

# **Proposals for the Research Excellence Framework – a critique<sup>1</sup>**

**Bahram Bekhradnia**

## **Introduction**

1. The UK higher education funding bodies have published their proposals for the design and conduct of the Research Excellence Framework (REF), which will replace the Research Assessment Exercise (RAE) as the method by which research will be assessed for the purpose of the allocation of their research funds. This is not just a technical matter. Something like £10 billion will be allocated as a result of the evaluations that are conducted, and this money is allocated highly selectively according to quality; moreover, the nature of the assessment process inevitably has a strong impact on the behaviour of universities, the academics who work in them and the nature of the research that they do.

2. It needs though to be borne in mind that unlike the research councils, the funding councils do not fund research activity directly. They fund universities, and that includes the capacity and infrastructure to conduct research. Most of their funds are allocated according to teaching criteria, and a smaller amount according to research criteria. Of course, those universities that receive more funds by virtue of performing well against the research criteria are able to invest more in research and indeed if they do not, or if they do but anyway perform poorly in the next exercise, they will receive less funding in the future. But it is not research activity as such that is funded by HEFCE. The REF proposals have to be seen in that context.

## **Background and description of the proposals**

3. In the 2006 Budget statement and accompanying document 'Science and innovation investment framework: next steps', the then Chancellor of the Exchequer announced the Government's intention to replace the Research Assessment Exercises (RAE) with an assessment process based on measuring how much Research Council and other external income universities had earned. The principal reason given for the change was that the RAE placed excessive burdens on universities, and was expensive.

4. There were a number of grounds on which the Government's original proposals for assessing and funding research were criticized at the time, the

---

<sup>1</sup> This report has been published to coincide with HEPI's conference on *Research Funding and Assessment: the Future*, held at the Royal Society on 14 October 2009 and generously sponsored by Elsevier.

greatest concern being that there was no measure of quality in those proposals. While the Research Assessment Exercise had been based on peer review – the assessment of the quality of research by experts in the disciplines concerned – peer review was entirely absent from the proposals: assessment was to be entirely quantitative. Nor was there any evaluation of the likely impact on academic behaviour as researchers came to understand that the only way of securing public money was to win private grants and contracts.

5. In November 2007 the UK higher education funding bodies responded to these criticisms with a consultation paper that retained quantitative indicators as the basis for the new system – including the measures based on external income that had previously been proposed – but proposed also that citation analysis should be added as a core feature, on the basis that citations provided a measure of quality. The behavioural concerns that had been raised about the earlier proposals were not addressed, nor were the likely behavioural consequences of the addition of citation analysis to the assessment process. Most important, they proposed a form of ‘light touch’ peer review for the non-science based disciplines, but peer review was still not part of the process that was proposed for the remainder.

6. The process now proposed is radically different, and will recognizably be a development of the previous Research Assessment Exercises. Most important, as with the Research Assessment Exercise, it is now proposed that all assessments will be undertaken by peer review: panels of academics and others competent to conduct the assessments will review outputs that are submitted by universities (generally academic papers) and make judgments about quality more broadly. The major new factor is that the assessment of ‘impact’ is now an important separate and explicit element – impact had not been mentioned in the earlier proposals. That has been added as an entirely new element that it is proposed should have major bearing on the overall quality assessment of a submission's quality. And the proposals are clear that by impact is meant impact beyond the academic environment – specifically economic and social impact.

7. It is proposed that a 5-point marking scale should apply as before – 4\*-1\*, and unclassified. The evaluation will have three components (based on assessments of the quality of outputs, on the impact of the research, and on the research environment of the unit that is submitted for assessment) and each will be assessed separately and then combined into an overall score. It will be this overall score that is used for funding purposes. Analysis by Evidence Ltd, and soon to be published by HEPI, shows how only a very small amount of the research published in the UK ranks among the world's most highly cited work. Assuming that there is some relationship between citations and quality (which is contested, but which is a reasonable supposition at an aggregate level), then this represents research that stands above anything else, and the present scoring system does not allow this to

be identified. The consultation paper acknowledges the criticism that so many submissions were graded 4\* in 2008 that the top grade failed to distinguish this truly outstanding research from the rest that scored the top grade. It proposes therefore that the 4\* grade should be redefined to include only such world-leading work. That would be one approach (though the proposed definition does not really acknowledge explicitly what a small number of outputs would be expected to meet this standard), but would lead to bunching further down the scale. Another approach – not proposed, but one that should be considered – would be to add a category at the top of the scale – a 5\* grade.

