

Evaluating and funding research through the proposed “Research Excellence Framework”

Bahram Bekhradnia

Introduction

1. In November 2007 the Higher Education Funding Council for England published a consultation paper concerning proposals for the future assessment and funding of research. HEFCE’s new proposals are a considerable improvement on the original proposals from the Government to replace the RAE with measures of research income as the basis for funding research in the future. The original proposals excluded any measure of quality as a basis for funding research and were widely criticised at the time.

2. In its response to the consultations on the original proposals HEFCE recognised the criticisms and proposed to use bibliometrics – and in particular citation analysis, which measures the extent to which an article is cited by others – as a measure of quality. The present proposals on which HEFCE is consulting – the Research Excellence Framework (REF) – are based on two reports by experts on citation analysis – a group based at Leiden University, and Evidence Ltd, a UK based company specialising in research policy and bibliometrics in particular. HEFCE also proposes to retain measures of research income as an ingredient in its funding of research (QR). The majority of HEFCE’s proposals – and certainly the greatest changes – are concerned with the assessment and funding of science and engineering related disciplines.

3. It is worth recalling the original aims of the Government’s proposals to replace the RAE, that HEFCE has inherited and now seeks to operationalise: to reduce the burden of assessment, to reduce cost and to find a new mechanism for providing QR funding that is less burdensome and costly than the RAE, that would nevertheless allow QR to be allocated by reference to the quality of the research that is carried out, and to recognise and reward quality wherever it is found.

4. The original proposals failed on all counts. They did not measure quality and they would have led to allocations of QR that bore little relation to the quality of the research, they would have greatly increased the cost of assessment, while shifting it from the centre to institutions and to individual academics, and they would have had a seriously distorting effect on behaviour. These objections were enumerated in HEPI report 25 “Using metrics to allocate research funds: initial response to the Government’s consultation proposals”¹, were widely repeated during the consultation on the

¹ All HEPI reports are available at www.hepi.ac.uk

original proposals and were implicitly accepted by HEFCE in its revised proposals. In particular HEFCE was then, and remains, clear that in future, as in the past, the new arrangements should assess quality and that QR funds should be allocated on the basis of quality.

5. From HEFCE's new consultation proposals it is clear that it has decided that quantitative indicators will be at the heart of the research assessment and funding system, and in particular that citation analysis will be a core feature. Although two options are offered for consultation, fundamentally they are similar and both take metrics and citations as a given element in the new arrangements. This report therefore also takes these as a given, and focuses on the consequences and implications of the approach proposed, and discusses some modifications that will be required in order to improve the proposals and make them fit for the purpose for which they are intended – assessing research quality and providing funding differentially, based on the quality of research.

6. The first thing to be said is that the proposals put forward by HEFCE are very much less potentially damaging to the UK's research base than the original proposals from the Government in March 2006. As was pointed out in the HEPI report in June 2006 "Using Metrics to Allocate Research Funds" a funding system based entirely on the amount of income brought in risks distorting the nature of the research activity carried out in universities and increasing the cost of compliance. All competitive funding methods, including the Research Assessment Exercise and HEFCE's new proposals, are bound to influence behaviour to some extent but the original proposals would have:

- Penalized academics who did research without a "customer" , making it difficult for them to finance research in fields which are, perhaps temporarily, unfashionable, or whose conclusions are likely to be unwelcome to funders of research
- Greatly increased the incentive to apply for Research Council grants, thereby increasing the failure rate even further than the current level of over 70 per cent and thereby increasing the cost
- Had a major impact on recruitment policies: staff with a track record of winning research grants would have become increasingly in demand, and others (in particular young staff or those returning from a career break) increasingly unattractive.

Quality Assessment?

7. One major problem with the new proposals is that citation analysis does not measure quality, so there must be real doubt about whether it can be used as a basis for allocating QR differentially according to differential quality. Indeed, the Leiden group that HEFCE retained to advise about the use of

citations is quite explicit in its report that citation analysis does not in fact measure quality. Section 3.3 of its report contains number of questions and answers, one of which is:

“Does citation analysis provide an objective measure of research quality?”

