Measuring Journals

John Ewing

Not everything that can be counted counts, and not everything that counts can be counted.

—Albert Einstein (1879–1955), attributed

The impact factor was created in the late 1950s as a way to measure a journal's value by calculating the average number of citations per article over a specific period of time. Since citations generally reflect the interest of scholars in an article, the impact factor ought to reflect the average interest in articles appearing in a journal. This seems to be a sensible use of citations.

But the impact factor has far outgrown its original purpose. A recent article [4] on the impact factor asserts that "impact factors have assumed so much power …that they are starting to control the scientific enterprise" and that they "play a crucial role in hiring, tenure decisions, and the awarding of grants." The same article quotes Eugene Garfield, the creator of the impact factor, who laments: "We never predicted that people would turn this into an evaluation tool for giving out grants and funding."

People misuse the impact factor because there are no explicit principles governing its interpretation. The impact factor is used to measure the value of things for which it was never intended (articles and authors, for example), and it is used to make faulty comparisons between unlike objects, including journals themselves.

This is not a new problem. For decades, scholars have complained about the misuse of the impact factor, and there is an extensive literature of such complaints and admonitions.[1] But in a world gone mad with an obsession to evaluate everything "objectively", it is not surprising that desperate and sometimes incompetent evaluators use a poorly understood, but easily calculated, number to comfort them. Simple numerical scales are seductive.

The latest fad for "objectively" measuring the value of (online) journals involves usage statistics. At a recent meeting of the Society for Scholarly Publishing, there were at least a dozen presentations on usage statistics for journals. The opening slide for one of these presentations (*Project Counter*) began with a clear statement: Libraries need online usage statistics to assess the *value* of different online products and services. Almost everyone in the audience, librarians and publishers alike, appeared to nod in agreement. Speakers promoted the idea that, in the coming years, usage statistics would be the key to measuring the *value* of electronic publications—the key to making the tough decisions about journal subscriptions. And they enthusiastically endorsed "codes of practice" to make certain we could trust these new tools.[2]

But in many respects usage statistics are even more flawed than the impact factor, and once again, the essential problem is that there are no explicit principles governing their interpretation. People already misuse usage statistics, and in this case, the misuse may have adverse consequences for scholarly publishing itself.

As scholars we ought to be troubled by any misuse of statistics; as mathematicians, we ought to feel a responsibility to set things straight.

John Ewing is executive director of the AMS. His email address is jhe@ams.org.

[1] *A good list can be found at* http://www.kmutt.ac.th/jif/public_html/impact%20factor_article2005.html *(last accessed June 2006). Also, see the more recent article by Vitali Milman in the* Notices *[5].*

[2] *The most prominent such code is part of Project Counter,* http://www.projectcounter.org *(last accessed June 2006).*

Impact factor

The impact factor is computed from data gathered by the Institute for Scientific Information (ISI), which publishes the Science Citation Index. ISI indexes many thousands of scientific journals, adding information about each article and its references to its database each year. Using that information, one can determine how often a particular article is cited in subsequent articles, at least for those that are indexed by ISI. (While ISI indexes many mathematics journals, for example, it indexes far fewer than half.)

Impact factor for Journal X, 2005

A = # citations in all ISI articles during 2005 to papers published in X during 2003–2004

B = # of articles published in X during 2003–2004

Impact Factor = A/B

For a particular journal and year, the journal impact factor is computed by calculating the average number of citations to articles in the journal during the preceding two years from all articles published in that given year (and indexed by ISI). If the impact factor of a journal is 1.5 in 2005, it means that on average articles published during 2003 and 2004 were cited 1.5 times in the collection of all ISI-indexed journals published in 2005.

Sample Journals	2003 IF
New England J Med	34.8
Nature	31.0
Science	29.2
Cell	26.6
Lancet	18.3
...	...
Journal of AMS	2.5
Annals of Math	1.5
SIAM J of App Math	1.4
Inventiones	1.3
Reine Angew Math (Crelle)	0.7

Is medicine twenty times more active than mathematics?

