

This is a longer version of comments published in the Times Higher Education Supplement, June 1, 2007.,

It is on the web at <http://www.dscience.net/goodscience/?p=4>

I should make it clear that the term 'bean counter' is not aimed at accountants (we need good honest accountants). Rather it is aimed at a small number of senior academics and HR people who do not understand how to assess people.

How to get good science

David Colquhoun, Department of Pharmacology, University College London (May 2007). [email me](#)

The aim of this article is to consider how a university can achieve the best research and teaching, and the most efficient administration.

My aims, in other words, are exactly the same as every university vice-chancellor (president/rector/provost) in the country.

Academics, like everyone else, are expected to do a good job. They are paid largely by taxpayers, and taxpayers have every right to demand value for their money. The problem is that it is very hard to measure the value of their output. Most of the ideas that have made life as comfortable as it is in the affluent West have their origins in science departments in universities, but it isn't possible to place a monetary value on, say, [James Clerk Maxwell](#)'s equations of electricity and magnetism, or on [Bernard Katz](#)'s work on synaptic transmission, Still less is it possible to measure the contributions of A. E. Housman, Stanley Spencer or Augustus John (all UCL people, as it happens).

This paper describes one example of what happens when universities change from being run by academics to being run by managers. It describes an effect of corporatisation in the medical school of Imperial College London, but the same trends are visible in universities throughout the world. The documents on which it is based were sent to me after I'd written "All of us who do research (rather than talk about it) know the disastrous effects that the Research Assessment Exercise (RAE) has had on research in the United Kingdom: short-termism, intellectual shallowness, guest authorships and even dishonesty" (Colquhoun, 2007). The problem is not so much the RAE itself (the last one was done much better than the assessment described below), but rather it is the effect that the RAE has had on university managers, who try to shape the whole university in their misperception about its methods. It is another example of [Goodhart's law](#). The problem arises when people with little understanding of scholarship, or of statistics, attempt to measure numerically things that cannot be so

measured. That is a plague of our age ([Colquhoun, 2006](#)), but it is a process loved by politicians, “human resources” people and university managers.

Imagine how you would feel if you were sent every year a spreadsheet that showed your publication score and financial viability, and showed these things for all your colleagues too. Well, you may say, there’s nothing wrong with knowing how you are doing. But imagine too that your publication score is entirely automated, with no attempt to measure the quality of what you are doing. And imagine that if your grants don’t cover your costs, you are in danger of being fired. And imagine that your meetings with senior colleagues consist of harassment about what journals you publish in, and how many grants you have, not a discussion of your scientific aims. Not so good, you may think. But this is exactly what has been happening at Imperial College Medical School.

Let’s take a closer look at how academics are being assessed.

Imperial’s “publication score”

The publication score that appears alongside that of your colleagues is calculated thus.

Multiply the impact factor of the journal by the author position weight, and divide by the number of authors. The author position weight is 5 for the first and last author, 3 for the second author, 2 for the third author and 1 for any other position.

This index is clearly the invention of an uninformed incompetent. That is obvious for a start because it uses the impact factor. The impact factor is a (superfluous) way of comparing journals. It is the invention of Eugene Garfield, a man who has done enormous harm to true science. But even Garfield has said

“In order to shortcut the work of looking up actual (real) citation counts for investigators the journal impact factor is used as a surrogate to estimate the count. I have always warned against this use”. [Garfield \(1998\)](#)

Garfield still hasn’t understood though. As the examples below show, the citation rate is itself a very dubious measure of quality. Garfield quotes approvingly