8. The answer to that question is unequivocal:

“No. Bibliometric indicators measure impact rather than quality.”

and elsewhere the report explains that “impact” is the extent to which the research is noticed by colleagues, but that that is different from the value of the research.

9. This is awkward in light of the commitment of HEFCE to continue to allocate research funds on the basis of quality. However, the Leiden report is helpful in this respect. At several points the report emphasizes that citations alone are unsuitable for evaluating the quality of research, but that used in conjunction with peer review they are a powerful tool. On page 31 for example the report states:

“Thus the outcomes of bibliometric analysis must be valued in a qualitative, evaluative framework that takes into account the contents of the work.”

10. It goes on to state on page 32:

“Therefore, as a general principle we state that optimal research evaluation is realised through a combination of metrics and peer review. Metrics, particularly advanced analysis, provides the tools to keep the peer review process objective and transparent. Metrics and peer review both have their strengths and limits. The challenge is to combine the two methodologies in such a way that the strengths of one compensates for the limitations of the other.”

11. And on page 40 it says:

“It is important to develop research information systems that cover more than just publication data – for instance, number of full-time equivalents (FTEs) for research, mission and objectives, age distribution data of research personnel, data on facilities, supporting personnel, data on funding. Such an extension to non-bibliometric data or, in general terms, institutional context data, allows us to gain a deeper insight into important aspects of research performance, particularly the ‘efficiency’ of research and its social characteristics.”

12. This is endorsed elsewhere by Evidence Ltd:

"Indicators do not replace expert decision making. They do not, by themselves, tell you where to put your money. And if you make the mistake of using them for those purposes, then the extent to which they remain good indicators is undermined²."

13. So HEFCE's appointed experts on citations warn first that citation analysis does not measure research quality and second that HEFCE should not use citation analysis alone to evaluate research but that it should be used to inform peer judgements. That is exactly the proposal contained in the report produced by HEPI in June 2006 which stated:

"Citations... could have a role in a streamlined peer review process... by informing peer review panels."

and concluded

"An alternative [to the Metrics proposal] is available – not the continuation of the RAE, but a system based on both peer review and metrics."

14. Although citation analysis does not measure quality and there are warnings from the experts that it should not be used except to inform peer review, there are circumstances in which it might be used without peer review to drive research funding. If it is found that the results of citation analysis closely mirror those of the RAE, then although it does not *measure* research quality it may be thought closely to *reflect* it nevertheless. If that is so then it may truly be the case that citation analysis will prove a robust and cheaper way to assess research and allocate QR on the basis of quality. But to achieve this, a close match between the results of citation analysis and the RAE will be essential; and that will have to be at a far more fine-grained level than the six broad groupings proposed by HEFCE, if in future excellent research is to be funded wherever it occurs, as has been the case in the past, and as remains HEFCE's intention.

15. The consultation paper understandably contains no simulations to show how assessments or allocations will fall under the proposed new system. The new research assessment exercise will produce results in a year, and it might seem premature and otiose to compare the results of a citation analysis with research assessment exercise results that are six years old. However, until such a comparison has been carried out, judgement must be suspended as to whether the results of citation analysis come close to mirroring the results of quality assessment and so whether citation measures will result in funding being directed towards the highest quality research. If they do, then although citation analysis does not *measure* quality it can be assumed to *reflect* quality, and that might provide a sufficient basis for funding. If they do not, then the

² Research Fortnight (November 2007), 290, pp. 18-19

basis proposed cannot be used to allocate funding differentially according to quality.

16. Although it would be unreasonable to expect the results of the Research Evaluation Framework to mirror exactly the results of the RAE, a very close match will be essential to create confidence that the system does what has been claimed. The RAE measures research quality. There has been criticism of the RAE on many grounds – that it is expensive, that it distorts research activity, that it unduly influences recruitment – but there is seldom any doubt that it is a process that enables assessments of quality. If we can be confident that the RAE outcomes are measures of quality, then it would be reasonable to expect that any other mechanism for assessing research that purports to make judgements about quality would have similar outcomes. We need to suspend judgement about whether the proposed citation mechanisms will do this until we have seen their results and compared them with the 2008 RAE results.