There are some obvious problems with this measure of quality. Comparing citations in different fields, for example, may be meaningless—the traditions and cultures of citing other work (including one's own!) vary greatly between disciplines or even between sub-disciplines. Also, looking at citations for only two years after publication may produce faulty results. In some fields (mathematics is one of them) citations frequently follow several years after publication.[3] Why two years? And why choose citations from journals published in a single year? Both are somewhat arbitrary choices, without any obvious justification.

The largest problem with the impact factor, however, is that many people do not know how to properly interpret it. The impact factor is an *average* number of citations—it says nothing (directly) about how many citations there are to individual articles, any more than the average family size says something about the size of individual families.

This is an elementary caution that one learns in a first statistics course. Nonetheless, otherwise well-educated scholars and administrators continue to make the mistake. For example, the Research Assessment Exercises carried out in the United Kingdom asks researchers to provide four "research outputs" for review. University administrators routinely advise scholars to choose outputs (that is, articles) from high-impact journals. Here's the way one well-known university justifies the advice:

> …it is clear that RAE panel members will have to examine a vast number of papers in a short period. They will also have to examine papers on topics on which they are not experts. It is, therefore, inevitable, that they will use a variety of heuristic measures in assessing papers. You should be sensitive to the likely heuristic measures when both preparing and selecting your RAE submissions.[4]

Administrators in another university simply state as a fact:

> Journal Impact Factors…are used in many subject areas to determine the reputation and standing of certain journals

[3] Mathematical Reviews *now has created its own citation database of the mathematical literature, built from reference lists from 325 mathematics journals published since 2000. Using this data, it computes the Mathematical Citation Quotient (MCQ), which like the impact factor is an average number of citations to a journal over a period of time. In this case, the period is five years instead of two.* `http://www.ams.org/mrcitations/search.html` *(last accessed June 2006).*

[4] *University of Edinburgh, Research Assessment Exercise 2008, Outputs,* `http://www.inf.ed.ac.uk/admin/rae/outputs.html` *(last accessed June 2006).*

> when selecting journal papers for submission to the RAE.[5]

And still another administrator writes:

> The RAE will be looking for the international significance of the research. This may be measured, at least approximately, in terms of journal impact factor.[6]

Those who draft the guidelines for the Research Assessment Exercise point out that impact factors of journals do not (directly) measure the quality of individual articles. Indeed, they explicitly admonish review panels:

> It will not use a formal ranked list of outlets, nor impact factors, nor will it use citation indices in a formulaic way.[7]

But guidelines that talk about weighting research outputs as 70%, research environment as 20%, and research esteem as 10%, don't sound convincing when they caution scholars about the misuse of data. A demand for unreasonable precision inevitably leads to unreasonable practice.

Using the average citation count to infer information about individual papers is especially foolish in the case of scientific journals. Studies of particular journals show that the distribution of citation counts is highly skewed, with a relatively small number of articles having high counts.[8] The average, therefore, is often far higher than the number of citations for a "typical" article.

Of course, while the impact factor says little directly about the importance of individual articles, it says even less about individual authors. Yet scholars and administrators continue to cite impact factors of journals as evidence of a scholar's ability.[9]

Even as a measure of journal quality, the impact factor is misinterpreted. At best, the impact factor measures the quality of articles published in the journal rather than quality of the journal itself. This is an important distinction. It's the difference between judging a Beethoven symphony by the average quality of each movement rather than the average quality of each measure. Which is better, a journal with two long articles, each with ten citations, or a journal with twenty short articles, each with one? Not everyone will agree on the answer, but the impact factor unequivocally asserts the first is ten times better than the second! Short articles have less content than long ones, and hence they will be cited less often (on average). Is it better to publish many short articles or to publish a few long ones?

If one believes that citations measure quality, then the impact factor is surely one way to measure the quality of *individual articles* in a journal. A prospective author might want to look at the impact factor of a journal before submitting a prized paper. But if one wants to measure the quality of the *journal itself*, then it makes more sense to consider the number of citations per thousand pages (or, even more accurately, per thousand characters). This gives a measure of the scientific value per unit of content. Ranking journals by the number of citations per thousand pages often gives a different picture than ranking by impact factor.

There are other problems with the impact factor. The data is not always completely reliable; ISI makes mistakes (although these seem to be relatively minor). Authors and editors can manipulate impact factors by self-citing or publishing review articles, which guarantee higher citation rates. The formula itself is flawed because the numerator includes citations to things such as editorials and letters, which are not included in the denominator. All these are reasons to be somewhat skeptical about the precision of the impact factor. But the central problem is its misinterpretation.

