

Tide or tsunami? The impact of metrics on scholarly research

Andrew G Bonnell

University of Queensland

Australian universities are increasingly resorting to the use of journal metrics such as impact factors and ranking lists in appraisal and promotion processes, and are starting to set quantitative 'performance expectations' which make use of such journal-based metrics. The widespread use and misuse of research metrics is leading to increased concern in scientific and broader academic communities worldwide. This paper reviews some of the most important recent responses to the so-called 'metric tide', with particular reference to the report of that name recently issued by the UK's Higher Education Funding Council for England, and other important statements such as the San Francisco Declaration on Research Assessment and the Leiden Manifesto. While there is a spectrum of views on research metrics in general, there is widespread agreement from authoritative sources that it is not appropriate to rely on journal-level metrics, such as journal ranking lists, for assessing the merit of individual scholars.

Keywords: metrics, research, publications, performance management, managerialism

A couple of items of evidence on the current impact of metrics on researchers in Australian universities:

Exhibit A

A message from the Associate Dean for Research in the Faculty of Business and Law, Swinburne University, on 10 March 2015 reminded staff as usual to submit their recent publications to the university's electronic databank for the annual Higher Education Research Data Collection (HERDC), but this time with the stipulation: 'Publications in unranked (ABDC, IS or Law rankings) outlets, either journals or conference papers, should not be submitted. Reporting these publications to HERDC has a negative impact on our ERA [Excellence in Research for Australia] rankings' (Trounson, 2015).

In fairness, it should be added that Swinburne's acting Deputy Vice-Chancellor for Research and Development

subsequently issued a statement that the Associate Dean's message 'does not reflect university-level processes or communications'. Official communications to staff had made clear that all research publications should be submitted to the research bank. 'The university takes seriously its reporting obligations and further communications to staff will reinforce that all publications should be submitted to our repository so all eligible publications can be included in HERDC,' the DVC stated. The Australian Research Council (ARC) expressed its concern over the reporting of this message, and reaffirmed that universities needed to make complete submissions to the ERA (Trounson, 2015).

Exhibit B

The University of Queensland maintains an online database that tracks grants, research higher degree

supervisions and completions, and publications, over a six-year period, ascribing a numerical value to each, to two decimal places, which is compared in bar graphs to the average for the School, Faculty or Institute, the University, and the academic level (A-E). It is updated daily. The points values for publications are derived from a journal ranking list that was adapted from the 2010 ERA journal list promulgated by the Australian Research Council after a round of internal consultation during which staff were able to add unlisted journals or lobby for changes in the rankings. While such tabulations always come with a caveat that they should not be used in isolation, in practice individuals whose output falls below the average, however high the overall level of performance of a School, may feel under pressure, or may be put under pressure in performance appraisals.

It should be noted that the University's Q-T index for teaching is even more problematical than the Q-R index for research, being derived directly from an unweighted average of student evaluation scores.

This is despite the fact that the ARC prepared an amended ERA journal ranking list for the 2012 round after widespread criticism of the 2010 list. Some of the more egregious flaws of the 2010 list were corrected in the 2012 list, but the latter also reflected the lobbying efforts of various groups including professional associations. In any event, the 2012 list was withdrawn in 2011, prior to the 2012 ERA round, after a fresh round of criticism and complaints that the lists were being misused (e.g. for individual performance management purposes). Since then, the ARC has persistently advised that, in the words of ARC CEO Aidan Byrne, 'ERA hasn't made use of journal rankings since 2010, and while some universities have continued to use them internally, it is the ARC's firm view that this should stop' (Trounson, 2015).

