

Time to abandon the gold standard? Peer review for the REF falls far short of internationally accepted standards.

 blogs.lse.ac.uk/impactofsocialsciences/2014/11/19/peer-review-metrics-ref-rank-hypocrisies-sayer/

11/19/2014

The REF2014 results are set to be published next month. Alongside ongoing reviews of research assessment, [Derek Sayer](#) points to the many contradictions of the REF. Metrics may have problems, but a process that gives such extraordinary gatekeeping power to individual panel members is far worse. Ultimately, measuring research quality is fraught with difficulty. Perhaps we should instead be asking which features of the research environment (a mere 15% of the assessment) are most conducive to a vibrant research culture and focus funding accordingly.



At the behest of universities minister David Willetts, HEFCE established an Independent Review of the Role of Metrics in Research Assessment in April 2014 chaired by James Wilsdon. This followed consultations in 2008-09 that played a decisive role in persuading the government to back down on previous plans to replace the RAE with a metrics-based system of research assessment. Wilsdon's call for evidence, which was open from 1 May to 30 June 2014, received 153 responses 'reflecting a high level of interest and engagement from across the sector' ([Letter to Rt. Hon. Greg Clark MP](#)). Sixty-seven of these were from HEIs, 27 from learned societies and three from mission groups. As in 2008-09, the British academic establishment (including the Russell Group, RCUK, the Royal Society, the British Academy, and the Wellcome Trust) made its voice heard. Predictably, '57 per cent of the responses expressed overall scepticism about the further introduction of metrics into research assessment,' while 'a common theme that emerged was that peer review should be retained as the primary mechanism for evaluating research quality. Both sceptical and supportive responses argued that metrics must not be seen as a substitute for peer review ... which should continue to be the "gold standard" for research assessment' (Wilsdon review, [Summary of responses to submitted to the call for evidence](#)).

The stock arguments against the use of metrics in research assessment were widely reiterated: journal impact factors cannot be a proxy for quality because 'high-quality' journals may still publish poor-quality articles; using citations as a metric ignores negative citation and self-citation; in some humanities and social science disciplines it is more common to produce books than articles, which will significantly reduce their citation counts, and so on. Much of this criticism, I would argue, is a red herring. Most of these points could easily be addressed by anybody who seriously wished to consider how bibliometrics might sensibly inform a research assessment exercise rather than kill any such suggestion at birth (don't use JIFs, exclude self-citations, use indices like Publish or Perish that include monographs as well as articles and control for disciplinary variations). What is remarkable, however, is that while these faults are often presented as sufficient reason to reject the use of metrics in research assessment out of hand, the virtues of 'peer review' are simply assumed by most contributors to this discussion rather than scrutinized or evidenced. This matters because whatever the merits of peer review in the abstract—and there *is* room for debate on what is by its very nature a subjective process—the evaluation procedures used in REF 2014 (and previous RAEs) not only fail to meet HEFCE's own claims to provide 'expert review of the outputs' but fall far short of internationally accepted norms of peer review.



Image credit: [Tulane Public Relations](#) (CC BY)

In sharp contrast with the evaluative procedures used in a range of other academic contexts in the UK and internationally, the REF relies entirely on in-house assessment by panels of (so-called) experts. On some panels, like history, just *one* assessor may read each output, something unheard of in peer review for journal and book submissions, research grant competitions, and tenure and promotion proceedings. Additionally, almost all REF panelists are drawn from British universities alone. David Eastwood admitted back in 2007 that ‘international benchmarking of quality’ was ‘one thing that the RAE has not been able to do’—which is cute, considering that REF panels award their stars on the basis of whether outputs are ‘world leading’, ‘internationally excellent,’ or merely ‘recognized internationally.’ Eastwood then still hoped to solve the problem with ‘bibliometrics, used with sensitivity and sophistication’ ([‘Goodbye to the RAE ... and hello to the REF’](#), *Times Higher Education*, 30 November 2007). HEFCE’s prohibitions on using journal impact factors, rankings, or the perceived standing of publishers (and humanities and most social science panels’ refusal to use *any* bibliometric data, including citations) reinforce the total dependency of REF evaluations on these panelists’ subjective judgments. Meantime the abandonment of RAE 2008’s use of external advisors where a panel felt it lacked specialist expertise and an overall cut in the number of panels from 67 in RAE 2008 to 36 in REF 2014 further reduced the pool of expertise available to panels, which were now also only ‘exceptionally’ allowed to cross-refer outputs to other panels.

If we could be confident that REF panels nevertheless ‘provide sufficient breadth and depth of expertise to undertake the assessment across the subpanel’s remit (including as appropriate expertise in interdisciplinary research and expertise in the wider use or benefits of research)’ (HEFCE, [REF 2014: Units of Assessment and Recruitment of Expert Panels](#), 2010, para 55) this might not be a problem. But we cannot. Nobody on the REF history panel, for example, has specialist knowledge of China, Japan, the Middle East, Latin America, or many countries in Europe (including once-great powers like Austria-Hungary, Spain and Turkey), though work on all these areas has likely been submitted to the REF. Whatever their general eminence in the historical profession, these ‘experts’ do not know the relevant languages, archives, or literatures. How, then, can they possibly judge the ‘originality’ of an output or its ‘significance’ in any of these fields? And on what conceivable basis can they be entrusted to determine whether it is ‘internationally excellent’ or merely ‘internationally recognized’—the critical borderline between 3* research that will attract QR funding and 2* research that will not?

