

# One scholar's crusade against the REF

The flawed research excellence framework is not a process of peer review in any meaningful sense, argues Derek Sayer, who appealed against his inclusion in the exercise

Published on  
December 11, 2014  
Last updated  
June 10, 2015

- —
- —
- —
- —
- —
- —
- —



Source: Getty

*My appeal was impelled by anger at seeing huge amounts of public money squandered on an exercise I came to believe had no real purpose*

One of the grand rituals of British academic life is about to reach its climax. As UK academics are acutely aware, the [results of the research excellence framework](#), as the research assessment exercise was renamed in good Orwellian fashion in 2007, will be published on 18 December. The change of name was intended to signal a radical shift in methods of assessing research quality, of course, but during the Higher Education Funding Council for England's consultations in 2007-09 the massed ranks of the British academic establishment succeeded in persuading the government to abandon its plans to replace what they described as "expert review" or "peer review" with metrics. Instead, REF 2014's 36 disciplinary panels have read and graded 191,232 research outputs by 52,077 academic staff located in 155 higher education institutions.

The REF panels' judgements will determine individual universities' share of Hefce's so-called QR (quality-related) research funding. But the rankings will also establish a pecking order that affects everything from institutions' ability to compete for external research grants to their capacity to attract top-notch faculty and graduate students. Readers of *Times Higher Education* will know that the REF casts a long shadow over British academic life. Pressure to maximise REF scores increasingly drives what is researched, how it is funded and where it is published. It influences hiring and promotion decisions, with "REFability" often trumping all other considerations. What began as a "light-touch" periodic audit in 1986 has spawned university bureaucracies that continually monitor and seek to manage individuals' research within REF priorities and timelines.

To be excluded from the REF is therefore a serious matter. In a June 2013 [University and College Union survey](#) discussed in *THE*, threats of redundancy for those not entered were reported by 29 per cent of respondents at [Middlesex University](#), 24 per cent at the [University of Leicester](#), 21 per cent at City University London, 18 per cent at [Queen's University Belfast](#), 13 per cent at the [University of Birmingham](#) and 11 per cent at the universities of Sussex, Cardiff and Warwick. Whether or not such threats are carried out, staff members not submitted to REF 2014 have had their reputations sullied, their confidence dented and their career prospects undermined. It was the exclusion of several of my colleagues in history at [Lancaster University](#) that led me to take the unusual step of [appealing against my inclusion in the REF](#).

My appeal was impelled by deep anger: anger as a citizen at seeing huge amounts of public money squandered on an exercise I came to believe had no real purpose beyond legitimating the replication of the UK's academic elites; anger as a scholar at seeing intellectual horizons narrowed, imaginations cramped and "risky" work

marginalised in the interests of maximising REF scores; and anger as a professional at seeing colleagues whose publications had passed the far more rigorous peer-reviewing procedures of international publishing houses and journals excluded from the REF on the basis of dubious university-commissioned evaluations. As a senior professor, I thought I had a moral duty to speak out against a process that in my view shames as well as damages Britain's universities. This article explains why.

The REF has been criticised inter alia for its monetary cost (officially £60 million, but likely much more), its opportunity costs in time that could have been spent in the classroom, the library or the lab, its destructive effects on collegiality and morale, its contribution to departmental closures and job losses, and the dangers posed to pure research by its introduction of "impact" as a dimension of evaluation. I agree with all these criticisms.

Far less attention, however, has been paid to the claim on which the REF stakes its entire legitimacy as a process of research evaluation – the claim that it is a process of expert peer review. I shall argue here that this claim is untenable. If I am right, there is no reason to give any greater credence to the evaluations published on 18 December than to, say, the [THE World University Rankings](#). Indeed, whatever one might think of the metrics upon which such rankings are based as measures of research quality, they have proved extremely reliable predictors of RAE performance in the past at a fraction of the REF's cost in time or money.



So let me begin by asking: what is generally practised as peer review in other academic settings, not just in the UK but internationally? Briefly, the key criterion in choosing reviewers is whether they are qualified by their own records of publication in the relevant field to evaluate a submission. Leading journals and university presses commission several reviews for each manuscript to ensure a spread of opinion and counteract possible biases. Similar principles apply in other settings where peer review forms part of evaluations. While the committees that decide research grant applications or tenure and promotion cases may not be made up of subject-matter experts, their judgements will invariably be informed by a range of specialist reviews.

Judged against these benchmarks, Britain's REF falls lamentably short. REF panels have from 10 to 30 members depending on the discipline, some of them "user members" drawn from companies, government or charities who are not involved in grading outputs. In all, about 1,000 assessors will have graded all 191,232 outputs for REF 2014 – the same number in total as the National Endowment for the Humanities in the United States uses to evaluate 5,700 applications for its 40 grant programmes. Peter Coles, head of the School of Mathematical and Physical Sciences at the [University of Sussex](#), calculates that each member of the physics panel must read 640 research papers in less than a year – in other words, about two a day. "It is...blindingly obvious," he concluded in a blog posted on 14 May, "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". One [RAE 2008 panellist](#) told *THE* that it would require "at least two or three hours" to read properly each of the 1,200 journal articles he had been allocated, that is, "two years' full-time work, while doing nothing else". [Another admitted](#): "You read them sufficiently to form a judgement, to get a feeling...you don't have to read to the last full stop."

