



Halt the avalanche of performance metrics

The increasing dominance of quantitative research assessment threatens the subjective values that really matter in academia, says Colin Macilwain.

The leaders of major universities around the world used to maintain a healthy scepticism towards league tables and the metrics that underpin them. But now, officials at institutions that do well in such assessments — partly on merit, and partly because they use the English language or have other historical advantages — are becoming beguiled with quantitative measures to rate the performance of academic staff. People who care about genuine quality in research and teaching need to resist that shift.

Universities evolved as self-governing bodies of academics. Originally, the president or vice-chancellor had a purely housekeeping role, once described in US parlance as assuring parking for the staff, sex for the students and sport for the alumni.

But lately — not least in Britain, where schemes such as the Research Assessment Exercise have come to dominate academic life — power has moved from the departments to the vice-chancellor. And university leaders, flanked by research managers and associated flunkies, want to use metrics to shift that balance still further.

Eight leading British universities are now energetically engaged in the joint development of a formidable computer tool that allows them to compare the performances of their researchers and departments against rivals, according to grant income, number of patents applied for, or pretty much any other criteria they choose. The tool is called Snowball (www.snowballmetrics.com) and the institutions signed up to it include the universities of Oxford and Cambridge, Imperial College London and University College London.

Like any metrics system, Snowball can, in theory, be used for good or ill. I suspect that in practice, however, it will end up being used mainly to exercise yet more control over academic staff, with every aspect of their professional lives tracked on the system.

Although Snowball was developed by people of genuine integrity who want to establish a fuller understanding of research performance, it shares a fundamental defect with other quantitative research-assessment tools: it is largely built on sand. It cannot directly measure the quality of research, never mind teaching, so instead it uses weak surrogates, such as the citation indices of individuals.

Citation indices — which rank research in terms of the average number of citations for articles — were robustly challenged earlier this year, when organizations led by the American Society for Cell Biology signed the San Francisco Declaration On Research Assessment (DORA), pledging to take a stand against the ever-expanding reach of journal-based metrics.

One of DORA's best ideas is to ask that citation databanks be available openly, for all researchers to use. I wish them luck with that. University

managers know that information is power — and they want not just the data, but to dictate how they are manipulated.

A major problem with metrics is the well-charted tendency for people to distort their own behaviour to optimize whatever is being measured (such as publications in highly cited journals) at the expense of what is not (such as careful teaching). Snowball is supposed to get around that by measuring many different things at once. Yet it cannot quantify the attributes that society values most in a university researcher — originality of thinking and the ability to nurture students. Which is not the same as scoring highly in increasingly ubiquitous student questionnaires.

Senior scientists have known for a long time that bogus measures of 'scientific quality' can threaten the peer-review system that has been painstakingly built up, in most leading scientific nations, to distribute funds on the basis of merit.

In the United States in 1993, for example, Congress passed the Government Performance and Results Act, which compelled federal agencies to start measuring their results. However, the US scientific establishment was strong and self-assured at that time, and successfully derailed the prospect that agencies such as the National Science Foundation (NSF) would start inventing numbers to 'measure' the work of its grant recipients. Instead, the NSF sticks to measuring things such as time to grant.

Nations with weaker scientific communities are less well-placed to fend off the march of metrics. The hazards are perhaps most immediate in places such as Italy, where peer review for grants has never fully taken hold, and China, where it

has rarely even been tried. There is a worrying tendency in developing countries, especially, for research agencies to skip the nuanced business of orchestrating proper peer review, and to move straight to the crude allocation of funds on the basis of measured performance. This bypasses quality and, bluntly, invites corruption.

But I see trouble ahead at the leading universities in the United Kingdom and the United States, too. Their reputations were built by autonomous academics, working patiently with students. If the name of the game becomes strong performance measured in numbers — as the vice-chancellors seem to want — it will kill the goose that laid the golden egg.

Defenders of Snowball say they are baffled that scientists, given what they do for a living, remain sceptical of research-performance metrics. But science seeks to identify and measure good surrogates, to test falsifiable hypotheses. Seen in that light, quantifiable research assessment does not measure up. Nevertheless, the snowball has started rolling down the mountain — and it is hard to see how its momentum will be arrested. ■

Colin Macilwain writes about science policy from Edinburgh, UK.
e-mail: cfmworldview@gmail.com

**BOGUS MEASURES
OF
'SCIENTIFIC
QUALITY'
CAN THREATEN THE
PEER-REVIEW
SYSTEM.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/tpfpwg