The problematic ratings game in modern science

THIS IS AN AGE OF QUANTITATION AND standardization. Driven by the inexorable logic of a technological culture, trains and aeroplanes must run on time, financial movements tracked and minutely analysed, claimed remedies need prove themselves in the fire of double-blind, prospective, randomized clinical trials (and then again in meta-analyses), and personal, scholarly accomplishment must be measured out, if not in coffee spoons, then in impact factors and citation rates.

In fact, the last item has become the basis for a major industry. Journals must prove their worth to readers, authors and advertisers. And, since governments and other funders of the knowledge sector require the assurance that they're getting the biggest bang for their buck, institutions and individuals are evaluated and graded against hard quantitative data derived from citation counts and 'impact factors', which, supposedly, provide the certitude they require. To determine whether such confidence is justified, we need briefly review the background to these developments and the potential limitations in the information they claim to provide.

In the knowledge industry, the response of peers is often the ultimate coin of the realm. Citations to the work of colleagues is increasingly used as a formal, and potentially more reliable, proxy for directly expressed opinion. The use of citations to understand the structure of science goes back at least 75 years,[1] but the main breakthrough came with the publication of a paper[2] on citation analysis in *Science* by Eugene Garfield in 1955. Driven by the explosion in computing power, this seed germinated over the succeeding decades[3,4] into a commercial enterprise, the Institute for Scientific Information (ISI®), which publishes *Current Contents* and the *Journal of Citation Reports* (*JCR*®)[5] relating to the scientific and scholarly literature. ISI was bought out by the Thomson Corporation in 1992 to form an important component of Thomson Scientific, which provides a range of products, on a commercial basis, dealing with information of scientific and commercial significance.[6]

Those engaged directly or indirectly in the practice of science are best acquainted with the products of Essential Science Indicators, a web-based tool which mines a vast database of journals and other products of scientific and scholarly activity for information, which can be used to evaluate the productivity and impact of institutions, journals, departments, groups and individuals. Such data, reduced to numbers, are widely used for purposes of evaluation and support,[7,8] and thus carry enormous implications for individuals and collectives. Is this a new form of arbitrary tyranny or do these numbers provide meaningful information to guide vital decisions? To answer such a question, we need briefly examine some of the indices currently used (Table 1) and the criticisms which have been levelled at their quality and implications.

The most widely used is the 'impact factor' (IF) of the ISI system, defined as the ratio between original, peer-reviewed papers published in a journal in the prior two years (the denominator) and the number of citations to these papers in the broad scientific or scholarly, indexed literature (including forms of publication other than original articles, such as reviews, commentaries, editorials, etc.) in the succeeding year (the numerator). This basic idea can be manipulated in various ways to apply to individuals, institutions, groups and even countries. Also, since both author and journal self-citation can be used to inflate impact factors, the IF can be calculated with either or both of these self-citations excluded. Citation records can be mined for other information. For example, the 'immediacy index' relates citations in the current year to articles in the same year; this indicates how topical ('hot') the journal's contents are perceived to be, or may say something about the discipline within which the journal publishes (see below). Half-lives of cited journals (that is, the time in years that includes 50% of all citations in the literature to the journal in question, starting with current date) or the 5-year IF, all add further dimensions, as does the trend graph of IFs (Table 1).

This message has been taken to new levels of refinement by Hirsch, a physicist, who devised the *h*-factor to capture both productivity and quality in a single number.[9] He defines *h* as the number of articles published by a given scientist which elicits at least *h* citations. This immediately indicates that as the number of articles with a high citation count from a given scientist increases, so does the threshold for inclusion. Thus, for an *h*-factor of 40 papers, each must collect at least 40 citations; for an *h*-factor of 60 papers, 60 citations each are required. Needless to say, this requires time and wide recognition, but enables scientists of the same 'scientific age' to be compared with one another through the use of a formula like $h \approx mn$, where m is the number of *h*-papers published by a scientist per year and n is the number of years a scientist has been working. The larger the value of m, the higher the visibility of the scientist.

Hirsch suggests some thresholds: an m of

Table 1. A selection of the indices provided by the *JCR*®.*

Impact factor (and 5-year impact factor)
Immediacy Index
Cited Half-life
Citing Half-life
Cited Journal List
Citing Journal List
Related Journals
Impact Factor Trend Graph

*From Thomson '*Journal Citation Reports*® on the Web v.4.0'.

1 denotes a successful scientist; an *m* of 2 an outstanding scientist, whereas 3 characterizes 'truly unique individuals'. A study of Nobel prizewinners in physics over the past 20 years, however, showed *m* to vary between 0.47 and 2.19 and *h* to fall between 22 and 79. Partly, the low *m* is an artefact created by the long time lapse between the period of maximum productivity and receipt of the prize, but it does raise some interesting questions regarding the evaluation of scientific merit.

In assessing the significance of all such derived indices, it is fundamentally important to know that each scientific field (discipline) has its own characteristic citation behaviour—that is, the number of citations authors tend to use, the importance attributed to novelty and the size of the universe of citation options, which is, roughly, proportional to the number of journals active in a field *and* indexed in the ISI system. The distributions are also highly skewed. For example, a high IF may be earned by a minority of articles published in a particular journal, the rest being unexceptional. Topic reviews are typically large earners of citations, so journals including such reviews in their standard fare rate higher than others that do not. There is evidence that publication in a high-impact journal itself enhances the citation score of individual articles (and thus authors): Stanbrook, quoted by Monastersky,[8] showed that the same item published simultaneously in 12 different journals generated a 100-fold difference in citation rates between the highest- and lowest-ranked journals.