But, like a virus released into the environment, once the journal rankings lists are out there, they can't be recalled by email. Data empower managers, and managers do not voluntarily relinquish the ability to assemble, deploy, and manipulate large datasets. Journal rankings are also used as a lever to seek to improve universities' ERA scores. As Frank Larkins' analysis of the different outcomes of the 2010 and 2012 ERA indicates, universities respond highly strategically to such ranking schemes, being very selective about what areas to submit for assessment. Consequently,

**...the message passed down the line,
and reiterated in academics' annual
performance appraisals and on other
occasions ... is that academics need to be
increasingly 'strategic' about where they
place their work.**

higher scores in the 2012 round do not automatically mean an increase in quality over a couple of years (inherently improbable given the time-frame) though they do reflect a degree of strategic 'gaming' (Larkins, 2013. Larkins is too discreet to use the word 'gaming' himself.)

Even if ERA scores bring relatively little funding, the indirect rewards from the reputational and marketing benefits of good ERA results lead to an increase in competitive behaviours and increased pressure on academics to perform in high-quality 'outlets'. While academic managers are not usually as crass as the Swinburne example, the message passed down the line, and reiterated in academics' annual performance appraisals and on other occasions, like applications for study leave, is that academics need to be increasingly 'strategic' about where they place their work. The status

hierarchy embodied in journal rankings, flawed and controversial though they might be, is reflected in the weightings used in points systems such as the Q-Index or in universities' increasingly quantified and explicitly articulated 'research performance expectations' for staff at specified academic levels of appointment, which refer to journal rankings. The Q-Index is calculated to two decimal places, providing an illusion of objectivity and precision. A number of universities have been promulgating such 'expectations' since the second ERA round.

Globally, indications are mounting that all is not well in scholarly publishing, and the misuse of metrics and attempts to exploit the shortcomings of systems of measurement are a frequent theme.

The Economist (2013) reported industrial-scale fraud, such as ghost-writing rackets, in China. The reasons given for this phenomenon are by no means isolated to China. *The Economist* wrote:

In the 1980s, when China was only beginning to reinvest in science, amassing publishing credits seemed a good way to use non-political criteria for evaluating researchers. But today the statistics-driven standards for promotion (even when they are not handed out merely on the basis of personal connections) are as problematic as in the rest of the bureaucracy.

A 'warped incentive system has created some big embarrassments', including mass retractions of dozens of articles by researchers who have been caught cheating.

The 'warped incentive scheme' derives from the fact that, as some Chinese scientists argue:

Some administrators are unqualified to evaluate research, ... either because they are bureaucrats or because they were promoted using the same criteria themselves. In addition, the administrators' institutions are evaluated on their publication rankings, so university presidents and department heads place a priority on publishing, especially for SCI [Science Citation Index] credits (Economist, 2013).

More recently, in April 2015, *The Lancet* issued a dramatic warning that 'reductive metrics' were leading to a crisis in scientific publishing. *The Lancet's* editor, Richard Horton (2015), wrote that the 'apparent endemicity of bad research behaviour is alarming':

much of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with an obsession for pursuing fashionable trends of dubious importance, science has taken a turn towards darkness.

Contributing factors causing this crisis include the fact that:

Universities are in a perpetual struggle for money and talent, endpoints that foster reductive metrics, such as high-impact publication. National assessment procedures, such as the Research Excellence Framework, incentivise bad practices.

The Academy of Medical Sciences, Medical Research Council, and Biotechnology and Biological Sciences Research Council are reported to be backing an investigation into the problem (Horton, 2015).

In February 2014, *Nature (News)* reported that the publishers Springer and the Institute of Electrical and Electronics Engineers (IEEE) had had to remove over 120 papers from their subscription platforms after French computer scientist Cyril Labbé had 'discovered that the works were computer-generated nonsense'. The gibberish papers came from 'more than 30 published conference proceedings between 2008 and 2013'. As *Nature (News)* reported, Labbé

is no stranger to fake studies. In April 2010, he used SCIdgen to generate 102 fake papers by a fictional author called Ike Antkare. Labbé showed how easy it was to add these fake papers to the Google Scholar database, boosting Ike Antkare's h-index, a measure of published output, to 94 — at the time, making Antkare the world's 21st most highly cited scientist.

