My Word

The mismeasurement of science

Peter A. Lawrence

Answer from the hero in Leo Szilard's 1948 story "The Mark Gable Foundation" when asked by a wealthy entrepreneur who believes that science has progressed too quickly, what he should do to retard this progress: "You could set up a foundation with an annual endowment of thirty million dollars. Research workers in need of funds could apply for grants, if they could make a convincing case. Have ten committees, each composed of twelve scientists, appointed to pass on these applications. Take the most active scientists out of the laboratory and make them members of these committees. ...First of all, the best scientists would be removed from their laboratories and kept busy on committees passing on applications for funds. Secondly the scientific workers in need of funds would concentrate on problems which were considered promising and were pretty certain to lead to publishable results. ...By going after the obvious, pretty soon science would dry out. Science would become something like a parlor game. ...There would be fashions. Those who followed the fashions would get grants. Those who wouldn't would not."

There is another kind of justice than the justice of number....... There is a justice of newborn worlds which cannot be counted. [1]

It is fun to imagine song writers being assessed in the way that scientists are today. Bureaucrats employed by DAFTA (Ditty, Aria, Fugue and Toccata Assessment) would count the number of songs produced and rank them by which radio stations they were played on during the first two weeks after release. The song writers would soon find that producing junky Christmas tunes and cosying up to DJs from top radio stations advanced their careers more than composing proper music. It is not so funny that, in the real world of science, dodgy evaluation criteria such as impact factors and citations are dominating minds, distorting behaviour and determining careers.

Modern science, particularly biomedicine, is being damaged by attempts to measure the quantity and quality of research. Scientists are ranked according to these measures, a ranking that impacts on funding of grants, competition for posts and promotion. The measures seemed, at first rather harmless, but, like cuckoos in a nest, they have grown into monsters that threaten science itself. Already, they have produced an "audit society" [2] in which scientists aim, and indeed are forced, to put meeting the measures above trying to understand nature and disease.

The journals are evaluated according to impact factors, and scientists and departments assessed according to the impact factors of the journals they publish in. Consequently, over the last twenty years a scientist's primary aim has been downgraded from doing science to producing papers and contriving to get them into the "best" journals they can [3]. Now there is a new trend: the idea is to rank scientists by the numbers of citations their papers receive. Consequently, I predict that citation-fishing and citation-bartering will become major pursuits.

Impact factors and citations

The impact factor of a journal reflects the number of times the average article is cited in the two years following publication [4,5]. Of course, there is some correlation between the quality of the work being assessed and the impact factor of the journal that publishes it, but, even so, there are many faults with this measure [6]. Note particularly that it is not the impact factor of your paper that is being added up, it is that of the journal [7]. Your paper may prove to be wrong and to have diverted and wasted the efforts of hundreds of scientists, but it will still look good on your CV and may land you a job. Truly original work usually takes longer than two years to be appreciated — the most important paper in biology of the 20th century was cited rarely for the first ten years [8]. An article by Ed Lewis [9], the keystone of the work that won him the Nobel Prize in 1995, was quoted little in the first two years and took six more to reach its peak rate of citations.

Crucially, impact factors are distorted by positive feedback — many citations are not based on reading the paper but by reading other papers, particularly reviews. One study even suggested that, of cited articles, only some 20% had actually been read [10]. Consider the 48 citations of one of our articles [11], only eight of which are appropriate to what is actually reported, leaving 40 that are either plain wrong (3) or incidental (37) — for these 37 a different article either could or should have been cited. Thus it may be that, in general, citations are determined more by visibility and convenience than by the content or quality of the work.

Nevertheless, citations are now being used to make quantitative comparisons between scientists. The 'H-index' [5] measures the total number of papers a scientist has authored and the number of citations those papers have received. If, over a lifetime of research, you have authored 50 papers that have been cited 50 or more times, then your H-index is 50. At least this measure relates to the papers themselves, which may be cited often and over many years, even if they were published in journals with low impact factors.

The use of both the H-index and impact factors to evaluate scientists has increased unethical behaviour: it rewards those who gatecrash their names on to author lists. This is very common, even standard, with many people authoring papers whose contents they are largely a stranger to. There are several methods;

perhaps by providing a reagent, by some kind of horse-trading between group leaders, or by the misuse of authority or power — often involving the exploitation of young scientists [12], for example, when group leaders place their names on projects initiated, executed and written up by junior members of their groups. This practice can strike back when there is a case of fraud or error, but then the senior scientists tend to claim that they were too distant from the experiments to be able to detect the danger. Note, however, that science prizes are rarely, if ever, refused because of a similar distance from the key experiments!