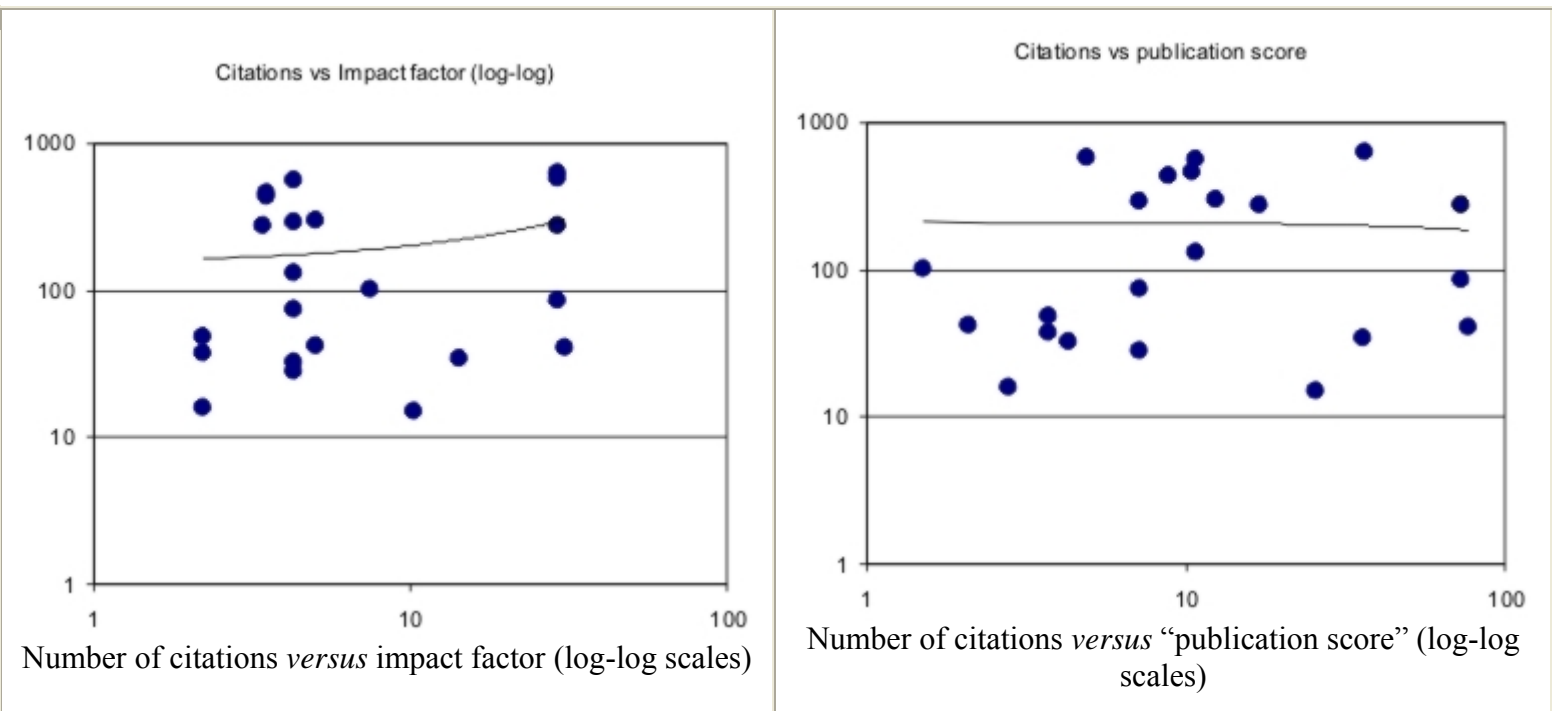
“Impact factor is not a perfect tool to measure the quality of articles, but there is nothing better, and it has the advantage of already being in existence and is therefore a good technique for scientific evaluation.” [\(Hoeffel, 1998\)](#)

And you can’t get much dumber than that. It is a “good technique” because it is already in existence? There *is* something better. Read the papers.

Try asking an impact factor enthusiast why it matters that the distribution of citation numbers for a given journal is highly skewed, and you will usually be met with a blank stare. One effect of the skew is that there is no detectable correlation between impact factor and citation rate (see, for example, [\(Seglen, 1997; Colquhoun, 2003\)](#)). The easiest way to illustrate the numb-skulled nature of this assessment is with a few examples.

Publication score versus citation

Take a selection of 22 my own publications (the selection is arbitrary: it spans a range from 15 to 630 citations and omits some of the dross). Figure 1A shows that the well-known lack of correlation between citations and impact factor is true for me too. Figure 1B shows the same for the publication score.