8. Each Research Assessment Exercise has built on and developed its predecessor, some with significant changes. For example, the 1992 exercise reduced the number of panels to 69 from more than 150, and evaluated applied research separately. In 1996 the publication count was dropped (in the 1992 exercise the number of publications per researcher was required to be submitted, giving rise to concerns that the exercise was giving rise to a ‘publish or perish’ culture). In 2001 the number of research outputs submitted was reduced to 4, and the most significant change occurred in 2008 with the new profiled grading system, which represented a big improvement on the sharply delineated step scoring that had preceded it, but which led to some reduction in the concentration of research funding because it more sensitively recognised research strengths that were previously hidden.

9. The present proposals represent a continuation of that tradition. The underlying process is recognizably built on the RAE, and in particular is based on peer review and not on metrics as had originally been proposed, but with significant changes. HEPI’s December 2007 report<sup>2</sup> explained that there was a fundamental difference between metrics informed by peer review, which characterised the funding bodies’ previous proposals, and peer review informed by metrics. These new proposals represent a fundamental and welcome shift to the latter.

10. Those quantitative measures that formed the basis for the original proposals (Research Council and industry income in particular) have been dropped, and are barely mentioned in the proposals (their only mention is as one of a number of indicators that might be indicative of impact and of the research environment). However, citation information is retained, but as an optional feature that the panels may call upon to inform their judgments. Although in principle panels have always been able to call for specific data to inform their judgments, the innovation in these proposals is that citation information is systematized into the process – and that too

---

<sup>2</sup> ‘Evaluating and funding research through the proposed Research Excellence Framework’, quoting the June 2006 report ‘Using metrics to allocate research funds: initial response to the Government’s consultation proposals’.

is something that was proposed in the 2007 HEPI report. The consultation paper is open about how citations might be used, but offers the possibility of citation information being provided for each output (i.e. each of the publications submitted). One of the criticisms of HEFCE's previous proposals for the use of citations as a measure of quality was that they only provide meaningful information at high levels of aggregation, and if there are doubts about whether citations provide meaningful information at the level of the department, then they certainly do not provide it at the level of the individual paper. This proposal should not be implemented.

11. The consultation paper contains great deal of detail, some of it important. Much is in common with previous practice, and in this report we look closely at four of the most significant proposals where further consideration is needed, and it is to be hoped modifications will be made:

- a. The proposals for assessing impact;
- b. The proposal to reduce the number of panels by half;
- c. The arrangements for determining the assessment criteria that panels will use to assess quality;
- d. The implications of the proposal to assess the research environment as a separate element.

12. But first, one general observation is offered. An important driver for these proposals is to simplify the assessment process – that in itself is admirable. But it is worth bearing in mind the reasons for some of the features that were added to what was originally a rather simple research assessment exercise. Many – in fact virtually all – of the refinements of the RAE (the rules for the submission of staff who had recently moved, for example, and for early career staff, etc) were introduced in order to make the process fairer and more sensitive, generally in response to suggestions from the academic community. So while the desire to simplify the process is understandable, if that is at the expense of making the system less robust and less fair, then this is unlikely to commend itself. That was one of the criticisms of the previous proposals for 'light-touch' peer review in the non-science and engineering disciplines. The need for a process that is robust and acceptable to key stakeholders is recognized in the most recent report, which, reasonably, remarks also that this has to be balanced against the need to operate efficiently and to avoid undue complexity. However, a process that is not acceptable to those being assessed is unlikely in the long run to be a success.

13. There is no suggestion that any of the proposals to simplify the process are driven by financial considerations, although the original proposals identified cost

savings as the principal motivation for the proposed changes. Indeed the consultation paper refers to a report by PA Consulting that assessed the total cost of the 2008 RAE (the direct costs of conducting the exercise and the indirect cost to institutions) at £56.7 million, which is around 0.5 per cent of the total amount of money that will be allocated on the basis of its results. This compares with the cost of the Research Council processes, which are estimated by RCUK to be about 20 per cent of the money the Research Councils allocate. The RAE was not an expensive system, and reducing cost should not be an important consideration in devising the new REF.

## **Impact**

14. The biggest innovation in the proposals for the future is the introduction of 'impact' as an explicit and separate element in the grading system. Whereas previously panels were able in their assessments of overall quality to take into account the impact that the research that was being evaluated had had, they are now required to assess this separately, and the funding bodies' initial proposal is that 25 per cent of the final score that will be used for determining funding will be based on the impact score. Impact is here defined as impact on the economy and on society more generally, but explicitly excludes academic impact. Nor is there any acknowledgement that the impact criterion may be more appropriate in some disciplines than others. Indeed, there is an explicit statement that it is expected that in all disciplines the impact factor will count for a similar amount.