REF panelists are unlikely to have the time to do a proper assessment anyway. In all, around 1000 evaluators will have graded all 191,232 outputs for REF 2014 in under a year—the same number *in total* as the US National

Endowment for the Humanities uses to evaluate 5700 applications for its 40 grant programs! Peter Coles calculates that each member of the physics panel must read 640 research papers, i.e., about two a day. 'It is ... blindingly obvious,' he concludes, 'that whatever the panel does do will not be a thorough peer review of each paper, equivalent to refereeing it for publication in a journal' ('[The apparatus of research assessment](#)', *LSE Impact Blog*, 14 May 2014). One RAE 2008 panelist told *Times Higher Education* that it would require 'two years' full-time work, while doing nothing else' to read properly the 1200 journal articles he had been allocated ('Burning questions for the RAE panels', 24 April 2008). Another admitted: 'You read them sufficiently to form a judgment, to get a feeling ... you don't have to read to the last full stop' ('Assessors face "drowning" as they endeavour to read 2,363 submissions', 17 April 2008).

For major academic journals the process of review is often double blind. Though university presses and other academic book publishers divulge authors' identities to reviewers, they will often also first consult with the author on appropriate reviewers. Reviewers are required to provide substantial comments in either case. The REF, by contrast, makes no attempt to protect authors' anonymity—something we might think especially important when judgments may lie in the hands of a single assessor. And far from providing comments justifying their grades, RAE 2008 subpanels shredded all documents showing how they reached their conclusions and ordered members ordered to destroy personal notes in order to avoid having to reveal them under Freedom of Information Act requests. It is difficult to think of a procedure that would make it easier for evaluators to further ideological agendas or settle personal scores, should they be so inclined.

Metrics may have problems. But a process that willfully ignores whether an output has gone through any peer review before publication, where it has been published, and how often it has been cited in favor of the subjective opinions of evaluators who may have no specialist expertise in the field and then systematically erases all records does not strike me as a very defensible alternative. It also gives extraordinary gatekeeping power to the individuals who sit on REF panels. This is worrying because the mechanisms through which panelists are recruited are tailor-made for the sponsored replication of disciplinary elites. All applicants for panel chairs have to be endorsed by learned societies, chairs in turn 'advise' on the appointment of panel members, and at least one third of panelists have to have served in a previous RAE. Apart from being disproportionately older, white, and male compared with the UK academic profession in general, these may not always be the scholars best placed to identify cutting-edge research, especially where such research crosses or challenges disciplinary boundaries. NEH, we might note, prohibits its evaluators from serving in successive competitions to reduce this risk.

Were Wilsdon's committee to assess what might be achieved employing appropriately sophisticated metrics against what is actually done in the REF rather than comparing a crude caricature of metrics with an idealized chimera of peer review, I think it would have to take a different view of the merits of the two systems than that put forward in most responses to the call for evidence. For its REF process to be comparable to what is understood elsewhere as peer review, HEFCE would have to use subject matter-specific experts from an international pool, commission a minimum of two reviews of each output, and not overload reviewers with too many outputs for them to read them properly in the timeframe available. This would be even costlier in public money and academics' time than the present REF. To replace the REF with metrics, on the other hand, would yield a process that is (in Dorothy Bishop's words) 'transparent and objective, it would not require departments to decide who they do and don't enter for the assessment, and most importantly, it wins hands down on cost-effectiveness' ('An alternative to REF2014?' [Bishopblog](#), 26 January 2013).

Metrics have in fact proven to be highly reliable *predictors* of RAE performance, irrespective of whether or not they provide valid *measures* of research quality. It is not without irony that there is considerable overlap between RAE scores and major commercial university rankings, even though the research component in the latter relies primarily on bibliometrics. The *Times Higher Education* bases 30% of its [World University Rankings](#) on citations. Eleven British universities made the top 100 in its 2013–2014 rankings. Of these, eight were also in the top 10 in RAE 2008 and the other three in the top 20. Knowing this, given the choice between the intellectual charade of REF 'expert peer review' and appropriate metrics (the Web of Science is of little use in history) I would unhesitatingly choose the

latter. It is infinitely cheaper, much less time consuming, and does not have the negative consequences for collegiality and staff morale of the present system. My own department must be one of many in which some colleagues are now no longer talking to one another because of a breakdown in trust over staff selection for REF 2014—hardly a framework for research excellence.

The conclusion I would rather draw, however, is that peer review vs. metrics is in many ways not the issue. Neither is capable of measuring research *quality* as such—whatever that may be. Peer review measures conformity to disciplinary expectations and bibliometrics measure how much a given output has registered on other academics' horizons, either of which *might* be an indicator of quality but neither of which has to be. It seems rather silly to base 65% of the REF ranking on something that we cannot measure and that may be inherently unmeasurable because it is a subjective judgment. Perhaps we should instead be asking which features of the research environment (which currently counts for a mere 15% of the REF assessment) are most conducive to a vibrant research culture and focus funding accordingly. Library and laboratory resources, research income, faculty members' involvement in conferences, journal or series editing, and professional associations, PhD student numbers and the intellectual life of a department as reflected in research seminars and public lectures are all good indicators of research vitality. They are also measurable.

Professor Sayer's book [Rank Hypocrisies: The Insult of the REF](#) is available for pre-order on Amazon now.

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Author

Derek Sayer is Professor of Cultural History at Lancaster University and Professor Emeritus (Canada Research Chair in Social Theory and Cultural Studies) at the University of Alberta. His new book *Rank Hypocrisies: The Insult of the REF* is to be published by Sage on December 3.

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.