The highest publication score (77.3) was for a two page perspective in *Science*, with a mere 41 citations ([Sivilotti & Colquhoun, 1995](#)). As perspectives go, it was fine. But it seems that this was 7.2 times more valuable than my best ever paper (on which I was recently asked to write a classical perspective) which has a publication score of only 10.7 (but 565 citations) ([Colquhoun & Sakmann, 1985](#)). My lowest publication score (in this selection) is 2.08. That is for a [Hawkes et al., \(1992\)](#), a mathematical paper which provides the method needed for maximum likelihood fitting of single channel recordings, without which most of my experimental work could not have been done; its mathematical difficulty may account for its modest number of citations (42) but its value for our work has been enormous after the maths was put into a computer program that can be used by the semi-numerate.

Citations versus value: a real life story

The dimwitted nature of the publication score, and also of using citation rates, can be illustrated in another way. Consider some of the background to a couple of examples; these are the real life facts that are ignored by bean counters.

Colquhoun & Sakmann (1981) got a score of 73.2 and 278 citations. It was a 3 page *Nature* letter, a first stab at interpretation of the fine structure of single channel openings. It wasn't bad, but since *Nature* papers are so short they mostly can't be thought of as real papers, and four years later we published the work properly in the

Journal of Physiology ([Colquhoun & Sakmann, 1985](#)), the result of 6 years work (57 pages, 565 citations). For this Imperial would have awarded me a publication score of a mere 10.7.

Here is another interesting case. If we exclude chapters in *Single Channel Recording* ([Neher & Sakmann, 1983, 1995](#)) which apparently don't count, my most highly cited paper is Colquhoun, Neher, Reuter & Stevens (1981), This has 630 citations and a publication score of 36.6 for me, though only 14.6 for Harald Reuter. The reality behind this paper is as follows. In the early days of gigohm seal Harald Reuter decided that he wanted to learn the method, and to achieve this he invited three of us who already had some experience of the method to spend part of the summer vacation in Bern. We had a wonderful summer there, and being somewhat overmanned it was not very stressful. It would, I think, be fair to say that all four of us did much the same amount of work. While recording we noticed a type of channel that was opened by intracellular calcium, like the calcium-activated potassium channel that was already well known in 1981. This one was a bit different because it was not selective for potassium. We hadn't expected to get a paper out of the vacation job but it seemed novel enough to write up, and 1982 being a year when anything with "single channel" in the title, however trivial, sailed into *Nature*, and because we had a limited amount of data, we sent it there. Because we had all contributed much the same amount of work, we put the authors in alphabetical order. The analysis of the results, such as it was, was crude in the extreme (paper charts unrolled on the floor and measured with a ruler). If we hadn't seen this particular channel subtype, someone else would have done with a year or two. It just happened to be the first one of its type and so has been cited a lot, despite being scientifically trivial. This example shows not only the iniquitous uselessness of the publication score used by the Imperial; it also shows dramatically the almost equal uselessness of counting citations.

How not to get Nobel prizes

Employees of Imperial medical school are told

The divisional minimum benchmarks are:

"To have sufficient papers in top rated journals in the speciality to ensure 4 publications in the RAE review period (2001-2007) with an Impact Factor of at least 5 and with no overlap with co-authors from Imperial"

The "productivity" target for publications is to:

"To publish 3 papers per annum including one in a prestigious journal with an impact factor of at least 5"

Unfortunately Dr X has published only two papers in 2006 . . .

Let's see who lives up to their "productivity" criterion.

Take, for example two scientists who command universal respect in my own field,

Erwin Neher and Bert Sakmann. They got the Nobel Prize for Physiology or Medicine in 1991. In the ten years from 1976 to 1985, Sakmann published an average of 2.6 papers per year (range 0 to 6).

In six of these ten years he failed to meet the publication target set by Imperial, and these failures included the years in which the original single channel paper was published (Neher & Sakmann, 1976) and also the year when Colquhoun & Sakmann (1985) was published. In two of these ten years he had no publications whatsoever. On the other hand, a paper in 1981 in a journal with an ‘unacceptable’ impact factor of 3.56 has had over 15000 citations ([Hamill *et al.*, 1981](#)). This paper would have earned for Sakmann a publication score of a miserable 0.71, less than 100th of our perspective in *Science*.



Sakmann in Göttingen, 1980. He and Neher did the work *themselves*. (Picture from [here](#))

All this shows what is obvious to everyone but bone-headed bean counters. The only way to assess the merit of a paper is to ask a selection of experts in the field.