15. This proposal needs to be handled extremely carefully. While it is understandable that those in government who provide funding for research wish to see some economic and societal benefit, it seems unduly limited not to be able to value outstanding research that, for example, may change the face of a particular, but perhaps narrow, aspect of an academic discipline as highly as other research whose academic impact may be less, but whose societal impact is greater. Despite the fundamental continuity with previous exercises, the extent of the change that the introduction of 'impact' introduces is significant, and greater than any changes made previously. In the same way as the 1992 changes gave the message that it was not sufficient to produce good research if it was not communicated widely through publication, so the addition of 'impact' as a requirement to achieve top grades will mean that producing and disseminating excellent research will not be sufficient if it cannot be shown to have had impact beyond the academic community. In the same way as the 1992 changes fundamentally altered both the productivity and the publication behaviour of academic staff, so the proposed impact requirement will influence behaviour in ways that can only be speculated about at present.

16. While it is entirely legitimate that this should be so, those implementing the changes need to be confident that the behavioural changes that will follow are

anticipated and benign – concern has already been expressed, for example, about the likely willingness of institutions to submit early career researchers where they might make up a high proportion of the submission. Despite drawing the definition of ‘impact’ very widely, to include economic, social, public policy, cultural and quality of life impacts, aspects of the proposals need to be rethought. As they stand at present, the historian that does good but not extraordinary work on Henry VIII could be valued more highly, if he presents a television series on the subject, than his colleague who spends 17 years producing a book that changes the way that historians see their subject, but which does not have wider impact beyond the academic community. The funding bodies need to be careful to avoid the charge of Philistinism. In particular, while it is entirely legitimate to value and reward impact in all disciplines including the humanities, it does not follow that either the scope for wider impact or the benefit of such wider impact is equal in all disciplines. Nor is it obvious why outstanding academic research cannot be assessed at the highest level in its own right. The requirement to demonstrate impact to achieve the highest scores should not be applied uniformly and rigidly.

17. Beyond the point of principle, there are, as the consultation paper acknowledges, serious methodological questions about the assessment of impact – the timescales, for example, for research to have an impact, and the attribution of impact to a particular piece of research. The funding bodies propose, sensibly, to investigate these issues and the methods for assessing impact more generally. Pending these investigations, and while the wider academic community debates the implications of a separate impact assessment, as well as the proposal that this should be applied equally across all disciplines, it seems unnecessary, and probably unwise, to place a weight as heavy as 25 per cent of the total score on an untested and experimental process. It needs to be remembered that the REF is a process that will determine the allocation of billions of pounds, and such a process needs to be as robust and credible as possible.

18. One incidental issue that concerns the assessment of impact is that the process that is proposed for universities to demonstrate impact will put a serious additional burden on them. The consultation paper rightly concludes that no robust quantitative measures of impact are available, and that therefore the assessments will need to be based on the judgments of peer review panels, informed by evidence provided by universities. For each submission it is proposed that in addition to an overarching impact statement each submission should also include one case study for every 5-10 staff submitted. If impact is to be assessed then something like this may be inevitable. However, in an environment when the funding bodies are seeking to reduce the burden that they impose on universities, and where the RAE was blamed – very largely unjustifiably, but this was claimed as one of the reasons for its reform – for the burdens it imposed on universities, the

introduction of impact as a factor in the REF will almost certainly substantially increase that burden.

## **Number of panels**

19. It is extraordinary that, in this otherwise open and accommodating consultation paper that contains a number of propositions that are made for comment and for alternative proposals, the only thing largely ruled out is to retain a significantly greater number of units of assessment than the number proposed – 30, compared to the 66 in 2008. The consultation paper says that the funding bodies are “committed to substantially reducing the number of UOAs... even if this means combining relatively discrete fields into a single panel”. Yet the arguments against a larger number of units of assessment are not at all compelling. Two reasons are put forward:

- a. Universities under the old system sometimes had a choice of panels to which to submit a particular unit. That is undoubtedly true. With 66 panels there would have been more choice at the margins than with 30 panels, but for the great majority this issue did not arise.
- b. Some panels had relatively light workloads. That is also true. The proposed solution is to increase the workload of some of the others!

20. Although not put forward as a reason for reducing the number of panels, a further consideration is that the greater the number of panels, the greater the difficulty in achieving consistency, particularly consistency of marking. It is always going to be difficult to ensure that, for example, a top grade in philosophy has identified work of the same outstanding quality as a top grade in physics.