Changes in behaviour

Unfortunately, the use of these measures is having damaging effects on perceptions and on behaviour; these I list below. Please note that I am not saying that all science has gone rotten, I am describing trends and extremes; there are many principled researchers and teachers, but they are having an increasingly difficult time.

First, there is the nature of scientific reporting. It has become vital to get papers into high impact-factor journals; just one such paper can change the prospects of a postdoc from nonexistent to substantial (because of the weight put on such papers by grant-awarding bodies). Two or three such papers can make the difference between unemployment and tenure. These facts have cut a swathe through scientific thinking like a forest fire, turning our thoughts and efforts away from scientific problems and solutions, and towards the process of submission, reviewing and publication. Grisly stories of papers that have been bounced down a cascade of journals from high impact factor to lower and lower ones are now the main dish of scientific discourse. It is not unusual for a scientist to spend as much as a year trying to get a paper first past editors and then reviewers, and if rejected, recrafting the paper to get round the more trenchant criticisms, writing tortuously argued rebuttals, and then hounding editors to find a more sympathetic reviewer. If these tactics fail with one journal, they doggedly re-enter battle with the next. This is a massive waste of time and energy that, even so, can bring career rewards. Therefore, I would like the granting agencies to investigate the time and effort leaders of the groups that they fund are spending on this paper chase — for these agencies are largely responsible for it. Would it not make more sense if, from the beginning, a paper were sent to a journal that was likely to accept it? The idea that one should treat publication as some kind of all-comers boxing challenge is relatively recent.

Second, trying to meet the measures involves changing research strategy: risks should not be taken as this can mean long periods trying out new things, good for the originality of research but bad if a grant has to be renewed. Follow fashion and work in crowded halls — in there you can at least count on being noticed, whereas if you venture into the unknown you risk interesting no one. A paper in a new subject will ring no bells in the heads of editors and they know, if they select the paper for review, it may be difficult to find referees. Link or pretend to link your work to medicine, as the huge medical literature can yield many citations for any paper published in a prominent journal (editors are not unaware of this as one of their remits is to increase the impact factor of their own journal). For the same reasons, choose the most popular species; it may be easier to publish unsound but trendy work on humans than an incisive study on a zebrafish.

Third, there is the presentation of the results: hype your work, slice the findings up as much as possible (four papers good, two papers bad), compress the results (most top journals have little space, a typical *Nature* letter now has the density of a black hole), simplify your conclusions but complexify the material (more difficult for reviewers to fault it!), mine rich sources of data, even if they lack originality. Most damagingly, I find it has become profitable to ignore or hide results that do not fit with the story being sold — a mix of evidence tends to make a paper look messy and lower its appeal.

Fourth, there is the way that science is done and papers are authored. These measures are pushing people into having larger groups. It is a simple matter of arithmetic. Since the group leader authors all the papers, the more people, the more papers. If a larger proportion of young scientists in a larger group fail, as I suspect, this is not recorded. And because no account is taken of wasted lives and broken dreams, these failures do not make a group leader look less productive. I believe the need to maximise publications is also causing students to be treated more like technicians. Thus, increasingly, students are told what to do, so they miss out on learning how to become researchers. Also, group leaders often write up their students' work: if the writer is not fully aware of the results or exactly how they were obtained this can conveniently permit a more adventurous interpretation of the data. Ignorance can be turned into bliss if the outcome is publication in a higher impact journal. The downside is that papers may be less truthful, not something that is easily measured. Certainly, the student does not get a proper education and may not find his or her PhD period enjoyable — however, even this can be advantageous for the group leader, as that student is less likely to go on and become a competitor!

Fifth, leaving your lab to network is being rewarded; it can pay to build tacit webs of mutual support amongst colleagues, some of whom will review your papers. Attending many meetings, giving lectures and networking can raise your profile in the minds of editors (who frequently attend these meetings) and who may therefore be more likely to send your paper out to review, quantitatively their most important decision. (In the top journals 90–95% of submitted manuscripts are rejected by editors but 30–50% of reviewed papers are eventually accepted). It is no wonder that some of our most successful scientists spend bizarre amounts of their time touring.

Sixth, there is an unexpected effect [13]. The struggle to survive in modern science, the open and public nature of that competition, and the advantages bestowed on those who are prepared to show off and to exploit others have acted against modest and gentle people of all kinds — yet there is no evidence, presumption or likelihood that less pushy people are less creative. As less aggressive people are predominantly women [14,15] it should be no surprise that, in spite of an increased proportion of women entering biomedical research as students, there has been little, if any, increase in the representation of women at the top [16]. Gentle people of both sexes vote with their feet and leave a profession that they, correctly, perceive to discriminate against them [17]. Not only do we lose many original researchers, I think science would flourish more in an understanding and empathetic workplace.