Nothing else works. *Nothing*.

It seems to have escaped the attention of bean-counters that this is precisely what has always been done by good grant giving agencies and search and promotion committees. Academics have always been assessed. But before HR departments and corporate-academics got involved, it was done competently. Now a whole branch of pseudo-science has appeared which devotes itself to trying to find ways of assessing people without bothering to find out what they have done. “Bibliometrics” is as much witchcraft as [homeopathy](#). How long, one wonders, will it be before somebody coins the term ‘bibliomics’? (Oops, a Google search shows I’m too late, some numbskull has already done it).

How to get good science

Universities will have to decide what sort of science they want. They can bend their policies to every whim of the RAE; they can bow to the pressures for corporatisation from the funding council. Or they can have creative scientists who win the real honours. They cannot have both. If they want to have the latter they will have to have universities run by academics. And they will have to avoid corporate and commercial

pressures. They will have to resist the pressures to remove power from their best researchers by abolishing eminent departments and centralising power at a higher level. We have seen what this approach has done to the NHS, but it is a characteristic of the corporatising mentality to ignore or misuse data. They just *know* they are right. It is also the box-ticking culture of managerialism that has resulted in approval of BSc degrees in anti-science (Colquhoun, 2007). Impressive sounding validation committees tick all the boxes, but fail to ask the one question that really matters: is what is being taught nonsense?

The policies described here will result in a generation of 'spiv' scientists, churning out 20 or even more papers a year, with very little originality. They will also, inevitably, lead to an increase in the sort of scientific malpractice that was recently pilloried viciously, but accurately, in the [New York Times](#), and a further fall in the public's trust in science. That trust is already disastrously low, and one reason for that is, I suggest, pressures like those described here which lead scientists to publish when they have nothing to say. I wrote recently (Colquhoun, 2007) "All of us who do research (rather than talk about it) know the disastrous effects that the Research Assessment Exercise has had on research in the United Kingdom: short-termism, intellectual shallowness, guest authorships and even dishonesty". Now we can add to that list bullying, harassment and an incompetent box-ticking style of assessment that tends to be loved by HR departments. This process might indeed increase your RAE score in the short term (though there is no evidence that it does even that). But, over a couple of decades, it will rid universities of potential Nobel prize winners.

References

Many of these papers are available [from here](#)

Colquhoun D (2003). Challenging the tyranny of impact factors. *Nature* **423** , 479 [[download pdf](#)].

Colquhoun D (2006). Playing the numbers game. A book review, of Does Measurement Measure Up? How Numbers Reveal and Conceal the Truth by John M. Henshaw. *Nature* **442** , 357.

Colquhoun D (2007). Science degrees without the science. *Nature* **446** , 373-374.

Colquhoun D, Neher E, Reuter H, & Stevens CF (1981). Inward current channels activated by intracellular calcium in cultured cardiac cells. *Nature* **294** , 752-754.

Colquhoun D & Sakmann B (1981). Fluctuations in the microsecond time range of the current through single acetylcholine receptor ion channels. *Nature* **294** , 464-466.

Colquhoun D & Sakmann B (1985). Fast events in single-channel currents activated by acetylcholine and its analogues at the frog muscle end-plate. *J Physiol (Lond)* **369** , 501-557.

Garfield E (1998)

[http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101\(6\)p413y1998english.html](http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101(6)p413y1998english.html)

Hamill OP, Marty A, Neher E, Sakmann B, & Sigworth FJ (1981). Improved patch clamp techniques for high resolution current recording from cells and cell-free membrane patches. *Pflügers Arch* **391** , 85-100.

Hawkes AG, Jalali A, & Colquhoun D (1992). Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Philosophical Transactions of the Royal Society London B* **337** , 383-404.

Hoeffel, C. (1998) Journal impact factors [letter]. *Allergy* 53, 1225

Neher E & Sakmann B (1976). Single channel currents recorded from membrane of denervated frog muscle fibres. *Nature* **260** , 799-802.

Seglen PO (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal* **314** , 498-502. [[Download pdf](#)].

Sivilotti LG & Colquhoun D (1995). Acetylcholine receptors: too many channels, too few functions. *Science* **269** , 1681-1682.