[...] Labbé says that the latest discovery is merely one symptom of a 'spamming war started at the heart of

science' in which researchers feel pressured to rush out papers to publish as much as possible (*Nature (News)*, 2014).

Rising concern at the misuse of research metrics, and the negative effects of such metrics, prompted the Higher Education Funding Council for England (HEFCE) to undertake a major independent review of '*The Metric Tide*', which appeared in July 2015 (Wilsdon, *et al.* 2015). The steering group supporting Professor James Wilsdon of the Science Policy Research Unit, University of Sussex was highly distinguished, including Dr Liz Allen, Head of Evaluation of the Wellcome Trust, Sir Phillip Campbell, editor-in-chief of *Nature*, Dr Ian Viney, MRC Director of Strategic Evaluation and Impact of the Medical Research Council, London, and scholars from several UK universities as well as Leiden University. In some respects, the report is relatively conservative. It seeks to refine metrics and the UK's Research Excellence Framework (REF), not to abolish them. Nonetheless, it enunciates some grave criticisms of the current misuse of metrics.

Headline findings of *The Metric Tide* report include:

- Across the research community, the description, production and consumption of 'metrics' remains contested and open to misunderstandings. [...]
- Peer review, despite its flaws and limitations, continues to command widespread support across disciplines. Metrics should support, not supplant, expert judgement. Peer review is not perfect, but it is the least worst form of academic governance we have [...].
- Inappropriate indicators create perverse incentives. There is legitimate concern that some quantitative indicators can be gamed, or can lead to unintended consequences; journal impact factors and citation counts are two prominent examples (Wilsdon *et al.*, 2015, p. viii. Bold type in original).

Existing metrics systems were found to be in need of further development, and could not at present be relied on to replace more qualitative processes, such as narratives of case studies (Wilsdon *et al.*, 2015, pp.ix-x).

The first, overarching recommendation of the twenty recommendations in the Wilsdon report is:

'The research community should develop a more sophisticated and nuanced approach to the contribution and limitations of quantitative indicators.' (Wilsdon *et al.*, 2015, p. viii.)

The fourth recommendation is of particular interest here in the light of the quantitative performance management practices that are rapidly being adopted in Australia:

HR managers and recruitment or promotion panels in [higher education institutions] should be explicit about the criteria used for academic appointment and promotion decisions. These criteria should be founded in expert judgement and may reflect both the academic quality of outputs and wider contributions to policy, industry or society. Judgements may sometimes usefully be guided by metrics, if they are relevant to the criteria in question and used responsibly; article-level citation metrics, for instance, might be useful indicators of academic impact, as long as they are interpreted in the light of disciplinary norms and with due regard to their limitations. Journal-level metrics, such as the JIF [Journal Impact Factors], should not be used. (HR managers, recruitment and promotion panels, UUK [Universities UK] [to note]). (Wilsdon *et al.*, 2015, Recommendation 4, first sentence bold in original, last sentence: emphasis added.)

After the introductory chapter, a second chapter of *The Metric Tide* charts the rise of the field of 'scientometrics' and compares different national practices of institutionalised evaluation of research (including Australia's ERA). One of the chapter epigraphs, like some others in the report, reveals a subversive current that emerges from time to time:

'A timid, bureaucratic spirit has come to suffuse every aspect of intellectual life. More often than not, it comes cloaked in the language of creativity, initiative and entrepreneurialism.' David Graeber (Wilsdon *et al.*, 2015, p. 12)

Another chapter epigraph, perhaps inevitably, cites Douglas Adams from *The Hitchhiker's Guide to the Galaxy* on the meaning of life being the number 42. (Wilsdon *et al.*, 2015, p. 30).