Usage Statistics

Online journals present new opportunities to collect data—and also to misuse it. One can collect several kinds of data, but the most prominent are the numbers of full-text article downloads (frequently restricted only to html and pdf formats). This is the data that many librarians demand in order to judge the value of the journals they buy, so they can subscribe to only the journals that are worthwhile. As one activist librarian writes: "What good librarians should and do look at is not primarily the opinions of their users, but the actual behavior of their users."[10] His conclusion: You need usage statistics to be a good librarian.

What's wrong with measuring the value of journals by usage statistics? As for the impact factor, there are no explicit principles guiding interpretation.

[5] *Leeds University,* http://www.leeds.ac.uk/library/teams/rae/ *(last accessed June 2006).*

[6] *Kings College, London,* http://www.kcl.ac.uk/research/rae/notes.html *(last accessed June 2006).*

[7] *RAE 2008 Guidelines, p. 22,* http://www.rae.ac.uk/pubs/2005/03/ *(last accessed June 2006).*

[8] *A recent editorial in* Nature *pointed out that 89% of the 2004 citations were to just 25% of the articles [6]. Analysis of mathematics journals suggests that about 25% of the articles typically account for 75% of the citations.*

[9] *For example, in Italy, impact factors were suggested as a way to eliminate subjective judgments when making certain academic appointments [2].*

[10] *David Goodman, Palmer School of Library and Information Science, Long Island University,* http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/1446.html *(last accessed June 2006).*

Should we judge the value of a journal by how often users download articles? Over what period of time? Is an article that is downloaded ten times in its first year more valuable than an article downloaded one hundred times in the first fifty? Articles play many different roles in research—to generate ideas, for background reading, or to supply a crucial piece of the puzzle. Which of these purposes is more important? Is a review article that is casually perused by hundreds of readers more important than an article that leads to a stunning breakthrough by one researcher? Equating value with usage statistics ignores such questions and suggests there is a simple way to measure value—a few numbers that make it easy to compare journals.

The term "usage statistics" is itself misleading. Do these statistics really measure usage? Think of your own browsing habits and ask whether you "use" all the items on which you click. Browsing the Web is designed to make it easy to casually download much more than you actually use. Many users of online services download large numbers of files as a matter of routine (often automatically by script). If users download files but don't read them, are they "using" them? How often does this happen? Is it increasing? We have no way of answering these questions, and hence no way of knowing whether "usage statistics" really measure "usage" in its common meaning.

Whatever the meaning of usage statistics, we know that they are notoriously unreliable. The Web is deceptive because it creates the appearance of a user logging into a session and using a resource, but that's not the way it works.[11] Calling each journal page or file to your computer requires a series of messages to be sent and received. Sometimes the messages are exchanged with the publisher's server (recording the usage), but sometimes they are not (when a page or file has been cached on some site in between).[12] Sometimes the series of messages gets interrupted because there is too much traffic or some other problem on the network, and the exchange begins again. And sometimes there are flaws in the browser software, causing requests to be sent repeatedly (the "three-times" flaw for PDF files using Internet Explorer is a famous example).[13] Codes of practice try to get around these technical problems by creating precise rules for how to count, but reading the details of such codes only reinforces the notion that usage statistics are imprecise. Many experts agree.

[11] *For an excellent description of the way in which the Web works and the limitations of statistics, see* `http://www.analog.cx/docs/webworks.html` *(last accessed June 2006).*

[12] *For caching and a discussion of other problems with statistics, see [3].*

[13] *See Microsoft Knowledge Base article Q293792.*

The misuse of the impact factor has changed the practices of a few journal publishers. Already some unscrupulous publishers have tried to boost their impact factors by adding self-citations or publishing more review articles. Usage statistics threaten to foster far worse practices among publishers, even those with principles.

Many online journals provide ample material to users, whether or not they are subscribers. These include bibliographic information (title, author, etc.), abstract, and (sometimes) the references with links. This makes it easy for scholars to navigate the literature, even when their institutions do not subscribe to all journals—a kind of open access within the existing subscription system of journals. Allowing people to see this material without "downloading" the article, however, will adversely affect the journal's usage statistics. Surely if usage statistics are equated with "value", publishers will change their practice in order to protect their journals.