Who is to blame and what to do?

The main villains are fashion, the management cult and the politics of our time, all of which favour numerical evaluation of ‘performance’ and reward compliance. Over recent years, within governments and outside them, people have lost sight of the primary purposes of institutions, and a growing obsession with internal processes has driven more and more bureaucracy — such as increasingly complex grant applications and baroque research assessment exercises — at the expense of research effort. This bureaucracy is placing heavy demands on scientists that lay waste their sense of purpose and attack their self-esteem. But scientists of all ranks, senior as well as junior, are also to blame as we have meekly allowed this to happen. But can we now start to fight back? We need to raise awareness of the problems and make changes locally. For example, appointment committees need to keep uppermost in their minds that they are not hiring a number but a person with a mix of abilities, of which, in research, originality is the most important. I am not alone in thinking that originality is not measured by the impact factors of journals or past citations to the author (see http://voltaire.members.beeb.net/goodscience.htm). Some centralised measures may be helpful for comparison across a country, or even worldwide but they should not be overvalued, as they are now. For hiring, promotion and tenure one should find out more about the candidate, for example by reading their work, or listening to their lecture.

To improve the standard of behaviour, I think we need a code of ethics as well as some means to enforce it, especially with regard to publication — at the moment people are not agreed on what is good or even acceptable practice, for example, in authorship. The crucial problem is, I think, annexation of intellectual property, which is often sanctioned by the scientific community and actually promoted by granting agencies (for example, I did not author my graduate student’s paper in a high impact journal and, as a result, my University cannot use that paper in its Research Assessment Exercise). A public discussion on what justifies authorship would therefore be a good start. There are also difficulties with the assessment of manuscripts: anonymous referees who murder papers for gain, who take advantage of privileged information or who share the contents of reviewed papers with others should be held to account. Authors who feel they have been cheated of opportunity have no resource but to try another journal, even though the delay may prove fatal for the paper and kill their prospects of employment or support. What can an author do for redress, if they feel a reviewer or a journal mishandles their paper? At the moment, a journal can accept a paper, ask an author to spend weeks revising it, fix a publication date, send proofs and then pull the paper (it happened to me with *Science*). One possible approach might be for the large granting agencies to set up an Ombudsman, to whom those who feel they have been wronged by maladministration could appeal. It is time to help the pendulum of power swing back to favour the person who actually works at the bench and tries to discover things.

References

1. Merton, T. (1966). Raids on the Unspeakable (New York, NY: New Directions Publishing Corporation).
2. Power, M. (1997). The Audit Society: Rituals of Verification (Oxford, UK: Oxford University Press).
3. Lawrence, P.A. (2003). The politics of publication. Nature *422*, 259–261.
4. Garfield, E. (1996). How can impact factors be improved? Brit. Med. J. *313*, 411–413.
5. Hirsch, J.E. (2005). An index to quantify an individual’s scientific research output. Proc. Natl. Acad. Sci. USA *102*, 16569–16572.
6. Seglen, P.O. (1997). Why the impact factor of journals should not be used for evaluating research. Brit. Med. J. *314*, 497–511.
7. Editorial (2005). Not-so-deep impact. Nature *435*, 1003–1004.
8. Olby, R. (2003). Quiet debut for the double helix. Nature *421*, 402–405.
9. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. Nature *276*, 565–570.
10. Simkin, M.V., and Roychowdhury, V.P. (2003). Read before you cite! Complex Syst. *14*, 269–274.
11. Casal, J., Struhl, G., and Lawrence, P.A. (2002). Developmental compartments and planar polarity in *Drosophila*. Curr. Biol. *12*, 1189–1198.
12. Lawrence, P.A. (2002). Rank injustice. Nature *415*, 835–836.
13. Lawrence, P.A. (2006). Men, women, and ghosts in science. PLoS Biol. *4*, 13–15.
14. Geary, D.C. (1998). Male, Female: The Evolution of Human Sex Differences (Washington, DC: American Psychological Association).
15. Baron-Cohen, S. (2003). The Essential Difference. Men, Women and the Extreme Male Brain. (London: Allen Lane).
16. European Technology Assessment Network Working Group on Women and Science (2000). Science policies in the European Union: Promoting excellence through mainstreaming gender equality, European Commission: Brussels.
17. Babcock, S., and Laschever, S. (2004). Women don’t ask: Negotiation and the Gender Divide (Hoboken, New Jersey: Princeton University Press), p. 240.

Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK, and MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.
Email: pal@mrc-lmb.cam.ac.uk