One notable comment stresses the imperfect state of the common sources of bibliometric data:

As PLOS [Public Library of Science] noted in its response to our call for evidence, 'there are no adequate sources of bibliometric data that are publicly accessible, useable, auditable and transparent' (Wilsdon *et al.*, 2015, p. 17).

Arguably, this situation is even worse in the humanities, where many citations are in books or book chapters. The Wilsdon report acknowledges that: 'Research evaluation in book-oriented fields is more challenging than for article-based subject areas', for such reasons, and also finds that 'some academic books are primarily written for teaching (e.g. textbooks) or cultural purposes (e.g. novels and poetry) and citation counts of any kind may be wholly inappropriate for these' (Wilsdon *et al.* 2015, p. 40). The Wilsdon report considers disciplinary variations in a separate chapter, noting the differences in research and publishing culture between disciplines and cautioning

that 'Metrics should not become the "tail that wags the dog" of research practice in all disciplines' (Wilsdon *et al.*, 2015, p. 50).

Elsewhere, in noting the limitations of citation indices, the report notes that 'bibliometrics often do not distinguish between negative or positive citation, highly cited literature might attract attention due to controversy or even error' (Wilsdon *et al.*, 2015, p.5) There is also the question of publications in languages other than English, which are often under-represented in citation indices (Wilsdon *et al.*, 2015, p.52), and, for that matter, in journal ranking lists compiled in English-speaking countries.

In its consideration of current trends in bibliometrics, the Wilsdon report (2015, p. 35) finds that:

The use of journal-level indicators for assessing individual publications is rejected by many bibliometricians. It is argued that the distribution of citations over the publications in a journal is highly skewed, which means that the JIF and other journal-level indicators are not representative of the citation impact of a typical publication in a journal.

At the same time, the report noted that 'some bibliometricians agree with the use of journal-level indicators in the assessment of very recent publications', but mainly, it seems, as a default option where there has not been sufficient time for citation statistics for individual articles to accumulate (Wilsdon *et al.*, 2015, p. 35). The Wilsdon report notes that there have been significant concerns raised in recent statements representing the voices of many in the scientific community over the fact that

the application of indicators at inappropriate scales features prominently in recent statements, such as DORA [San Francisco Declaration on Research Assessment] and the Leiden Manifesto. Too often, managers and evaluators continue to rely on metrics that are recognised as unsuitable as measures of individual performance, such as journal-level indicators (Wilsdon *et al.*, 2015, p. 48).

One of the Wilsdon report's conclusions reflects a strong condemnation of the misuse of inappropriate indicators, such as journal rankings and JIFs:

Inappropriate indicators create perverse incentives. Across the community, there is legitimate concern that some of the quantitative indicators already being used to support decisions around research excellence and quality can be gamed and can lead to unintended consequences. The worst example of this is the widespread use of JIFs, where group (journal-level) metrics are ascribed to its non-homogenous constituents (articles) as a proxy for quality. There is also a very real possibility of existing or emergent indicators being

gamed (for example through 'citation clubs', salami-slicing of papers to increase citation counts, and battles over author positioning). These consequences need to be identified, acknowledged and addressed (Wilsdon *et al.*, 2015, p. 138. Bold type in original).

The chapters in the Wilsdon report on 'Management by metrics' and 'Cultures of counting' contain some sharp analysis. The 'import of more corporate styles of management', 'greater competition for scarce resources' and the extent to which higher education has become an 'export industry' are all identified as factors that are driving more metric-driven management practices (Wilsdon *et al.*, 2015, p. 68). Some publication metrics feed directly into some ranking systems, such as the Academic Ranking of World Universities (ARWU, formerly Shanghai Jiao Tong) and university managers perceive a direct link between success in such internationally publicised ranking lists, despite their often glaring methodological flaws, and the capacity to charge international students higher fees than less highly-ranked institutions. The Wilsdon report also notes that while 'pressures to incorporate metrics into research assessment within universities may have originated in response to external forces', such information-gathering processes can quickly take on a life and dynamic of their own (Wilsdon *et al.*, 2015, p. 69).