Such limitations, and other real or potential defects in the method and the principles behind the calculations in the current practice of citation analysis, have been documented in a series of papers. [7–11] These are summarizad in Table 2, drawn mainly from ref. 10. The errors introduced are far from trivial. Retracted reports continue to draw hundreds of citations, and language, regional and institutional biases are very real. Dishonest and unethical behaviour by editors and authors alike have been recorded, and, even when falling short of such transgressions, inappropriate reliance on the IF can introduce distortions into the practice of science and cause individual injustice. Additional complexity is introduced when one considers non-traditional contributions to science, like software programs, or informal contributions falling short of authorship.[11]

Table 2. Real and potential biases in the impact factor.*

- Coverage and language preference of the database
- The procedures used to collect citations
- The algorithm used to calculate the IF
- Citation distribution of journals
- Online availability of publications
- Citations to invalid articles
- Negative citations
- Editorial selectivity according to type of article, language, region and institution
- Publication lag
- Citing behaviour across disciplines
- Unethical biases introduced by editors

*Modified from ref. 10 mainly.

It is sobering to consider that more than half of the 5968 science journals included in the ISI of 2004 had IFs below one.[8] This includes 17 of the 18 of the South African journals covered by the *JCR*®. The *South African Journal of Science*, with an IF of 0.93 and the highest category ranking (2nd quintile) of all South African journals, fared well in this assessment.[12] Not surprisingly, *Nature* and *Science* have IFs around 30–32 and lead the field of non-review journals in all 11 disciplines examined by *ScienceWatch*[13] for the period January 1999 to August 2004. Given that the citation records of South African scholars (the individual scores of articles published, and the IFs of the journals in which they appeared) are used for such career-vital decisions as tenure, grants and recognition, the message is a no-brainer for the ambitious scientist or scholar plying his or her trade in the global marketplace of science: aim for the top journals, by hook or by crook, as measured by their impact factors, even if these are mostly grossly misused in terms of the caveats mentioned above.

Within the field of scientometrics, a number of alternative measures have been devised but are not widely known or used.[10] Thomson's ISI has been slow to introduce procedures which limit known biases. It seems highly desirable that, in a developing country like South Africa, one of the recommendations in the recent report on 'A Strategic Approach to Research Publishing in South Africa'[1] should be implemented: namely, that an independent, even if virtual, national centre with the necessary expertise and research orientation be established, available to all those entrusted with the responsibility of decision with regard to funding or the recognition of research productivity and quality. While perfect assessments can never be reached, the use of optimized, objective methods is the main counterweight to intrusion of old-boy networks, political influence and the subjective opinions of friends, enemies and competitors. Such methods may have yielded a rough justice when the scientific and scholarly universe was limited, but in today's crowded academic space we depend more heavily on the neutrality of objective indices. They had better be the best available.

The helpful comments of Wieland Gevers are gratefully acknowledged.

Michael Berger

1. Gross P.L.K. and Gross E.M. (1927). College libraries and chemical education. *Science* **66**, 385–389, quoted in *The ISI Impact Factor.* Online: http://scientific.thomson.com/free/essays/journalcitationreports/impactfactors/
2. Garfield E. (1955). Citation indices for science: a new dimension for documentation through association of ideas. *Science* **122**, 108–111.
3. Garfield E. (1972). Citation analysis as a tool in journal evaluation. *Science* **178**, 471–479.
4. Garfield E. (1986). Which medical journals have the greatest impact? *Ann. Int. Med.* **105**, 313–320.
5. Online: http://scientific.thomson.com/products/jcr/
6. Online: http://scientific.thomson.com/products/
7. Wikipedia, *Impact Factor.* Online: http://en.wikipedia.org/wiki/Impact_Factor
8. Monastersky R. (2005). The number that's devouring science. *The Chronicle of Higher Education.* Online: http://chronicle.com/free/v52/i08/08a01201.htm
9. Hirsch J.E. (2005). An index to quantify an individual's scientific research output. *Proc. Natl Acad. Sci. USA* **102**, 16569–16572.
10. Dong P., Loh M. and Mandry A. (2005). The "impact factor" revisited. *Biomedical Digital Libraries.* Online: http://www.bio-diglib.com/content/2/1/7
11. Perkel J.M. (2005). The future of citation analysis: the challenge is to track a work's impact when published in nontraditional forms. *The Scientist.* Online: http://www.garfield.library.upenn.edu/papers/futureofcitationanalysists102405.pdf
12. Pouris A. ((2004). Evaluating the impact and visibility of the *South African Journal of Science. S. Afr. J. Sci.* **100**, 515–517.
13. *ScienceWatch* (2005). Hottest journals of the millennium (so far). Online: http://www.sciencewatch.com/jan-feb2005/sw_jan-feb2005_page1.htm
14. Academy of Science of South Africa (2006). *A Strategic Approach to Research Publishing in South Africa.* Online: http:// www.assaf.co.za/reports/