21. All these are true, but none seems a particularly compelling reason for reducing the number of panels. Against these are some rather serious arguments about the integrity of the exercise. One of the strengths of the Research Assessment Exercise has always been that those being assessed have – sometimes grudgingly and often with grumbles – recognised the competence and the authority of those by whom they are assessed. Sometimes, new disciplines have emerged, and new RAE panels were created to recognize them (for example, biotechnology) and other disciplines were amalgamated in the RAE (for example botany and zoology into biological sciences, as eventually was biotechnology when the academic imperative for a separate discipline subsided) but always in response to academic considerations and specifically in order to maintain the integrity and credibility of the exercise.

22. It is true that one consequence of reducing the number of panels would be to reduce the number of small panels, but only by increasing the workload of others.

The workloads of many of the panels in 2008 were already, as the paper acknowledges, very great, and panel members themselves have expressed concerns about the sometimes perfunctory and superficial way they had to assess work, in some cases without reviewing in detail the outputs submitted. The proposal to reduce the number of panels would, on average, increase the amount of work that each remaining panel had to review. That is the reason for the proposal made in the consultation paper possibly to reduce from 4 to 3 the number of outputs submitted, but that would be at the expense of even less rigorous assessments. That would seriously undermine the validity of the exercise and anything that risks undermining its integrity and credibility should be avoided.

23. It is therefore puzzling that the funding bodies appear to be so very adamant on this point. “Substantially reducing” the number of panels may be the right thing to do. And as so often there will be a trade-off between the disadvantages of different approaches – a more balanced workload between panels, for example, and the academic integrity of the assessments. These should be tested with the academic community, and it is to be hoped that those responding to the consultation – particularly the subject bodies – will not be inhibited from commenting for or against the proposal to halve the number of subject panels.

### **Assessment criteria**

24. The intention of achieving greater consistency between panels is admirable, and it is encouraging that the experience in 2008, of main panels that oversaw the work of sub-panels, is judged to have had this effect. Whether the proposed reduction in the number of main panels to as few as just four will result in coherent units needs to be judged by those with knowledge of the subjects concerned – that should be an academic, not a bureaucratic judgment. The proposal, though, that risks undermining the REF as an exercise where academic judgments are made by those competent to make them, and replacing these with bureaucratic judgments, is that panels will be closely fettered in setting the criteria for their judgments.

25. The consultation paper proposes that panels’ discretion in setting the criteria for their assessment of the quality of research outputs should be limited to measures to manage variable workloads, and whether, and if so how, they will use citation information. The reason offered for this is to improve consistency between panels. This is understandable, but to insist that panels in quite different disciplines adopt identical criteria for assessing quality in their subject risks sacrificing academic integrity and coherence for bureaucratic simplicity. How this works in practice remains to be seen, and may not be an issue. What is essential though is that there is acknowledgement that different measures of quality, and criteria for assessing this, are appropriate in different disciplines – and that these need to be the subject of academic judgments made by those conducting the assessments and endorsed by the academic community that is subject to the assessment. One of the



strengths of the RAE was that the academic community in each subject was consulted about the criteria by which they were to be judged and had some degree of ownership of that. There seems no reason for the new arrangements to depart from this.

## **Research environment**

26. The justification for including research environment as an element in the assessment is to ensure that more funds go to units that work in a sustainable environment. The logic behind this needs to be tested. The research assessment procedure – whether the REF or the RAE before it – is one that allocates funds for the future based on performance in the recent past. However, if the environment is not one that can sustain high-quality research then the unit concerned will not perform highly next time, and will lose funding. Rewarding the “research environment” must not be allowed to become a justification for rewarding those who have always had substantial research funding in the past, regardless of their present quality.

27. This element in the method must not become a covert way of increasing the concentration of funding in those universities that have previously been better funded for research, at the expense of those carrying out better research now. In fact, it is arguable that those who have performed highly without a strong research environment have performed exceptionally, and should receive additional funds to help them build up the research infrastructure to enable them to continue to perform well and develop further in future.

## **Conclusion**

28. A number of criticisms of the funding bodies’ proposals for the new REF have been made in this report, and suggestions for improvements offered. However, it would be quite wrong to lose sight of the immense improvement that these proposals represent compared to all those proposals for the REF that have been made previously. The previous proposals – and particularly the original proposal to rely exclusively on quantitative indicators would have been seriously damaging to the health of the research base and would have had unpredictable, but almost certainly harmful, effects on the behaviour of universities and their staff. The present proposals have reverted to peer review as the basis for the assessments, which will continue substantially to be based on the quality of the research submitted. There are a number of ways in which the proposals can be improved – some of them significantly so – but fundamentally, the approach is a sound one and the funding bodies are to be commended for having responded to the widely expressed concerns about their previous proposals.