Under the heading 'Cultures of counting', the report states that management systems with a strongly quantitative dimension have made decision-making 'more transparent' and have 'allowed institutions to tackle genuine cases of underperformance'. These claims might be contested – procedures to tackle genuine lack of performance pre-exist metric-driven management systems, and decisions solely based on metrics would risk being unsafe in the light of all the qualifications that the report itself raises on the use of metrics. At the same time, the report notes:

many within academia resist moves towards greater quantification of performance management on the grounds that these will erode academic freedoms and the traditional values of universities. There is of course a proper place for competition in academic life, but there are also growing concerns about an expansion in the number and reach of managers, and the distortions that can be created by systems of institutionalized audit (Wilsdon *et al.*, 2015, p. 79).

The report cites concerns that 'metrics are widely seen as absolving research managers of the responsibility for making assessments based on more accurate and complete information, and as contributing to mistrust of research management more generally' (Wilsdon *et al.*, 2015, p. 80). Regrettably, after noting widespread concern at the negative effects of the 'cultures of counting', the authors of

The Metric Tide go on to state: 'It is beyond the scope of this report to resolve all of these issues' (Wilsdon *et al.*, 2015, p. 80). It does canvass a range of these issues, however.

There is the obvious factor of the observation effect:

Researchers are not passive recipients of research evaluation but play an active role in assessment contexts. Therefore, any system used to assess research, whether peer review or indicator-based, that affects money or reputation will tend to influence researchers' behaviour in two ways (Wilsdon *et al.*, 2015, p. 81).

The first of these two kinds of effects is goal displacement: chasing the metrics becomes the goal of researchers rather than the metrics measuring whether the research itself has been successful. The second effect relates to 'a change in the research process itself in response to assessment criteria' (Wilsdon *et al.*, 2015, p. 82). Here, the question of 'gaming' arises. The report is sceptical of claims that the UK's Research Assessment Exercise (RAE) and REF resulted in widespread 'gaming', but concedes that 'it isn't always entirely evident what distinguishes gaming from strategizing' (Wilsdon *et al.* 2015, p. 83).

Other concerns that are noted in this section of the report include possible biases against interdisciplinarity, the extent to which the production of journal ratings is a 'highly political task' (Wilsdon *et al.* 2015, p. 83, here citing Pontille and Torny, 2010), the pressure on researchers to stop doing certain kinds of work (such as book reviews, encyclopaedia entries), or 'task reduction' (Wilsdon *et al.* 2015, p. 85), 'increased levels of stress anxiety among academics' under increasingly metrics-based regimes of management, and the effects on knowledge production of factors such as the 'conservatism of metrics users' (Wilsdon *et al.* 2015, p. 85, here citing Butler, 2003; 2005). There are also possible negative effects in terms of equity and equal opportunity in reliance on research metrics, including gender bias, which is the result of a number of factors from the social distribution of carers' work to the fact that men are apparently more likely to cite their own work (or each other's) (Wilsdon *et al.* 2015, pp. 90-95). The gendered effect of metrics fostered by the ERA in Australia has been recently analysed by Lipton (2015, p. 69), who finds that such 'quality assurance measures' and the performance metrics on which they rely 'continue to reflect and valorise the ideal academic as male and masculine principles of knowledge production, which dominate structures of governance'.

