Research Metrics

Around the world, people are demanding accountability for research. Many governments have undertaken large-scale efforts to assess the quality of research *objectively* and many more plan such efforts in the future. Because the number of ways to measure the quality of research is limited, a handful of quantitative measures have come to dominate the assessment process. One of these, the Impact Factor, has played an especially important role.

Over the past year, mathematicians from a variety of countries have written to the International Mathematical Union asking the IMU to address the misuse of the Impact Factor in assessing the quality of mathematics research. The IMU executive committee has responded by forming a small *ad hoc* committee, jointly with the International Council on Industrial and Applied Mathematics (ICIAM) and the Institute for Mathematical Statistics (IMS). The committee will consider various ways to assess (quantitatively) research in the mathematical sciences, but in particular it will consider the widespread use (and misuse) of the Impact Factor.

We are seeking your comments about the Impact Factor. We want to document specific instances in which the Impact Factor is being used or has been misused, either by individual institutions or by larger assessment efforts. We also welcome your general comments on the Impact Factor and its use in assessing mathematics research, but we are especially interested in gaining insight about the ways in which the Impact Factor is currently used.

Please send your comments by e-mail to: research-metrics@ams.org .
To be most helpful, they should arrive by September 30, 2007.

John Ewing (IMU, chair), Peter Taylor (ICIAM), and Robert Adler (IMS).

Joint ICIAM/IMS/IMU-Committee on "Quantitative Assessment of Research" Terms of Reference 31 March 2007

The drive towards more transparency and accountability in the academic world has created a "culture of numbers" in which institutions and individuals believe that fair decisions can be reached by algorithmic evaluation of some statistical data; unable to measure quality (the ultimate goal), decision-makers replace quality by numbers that they can measure. This trend calls for comment from those who professionally "deal with numbers"-- mathematicians and statisticians.

Throughout the world, assessment of research has become increasingly important, and at the same time, of great concern. No doubt, academic achievement, research, and teaching all need to be evaluated in order to guarantee and maintain quality. Comparisons are necessary to define best practices and to set standards. Scholars judge other scholars when they referee articles, write letters of reference, or participate in regular evaluations. Administrators rate scholars when they decide about salaries or promotion. Librarians rank journals or books when they decide on which to spend their budgets. And politicians assess institutions based on many factors, including research output.

Making judgment is hard work. Everybody making judgments would like to be supported by measures that objectively describe "performance" and help to compare and rank whatever is being judged.

There is growing concern about various measures of research performance, especially those using citation data (for example, the "ISI impact factor"). Many of these measures are used in ways for which they were not designed and to make judgments that are unjustified by the data. Many of those defending the use of such measures argue that they are based on sound statistical data and employ transparent mathematical formulas, and hence they are objective. The precision of the formulas and the pretended exactness of the data, though, may only disguise the inappropriateness of the decision-making process.

The International Council of Industrial and Applied Mathematics (ICIAM), the Institute of Mathematical Statistics (IMS), and the International Mathematical Union (IMU), institutions representing the world wide communities of mathematicians and statisticians, are troubled by the possible misuse of mathematical concepts or statistical indicators. As societies representing mathematicians and statisticians, they feel a responsibility to provide a clearer understanding about the proper use of statistical data in assessing research -- especially research in the mathematical sciences.

ICIAM, IMS, and IMU therefore have established the joint committee "Quantitative Assessment of Research" and charged it with the following tasks:

1. To evaluate to what extent the ISI impact factor is a significant indicator for the quality of a researcher, a department, or similar institution in statistics or mathematics.

2. To determine to what extent the ISI impact factor can be used to compare the quality of research in mathematics with that in other disciplines.
3. To determine whether or not the ISI impact factor has any bias with respect to language, region, or length, source or field of publication, interdisciplinary work.
4. To examine these questions for a selection of other recently proposed measures based on citation data.
5. To propose suitable substitutes for these measures based on citations.
6. To list the possible dangers or advantages that the widespread use of impact factors and similar simple measures may have on publication behaviour, recruitment, balance between scientific disciplines, etc.

The committee is asked to create a summary of its finding to be endorsed and publicly distributed by the Executive Committees of ICIAM, IMS, and IMU. It is requested to support its conclusions by examples and statistical data to be provided in additional documents meant to provide evidence and a solid basis for the findings.

Academic achievement is a complicated mix of contributions to research, teaching, and supervision of students, as well as contributions to academic self organization/administration and to the scientific community. ICIAM, IMS, and IMU acknowledge that the evaluation of scientific quality is notoriously difficult, simplistic answers to complicated questions of judgment are unlikely. However, the committee is also asked to investigate whether it sees possible alternatives to measures based solely on citations that may help to evaluate research and academic achievement and indicate quality in a sensible way.

Some thoughts about the meaning of citations (sample)

What's the value of a single citation? That's a fundamental question that (as far as I can tell) not many people have thought much about. I think most people simply assume the answer is obvious: Citation implies use, and use implies importance. The assumed answer seems to underlie everything else that one does regarding citations, however, and it astounding that no one has seriously undertaken an analysis of this point.

Answers to this question may differ by discipline, I suppose, but at least in the mathematical sciences I think a citation represents one of three situations:

> i. The citing item (I'll call it the "domain item") uses the research done in the cited item (target item), building on the work in some way.
> ii. The domain item references a result by the author of the target item, which contains the first instance of the result
> iii. The domain item references a result that is contained in, or is referenced by, the target item.

Those three situations aren't precise, in the sense that they are mutually exclusive, but they seem to capture the typical ways in which authors use citations.

The most important thing to note is this: When most people think about citations (especially counts of citations), they implicitly think of citations representing situation (i) or possibly (ii) -- a citation is supposed to reflect the importance of the WORK done by the cited author and its connection to subsequent work. On the other hand, many citations are of type (iii).

It's hard to quantify the preceding statement, but you can get some information by looking at the most cited things. Math Reviews now has a citation database of nearly 3 million citations (which I'll refer to from time to time, mainly because I have access to it and therefore can get the information more easily than the ISI database). The citations come from over 300 journals, from 2000-present, and the citations refer to things that are included in the 2 million objects in the MR database -- books, journals, collections, etc. Looking at these nearly 3 million citations, one can search for the 100 most cited objects. And here is the important fact: *every one of the top 100 cited things is a book!*

This does not mean that most citations are of type (iii), of course, but it surely shows that many "highly cited" objects have that status primarily because they contain many type (iii) citations (as books do). I don't mean to diminish the importance of highly cited works like this. Gilbarg's and Trudinger's "Elliptic partial differential equations" (most cited) surely deserves to be recognized as a fundamental resource in the field, as does Harshorne's "Algebraic geometry" (second most cited). But the high citation counts here means something other than "the research contained in the work was important". And it isn't only books that attain their high citation counts through a preponderance of type (iii) citations.

High citations counts do not necessarily mean "important research"; high citations counts mean just that -- highly cited. It seems to me that this is a fundamental point one must make *before* talking about any of the other aspects of citation counting or statistics or ISI or …. Citations provide some measure of the quality of research, but at the most basic level they provide only an imperfect measure.