The root of this problem, and the source of many others, is that all REF assessment is done in-house. Panel members alone are responsible for evaluating outputs, and in some panels the volume of work is such that only one panel member reads each output. Hefce's prohibitions on using journal impact factors, rankings or the perceived standing of publishers, as well as humanities and social science panels' refusal to use any bibliometric data, reinforce this dependency on subjective opinions. In addition, REF 2014 abandoned the RAE's use of external "specialist advisers" in areas that panel members did not feel qualified to cover or which crossed disciplinary boundaries, and permitted cross-referral to other panels only exceptionally. Reduction of the number of panels from 67 in RAE 2008 to 36 in REF 2014 further stoked fears that panels might not "include sufficient breadth and depth of expertise to produce robust assessments", according to the 2010 Hefce document, *REF2014: Units of Assessment and Recruitment of Expert Panels*. Those fears proved well founded.

The academic evaluators on the history panel read close to 7,000 outputs from more than 1,750 researchers covering all periods of history and areas of the world. Sixteen are historians of Britain, three of whom also work in imperial history. There are six historians of individual European countries, two of the United States and one of Africa. Clearly, the chances of outputs in history being read by panellists who are experts in an author's country of research (let alone their period and substantive field) are very unevenly distributed. If you work on the history of China, Japan, the Middle East, Latin America, or – as in my own case – the Czech Lands, nobody on the panel knows the languages, the archives or the secondary literature. How, then, can they judge the "originality" of an output or its "significance" if they do not know the field? On what conceivable basis can they be trusted to determine whether an output is "internationally excellent" or merely "internationally recognised" – the boundary between 3\* research (which attracts QR funding) and 2\* (which does not) – especially when they are expressly forbidden to use any bibliometric or other contextual data?

The likelihood that any such broad panel could contain a range of expertise sufficient to produce credible evaluations of all the work that falls under its remit is

doubtful, but confining membership almost entirely to people working in British universities hardly improves matters. When he was chief executive of Hefce, [David Eastwood](#) admitted in the pages of *THE* that “international benchmarking of quality” was “one thing that the RAE has not been able to do” – which is rich, considering that REF panels award their stars on the basis of whether outputs are “world-leading”, “internationally excellent” or merely “recognised internationally”. Academic journals, publishing houses, funding agencies and tenure and promotion committees across the world rely on an international pool of referees as a matter of course. Indeed, many British universities’ procedures for promotion to professor specifically require input from some non-UK referees.



*This underbelly of the REF is difficult to document because its victims are reluctant to speak on the record and universities hide their selection criteria*

A final contrast between the REF and standard peer-reviewing procedures is worth highlighting. For major academic journals the process of review is often double blind. Protecting reviewers’ anonymity allows them to express their opinions freely while communicating their comments to authors makes reasons for publication decisions transparent. The REF, by contrast, makes no attempt to protect authors’ anonymity – something we might think especially important when judgements may lie in the hands of a single assessor. And far from providing authors with comments, all documents showing how RAE 2008 subpanels reached their

conclusions were shredded and members ordered to destroy personal notes in order to avoid having to reveal them under Freedom of Information Act requests ("[Panels ordered to shred all RAE records](#)", *THE*, 17 April 2008).

In response to changes in Hefce's QR funding formula in 2010-11 (which defunded 2\* outputs entirely), many universities launched "internal REFs" with a view to excluding academics who emerged with an inadequate "grade point average" from their REF 2014 submissions. The uniform and relatively transparent – if far from expert – processes for evaluating outputs through RAE panels that had been a hallmark of the system since 1992 were now supplemented by the highly divergent, frequently ad hoc and generally anything-but-transparent staff selection procedures of individual institutions. This underbelly of the REF is difficult to document because its victims are often reluctant to speak on the record and universities hide their selection practices behind firewalls of confidentiality. I shall confine myself here to what went on in my own department, but there is plenty to suggest that [Lancaster University](#) was far from the only institution to play fast and loose with Hefce's criteria for staff selection of transparency, accountability, consistency and inclusivity (google, for example, the results of Warwick UCU's survey).

All eligible staff members in Lancaster's history department were required to identify four outputs for submission to the REF. These outputs were first read by a "critical friend" – an eminent historian from another UK university – who was not a specialist in most of their fields or periods of research. If the critical friend gave staff members a passing grade (the threshold for which was never made public) the university included them in its REF submission. Much of what the university regards as 3\* or 4\* research in history has thus been certified as such by a single external reviewer, who is not an expert in many of the areas concerned.