The problem of 'task reduction' identified by *The Metric Tide* report has been evident in Australia for some time now, especially since the inception of the ERA. As

university managers assign points values corresponding to 'outputs' and to proxies for quality such as journal rankings, and as these points values seep into workload allocation processes, performance appraisal regimes, and publication 'incentive' schemes (extra research funding for publishing in the 'right' places), academics come under increasing pressure to rationalise their activities, and early career staff especially are warned off activities that do not get rewarded in metrics. In September 2014, a group of over 40 editors of journals published by Wiley in Australia signed an open letter, coordinated by Martha McIntyre, drawing attention to the system of 'perverse incentives' under which 'the voluntary inputs of reviewing and editorial services to academic journals' were unrewarded and under-recognised at the same time as institutions put ever-increasing value on publication in peer-reviewed journals:

The ERA procedures effectively mean that certain research activities are rewarded while other academic activities are not; and that universities suffer financial consequences if their academic staff do not privilege the winning of large grants and publication of articles in prestigious, high quality journals over all other work. These journals have of course become prestigious precisely because of the hard work of successive editors, associate editors and reviewers, which, for the most part, is unpaid (McIntyre *et al.*, 2014).

Editors report that they are receiving increasing requests for special issues, which pose their own demands on reviewing; they also report increasing difficulty finding qualified people to undertake peer-reviewing of articles, and some journals are experiencing difficulties in finding editors (McIntyre *et al.*, 2014). (Disclosure: the author was a signatory of the McIntyre open letter in his capacity as a journal editor.) A recent article in the *Australian Universities' Review* by Franklin Obeng-Odoom came to the defence of book reviewers despite the lack of recognition and reward that attaches to reviewers, despite the fact that academics in book-based disciplines always crave good reviews for themselves. As Obong-Odoom puts it (2014, p. 78), 'One contradiction in the status quo is that academics expect to be served but they are discouraged from serving and hence are led down a line of being selfish'. This sums up in a nutshell the behavioural effects of the current incentive schemes which are largely driven by evaluation regimes and the metrics that underpin them.

While the Wilsdon report seems to have taken its brief to be the fine-tuning of research exercises such as the UK's REF, rather than advocating their abolition, it does note some significant concerns in its reflection on the REF. These include: 'the potential that some types of quantitative data could encourage particular behaviours

that were not necessarily positive. Examples ranged from the use of "citation clubs" to boost citations, to major distortions in the research endeavour, downplaying whole disciplinary areas' (Wilsdon *et al.*, 2015, p. 119). The report concludes that 'it is not currently feasible to assess the quality of research outputs using quantitative indicators alone' (Wilsdon *et al.*, 2015, p. 131).

The Wilsdon report cites, and partly follows, some recent statements by bodies representing significant numbers of scientists, which have articulated concerns at the misuse of metrics. The 2013 San Francisco Declaration on Research Assessment (DORA) followed on from the December 2012 conference of the American Society for Cell Biology, at which strong concerns were aired at the way in which current citation practices were having perverse effects on the scientific enterprise. The Declaration's primary general recommendation was:

1. Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions (Declaration on Research Assessment, 2013).

More specifically, it was recommended that institutions:

4. Be explicit about the criteria used to reach hiring, tenure, and promotion decisions, clearly highlighting, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published.

5. For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice.

In its recommendations to individual researchers, the Declaration reiterates the injunction:

15. When involved in committees making decisions about funding, hiring, tenure, or promotion, make assessments based on scientific content rather than publication metrics.

[and:]

18. Challenge research assessment practices that rely inappropriately on Journal Impact Factors and promote and teach best practice that focuses on the value and influence of specific research outputs (Declaration on Research Assessment, 2013).

As of 22 August 2015, the Declaration had over 12,500 individual signatories and 588 institutional signatories. The institutional signatories include the British Academy

and a number of national learned academies, such as the Austrian and Czech Academies of Sciences, as well as the Australian Academy of Science, the Association of Australian Medical Research Institutes, Neuroscience Research Australia, and the Association of Australian Cotton Scientists. Australia's National Health and Medical Research Council (NHMRC) has also signed the San Francisco Declaration (NHMRC, 2015) and even earlier, in April 2010, had issued a statement discouraging the use of Journal Impact Factors in applications or peer review of applications, stating: 'Journal Impact Factor is not a sound basis upon which to judge the impact of individual papers' (NHMRC, 2010). The NHMRC in 2015 has broadened this statement to read:

It is not appropriate to use publication and citation metrics such as Journal Impact Factors, the previous Excellence in Research for Australia (ERA) Ranked Journal List or h-index when assessing applications as these can potentially be misleading when applied to the peer review of publication outputs of individuals, and may also not be relevant to the project under consideration (NHMRC, 2015).

