of departments is much more complex.)

Yes, the four or five departments at the top of the present list are doubtless top departments. They are all very well known, and most active mathematicians would tend to agree that they are indeed

Adventures of a Survey Reviewer

"In our capacity as co-chairs of the Committee [for the Study of Research-Doctorate Programs in the United States], we invite you to serve as a reviewer of up to 50 programs in mathematics," read the May 4, 1993, letter. How could I refuse, even though the reviews were due no later than May 18, 1993, and I was in the midst of grading final exams. In addition to my general willingness to support important assessments like this one, I also had been one of the 348 mathematicians invited to participate in the study's 1982 predecessor—and one of the 223 who turned in their surveys. Actually, at that time I was slightly embarrassed with my responses. Since I always answer these things honestly, I was often forced to check the "little or no familiarity" box on the "Familiarity with Work of Program Faculty" and the "Don't Know Well Enough to Evaluate" box for the "Scholarly Quality of Program Faculty" for programs I knew would end up in Group I (even top 10), a reflection on my own mathematical parochialism. In the dozen intervening years (the survey for the 1982 study was done in spring 1981), nearly half of which I had spent as an associate secretary of the Society, I had hoped to have expanded my mathematical range sufficently to be able to render informed judgment on many more programs.

Alas, mathematics and the community of mathematicians had also expanded. Once again, I found myself having to select the unable-to-evaluate box for many programs, including ones which would end up at the very top. But now a new dilemma surfaced: some programs which I knew well and admired highly didn't even appear in the survey. It seems that the 1993 survey, like the 1982 one, had a grandfather clause: to be eligible for the current one, the program needed at least a certain score on the last one, or to have produced a certain number of Ph.D.s in the years 1987–92. The enforcement of the criteria was not clear to me: one of the programs I was asked to rate was not rated at all in the 1982 study and had produced only two doctorates in 1987–92; perhaps it was included from some great-grandfather clause based on the 1970 study.

I wasn't bothered by the inclusion. But troubled by the exclusions, I wrote the study co-chairs in the fall of 1993 and shared my concerns. And even though I had met their 2-week deadline in returning their survey, in the nearly two years since I have received no reply.

Andy Magid

tops, although they might change the order a bit or add one or two departments.

Is there a single best department? The report says that Berkeley and Princeton are tied for the top. Both are indeed very good. Berkeley, since about 1941, has had the best mathematical logic, while Princeton has had the best algebraic topology ever since there was such a subject (cf. Veblen, J. W. Alexander, and S. Lefschetz). But, hold on, it may not be that simple; there are now powerful topologists at Berkeley, and Princeton had outstanding logic (Alonzo Church, Kleene, Rosser, and Turing) at a time when there was precious little math at Berkeley. At Princeton, graduate students are turned out promptly; Berkeley has had many more such students, but some are intellectual orphans (in a long career, I have adopted five such Berkeley orphans). This treatment of graduate students is reflected in another index in the current report. But, again, Berkeley is the only American department of mathematics from which two women have been elected to membership in the National Academy of Sciences.

The upshot of all this is that there are all sorts of divergent comparisons. Berkeley and Princeton are not "tied"; they are, in postmodern lingo, "differently abled".

Anyhow, I meant to argue that Harvard math is better than either Princeton or Berkeley. Harvard has long been tops in algebraic geometry, a subject now of more central interest than topology or logic.

I hope you did not notice that my ratings really reflected how it was fifty years ago; I'm sure no other rater would do any such thing. I suggest an easy exercise for the reader: List the 50 departments which you know best (not those which the NRC may have handed you), and put these departments in your linear order of quality.

These details are exhibited to demonstrate the inevitable conclusion. There are indeed top departments, but there is no objective choice which is best. For less eminent departments, the real order, if any, may be much more uncertain.

Clearly, similar ambiguities in rank order occur in other academic fields. The present reputational rank order is an elaborate fashionable picture which has minimal correspondence with any real facts. These rank orders are concocted from outdated habits of thought and arranged in elaborate tabulations to please the human hunger for "information" and to nourish the prejudice that every aspect of life has its superbowl.

Therefore, such ratings are not fit for any serious use whatever. Let presidents, provosts, and deans beware. All real judgments must come from experts who are demonstrably qualified.