In all other cases, outputs would be subjected to further readings. In many cases, only one additional review seems to have been sought. The group responsible for choosing reviewers collectively lacked expertise in many areas of the department's research and did not consult with colleagues on appropriate reviewers for their work. If there were any clear guidelines for reviewers to follow, they were never made public. In one case, a Freedom of Information Act request uncovered evidence of the history research director meeting with an external reviewer to discuss an output before that reviewer had produced his or her appraisal. That such a conversation took place at all undermines any claim that this is a genuine process of independent review. The reviewer's eventual conclusion is full of ironies: "As a piece of research, this is without doubt fiercely intelligent and stimulating, if

rather demanding and non-conformist. But as a potential REF output, it is very risky, in its current form at least." Such procedures offer faculty members less protection against bias or error than the university offers its undergraduate students, whose scripts are anonymised and grades moderated by second markers and external examiners.

It is not surprising that so capricious a reviewing process leads to some bizarre outcomes. While one person was submitted on the basis of three forthcoming pieces that the university obligingly published as "working papers" on its e-prints repository to meet the REF deadline, another was excluded because an external reviewer reduced the critical friend's 3\* grade on one article to 2\*. External reviewers' grades seem routinely to have been preferred to the friend's (rather than the two being averaged or a third opinion sought), which begs the obvious question of why, if the friend's judgement was deemed so fallible in these cases, was it allowed to play so powerful a role elsewhere? This particular article had been accepted for publication in one of the world's top English-language history journals. Another colleague was excluded on the basis of the same portfolio of publications that had gained them promotion to a personal chair. For professorial promotions, the university requires a minimum of six external referees with international standing within the candidate's subject area, whose views, according to its 2014 guidance *Promotion to Readership and Personal Chair: Procedures and Criteria*, "can be especially useful in assessing the contribution of the candidate to and their standing (national and international) in scholarship and research".



The only permitted grounds for appeal against the university's staff selection decisions were procedural irregularities (which is difficult, when the procedure in question has never been made public) or discrimination as defined by the 2010 Equality Act and related legislation. In some cases, less than a week was allowed to prepare an appeal. The first stage of appeal, to the head of department and associate dean (research), breached Hefce's rule that "the individuals that handle appeals should be independent of the decisions about selecting staff", set out in its 2011 document *REF2014: Assessment Framework and Guidance on Submissions*. But most importantly, what could not be appealed were the substantive judgements on the basis of which, allegedly, individuals were excluded from the REF. "The judgements are subjective," explains Lancaster's *REF 2014 Code of Practice*, "based on factual information. Hence, disagreement with the decision alone would not be appropriate grounds for an appeal." This is the ultimate Kafkaesque twist. The subjectivity of the evaluation is admitted but only as a reason for denying any right of appeal against it.

This did not stop the dean of the faculty of arts and social science from advising heads of department to take personal responsibility for annual performance development reviews for all REFugees (as they have sardonically become known at Lancaster) in the future. Needless to say, my department is not a happy work environment right now. It will take years to repair the broken trust.

Is the REF game worth the candle? It depends for whom. The UK's academic establishment fought a tenacious campaign to retain this travesty of peer review in the Hefce consultations of 2007-09 despite its manifest inadequacies as judged by international norms. This suggests to me that informing QR funding decisions has long since ceased to be the principal objective of the REF. Had it been, appropriate metrics would surely long ago have been adopted with a big sigh of relief across the sector. This behaviour is inexplicable unless the Establishment had some considerable stakes in the process itself. Those stakes, I believe, had nothing to do with the merits of the REF as an exercise for evaluating research.

Speaking to *THE* ("[Evolution of the REF](#)", 17 October 2013), Sir Peter Swinnerton-Dyer, the architect of Britain's first "research selectivity exercise" in 1986, argued that Britain's RAE/REF regime long ago ceased to be a "tolerable process" for allocating QR funding. He is surely correct. "The rot really set in," in his view, when "vice-chancellors ceased to see the RAE as a funding mechanism" and regarded it instead as a "free-standing assessment of research quality" that would be "useful as a means to get rid of people not doing any research".

For Eastwood, on the other hand, this is the real point of the whole enterprise. "The RAE has...been the key instrument for performance management in institutions," he wrote in *THE* in 2007. "To this extent, the RAE has done more than drive research quality; it has been crucial to modernisation." He, too, is right.

What may be intolerable as a mechanism for funding allocation and indefensible as a process of research evaluation may work very well indeed as a disciplinary tool for university managers, not to mention a wonderful means of self-perpetuation for academic elites. The REF is an apparatus of empowerment (of some) and subordination (of others). It allows the activities of individual academics to be brought under an unprecedented degree of institutional control. This is a powerful repertoire of legitimation; one against which it is difficult to argue without appearing to engage in special pleading. The very laboriousness of the process is an earnest of its high seriousness. The rituals of the REF punctuate British academic life, lending a stately pomp and circumstance to what might otherwise be seen as no more than a vulgar bit of bean-counting – even if, at the end of the day, many of us can see (but are too cowed, cowardly or self-interested to admit) that the emperor has no clothes.