Of the many universities and university schools and institutes to have signed the San Francisco Declaration, the only Australian university signatories to appear on the DORA website's list are Murdoch University and the University of Queensland's Institute for Molecular Bioscience, although this fact does not seem to be publicised on their own websites.

The Leiden Manifesto proposed ten principles for the responsible measurement of research performance. It was composed by Diana Hicks (Professor in Public Policy, Georgia Institute of Technology), Paul Wouters (Leiden University), and three of their colleagues and was published in *Nature (News)* as a comment (2015).

The Leiden Manifesto states:

As scientometricians, social scientists and research administrators, we have watched with increasing alarm the pervasive misapplication of indicators to the evaluation of scientific performance.

[...]

Some recruiters request h-index values for candidates. Several universities base promotion decisions on threshold h-index values and on the number of articles in 'high-impact' journals. Researchers' CVs have become opportunities to boast about these scores, notably in biomedicine. Everywhere, supervisors ask PhD students to publish in high-impact journals and acquire external funding before they are ready.

In Scandinavia and China, some universities allocate research funding or bonuses on the basis of a number: for example, by calculating individual impact scores to

allocate 'performance resources' or by giving researchers a bonus for a publication in a journal with an impact factor higher than 15.

In many cases, researchers and evaluators still exert balanced judgement. Yet the abuse of research metrics has become too widespread to ignore.

The recommendations of the Leiden Manifesto include: '7) Base assessment of individual researchers on a qualitative judgement of their portfolio' (Hicks *et al.*, 2015).

There is thus a large and growing body of scientific opinion, and academic opinion more broadly, expressing concern about the growing tendency for metrics to be used inappropriately. In particular, there is condemnation from authoritative bodies such as the ARC, the NHMRC and the UK's HEFCE of the practice of using journal-level metrics and rankings for individual performance appraisal. Despite this, Australian universities continue down this path.

While writing this paper, I was, therefore, somewhat dispirited to read the latest upbeat aspirational statement from a leading Australian university: The University of New South Wales' (UNSW) August 2015 White Paper (p.8) stated its 'Objective No.1' in research as:

To establish UNSW as one of the top 50 research-intensive universities worldwide. UNSW will have leading researchers across all faculties and many of our staff will be amongst the world's most highly cited researchers. The number of publications appearing in leading journals will have doubled [by 2025].

The point of citing this is not to single out UNSW. The managers of all 'Group of Eight' major research universities would profess similar (probably identical) aspirations. Driven by competition for international student numbers, not to mention the quantified KPIs of individual managers, and seeking to justify charging higher fees than their competitors, academic managers chase rankings, and use crude quantitative levers to try to extract more and higher-profile publications from their staff.

It is not sustainable. Already, leading journals such as *The Lancet* and *Nature* are complaining of being spammed. Quantitative performance indices lock in overwork and undermine both collegiality and equity objectives as academics are pitted against each other in pursuit of rolling average output norms. Incentives for gaming and fraud mount, and the altruistic collegial behaviours on which the research eco-system depends are implicitly penalised. The fetishisation of journal rankings also undermines institutions' claims to support greater open access to research, puts a premium on conservative

publication practices, with the risk that innovation and interdisciplinary work will be marginalised, and potentially undermines academic freedom. With systemic public underfunding of higher education over a couple of decades at the root of the malaise of Australian universities, and little fiscal relief in sight, it is impossible to say when our research eco-system will either improve or implode.