Increasingly, publishers have instituted liberal copyright policies that permit authors to post their papers online, either on a homepage or on a preprint server. This too is a kind of open access that coexists with subscription journals. But there is evidence that making articles freely available in this way reduces (current) downloading. If publishers know that the value of their journal will be measured by the number of downloads, they will soon change this practice too. Publishers will close off their journals in order to protect them. Do we want that?

All these problems arise from a few numbers that are created from a flawed perception of how the Web really works, some subjective rules about how to compile them, and almost no advice about how to interpret them.

Alternatives

What's the conclusion? Surely not that the impact factor is useless—it's not. Citations provide some information about the value of articles and journals. But we should regard the impact factor as a way to measure the average quality of articles within a journal and nothing more. We should remember that measuring the quality of each article or even the entire journal itself requires much more information.

And surely not that usage statistics are completely useless—they have some use. But usage statistics are far less precise than proponents admit, and they convey only a small amount of information about a journal. Usage statistics alone do not measure value. And while usage statistics are only *slightly* useful, their misuse can be *enormously* damaging. They have the potential to roll back many of the progressive access policies of recent years.

The main conclusion is that we must stop seeking simplistic answers to complicated questions of judgment. The impact factor gives some information, but so do other statistics. (Some have advocated page rank, similar to the Google method of ranking Web pages [1].) Or one might compute the average number of citations per thousand pages or look at the number of citations to various year-ranges of the journal or examine the top 10% of heavily-cited papers. To judge articles, one might look at the number of citations to the article itself, again over various ranges. And to all these metrics of quality, one must add personal judgment: Scholars, rather than numbers, are frequently the best judges of quality. Distrust of "subjective" scholarly judgment is a modern disease—one that is profoundly anti-intellectual.

To evaluate scholars themselves, citations can play a role. But why try to capture this information by looking at complicated computations or indirect measurements like the impact factor? Simply examine the number of citations to the top papers of the person.[14] And there are many other ways to measure the value of a scholar's research in addition to citations. Once again, scholarly judgment is often the best measure of achievement.

How should librarians and scholars make tough decisions about journal subscriptions? They can revert to the time-honored method for assessing the value of any product: Ask the people who use it. This is imprecise and subjective; it is frustrating and time-consuming; it is not always easy to balance conflicting advice. But substituting a nonsensical number that bears little relation to the value of the journal, and that likely will promote regressive policies among publishers, surely isn't a way to solve these problems.

Making judgments is hard work. For many of us, it's a responsibility that accompanies our jobs. Scholars, administrators, and librarians need to accept that responsibility.

References

[1] JOHAN BOLLEN, MARKO A. RODRIGUEZ, and HERBERT VAN DE SOMPEL, Journal Status, (http://www.arxiv.org/abs/cs.GL/0601030 : Last accessed June 2006).

[2] L. CALZA and S. GARBISA, Italian professorships, *Nature* **374** (1995), pp. 492–.

[3] JEFFREY GOLDBERG, Why Web usage statistics are (worse than) meaningless, (http://www.goldmark.org/netrants/webstats/ : Last accessed June 2006).

[4] RICHARD MONASTERSKY, The number that's devouring science, *The Chronicle of Higher Education*, October 14, 2005, (http://chronicle.com/weekly/v52/i08/08a01201.htm: Last accessed June 2006).

[5] VITALI MILMAN, Impact factor and how it relates to quality of journals, *Notices of AMS* **53**(3) (2006), pp. 351–352, (http://www.ams.org/notices/200603/ comm-milman.pdf : Last accessed June 2006).

[6] *Nature*, **435** (June 2005), pp. 1003–1004.

[7] PER O. SEGLEN, Why the impact factor of journals should not be used for evaluating research, *BMJ* **314** (1997), pp. 497–, (http://bmj.bmjjournals.com/cgi/content/full/314/7079/497: Last accessed June 2006).

[14] *The citation database of* Mathematical Reviews *does this for each mathematician, listing the number of citations for the top ten most highly cited papers. It also lists the total number of citations and the total number of authors citing.*