Acknowledgement

I would like to acknowledge the valuable feedback provided by participants in the following symposia to which an earlier version of this paper was presented: Reclaiming the Knowledge Commons, held by the Centre for Values, Ethics and the Law in Medicine, University of Sydney, at State Library of NSW, Sydney, 26 August 2015, and the National Scholarly Communications Forum (sponsored by Australian Academy of the Humanities), Australian National University, Canberra, 7 September 2015.

Andrew G. Bonnell is an associate professor in the School of Historical and Philosophical Inquiry, University of Queensland, Australia.

Contact: a.bonnell@uq.edu.au

References

- Butler, L. (2003). Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publication counts. *Research Policy*, 32(1), 143-155.
- Butler, L. (2005). What happens when funding is linked to publication counts? In: H. Moed, W. Glänzel, & U Schmoch (Eds.), *Handbook of quantitative science and technology research*. Dordrecht: Springer, pp. 389-405.
- Declaration on Research Assessment (2013). Retrieved from <http://www.ascb.org/dora/>.
- Economist, The*. (2013). Looks good on paper. Retrieved from <http://www.economist.com/news/china/21586845-flawed-system-judging-research-leading-academic-fraud-looks-good-paper>.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & I. Rafols (2015) Bibliometrics: The Leiden Manifesto for Research Metrics, *Nature News*, Vol. 520, Issue 7548. Retrieved from <http://www.Nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351>; also <http://leidenmanifesto.org>.
- Horton, R. (2015). Offline: What is medicine's 5 sigma? *The Lancet* 385, April 11, p.1380.
- Larkins, F.P. (2013). ERA 2012 (Part 1): University Responses and Performances Compared with ERA 2010, paper for L.H. Martin Institute for Tertiary Education Leadership and Management, 2013. Retrieved from http://www.lhmartininstitute.edu.au/userfiles/files/Blog/FLarkins_HE%20Research%20Policy%20Analysis_ERA2012_pt1_Feb2013.pdf.
- Lipton, B. (2015). A new 'ERA' of women and leadership. *Australian Universities' Review*, 57(2), 60-70.
- McIntyre, M., S. Jones, M. Bartold, L. Frost, M. Shanahan, M. Weder, P. Jensen, et al. (2014). *Journal reviewing and editing: Institutional support is essential. Submission to institutions relevant to the higher education sector in Australia on behalf of Australian editors of academic journals*. September. Retrieved from http://www.air.asn.au/cms_files/00_Homepage/00_LatestNews/journal_reviewing_institutional_support_sept14.pdf.
- Nature (News)* (2014). Publishers withdraw more than 120 gibberish papers. 25 February. Retrieved from <http://www.Nature.com/news/publishers-withdraw-more-than-120-gibberish-papers-1.14763>.
- NHMRC (National Health and Medical Research Council) (2010). NHMRC removes Journal Impact Factors from Peer Review of Individual Research Grant and Fellowship Applications. Retrieved from https://www.nhmrc.gov.au/_files_nhmrc/file/grants/peer/impact%20factors%20in%20peer%20review.pdf.
- NHMRC (National Health and Medical Research Council). (2015). *Guide to Peer Review*. Retrieved from <https://www.nhmrc.gov.au/book/guide-nhmrc-peer-review-2015>.
- Obeng-Odoom, F. (2014). Why Write Book Reviews? *Australian Universities Review*, 56, 1, pp.78-82.
- Pontille, D., & Tomy, D. (2010). The controversial policies of journal ratings: Evaluating social sciences and humanities. *Research Evaluation*, 19, 5, pp.347-360.
- Trounson, A. (2015) Swinburne accused of research ratings ploy. *The Australian*, 1 April.
- UNSW (University of New South Wales). (2015). UNSW 2025. Statement of Strategic Intent. White Paper, UNSW, August, p.8. Retrieved from <https://www.2025.unsw.edu.au/whitepaper/UNSW-2025-White-Paper-13-3.pdf>.
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J. & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. DOI: 10.13140/RG.2.1.4929.1363.