17. A report produced by Evidence Ltd for Universities UK does include some analysis of correlations between citation metrics and RAE grades (Figures 3 and 4)³. This shows that there is a good relationship at a broad level: citation impact averaged across all units increases progressively⁴ with an increase in the peer reviewed grade. The data also reveal, however, that the variance between units within any grade is very high. Indeed, the scatter is such that a unit with a citation impact close to world average would be equally likely to have been awarded a 4 or 5 grade, or even to be an exceptional 3a or trailing 5*. It is no wonder that HEFCE does not propose to use citation analysis at the level of the department.

18. If there is no sufficient match in the post-2008 analysis, and if HEFCE truly wishes to allocate QR resources by reference to quality, then it will have to modify its proposals, in a way that has in fact been urged on it by its own advisors. HEFCE will need to use citations as an important input – in some cases perhaps the dominant input – in the work of the expert panels that it proposes to establish as part of the REF. It would be the task of these panels to interpret the citation analysis – but also any other information that seems relevant – in making assessments of research quality.

The role of expert panels

19. HEFCE has proposed the creation of expert panels in the context of the new Research Evaluation Framework, but it also proposes that the role of these panels should be limited to advising “on the selection and use of

³ *The use of bibliometrics to measure research quality in UK higher education institutions*, Adams, J., (Universities UK: London, 2007), p. 44.

⁴ Except for the average for all grade 2 departments which is higher than the average for all grade 3bs.

indicators within the framework for all disciplines". It will not take a great leap to give these panels the role of deciding on the ratings to be awarded, based largely, perhaps, on the citation analysis, but partly on other factors that they decide are appropriate. Such a modification of the HEFCE proposal will have the twin merits of making the system for the sciences much closer to that proposed for the humanities, and also bringing it into line with the recommendations of its chosen experts. Indeed, it seems inconceivable that HEFCE will fly in the face of its appointed experts and use citations in an inappropriate way. It would also mean that research can continue to be funded on the basis of relative quality, which is what HEFCE has intended all along.

Citation analysis

20. The HEFCE proposals are based on two reports produced by technical experts in the matter of citation analysis, and that is one of the problems. These will be highly mechanistic assessments, and therefore objective, and welcome to that extent. But key decisions – for example what subjects to group together⁵, what needs to be normalized⁶ and the values of the normalization, and what period to allow for the purpose of counting the citations, all of which can critically affect the results – will in the future be largely in the hands of technicians. This impression is heightened by the many issues which the consultation paper says are problematic, and on which HEFCE is still taking technical advice. As far as many individual academics are concerned – and their institutions – this element of research assessment and funding may in future be a black box. The assessment of quality – or rather what replaces it – is to be taken out of the hands of peers – a process that is known and understood in the academic world – and put in the hands of technical experts.

21. This may not matter. As with the RAE, academics and institutions may be able to work out the relationship between their inputs and the outcomes of the evaluation. In so far as they do not, then that will reduce the extent to which they are able to play games in order to maximise their standing in the exercise. That need not be a bad thing, but in that case it will be difficult to build confidence and buy in to a process that is not widely understood. More likely though is that they will work out how the new citation system works and how it can be manipulated to their advantage. If so, HEFCE can expect some fierce arguments about some apparently arcane issues that will decisively impact the REF results. In that case also, they can expect academics and institutions to modify their behaviour in ways that are as yet unimagined, in the same way as occurred with the RAE. In the consultation paper HEFCE

⁵ This complex and potentially contentious issue is discussed in the Evidence report for Universities UK: section 4.7 from page 27.

⁶ Because different discipline groups have different citation patterns citation rates will need to be normalised. That makes sense, but it will add a layer of complexity.

asserts its confidence that these will not undermine “the quality and impact of the research base”, but it would probably be more prudent to suspend judgment about the behavioural impact at present.

22. In passing it is worth mentioning that one of the criticisms of the RAE has been that it is retrospective, and effectively allocates QR funds on the basis of evidence of research quality that is up to six or seven years old⁷. That will be even more so in the case of citations, as recently produced research is bound to have fewer citations than research of the same quality produced years before. Indeed, the Leiden report proposes using a timeframe of 8-10 years (i.e. citation analysis of research produced 8-10 years ago). There must be real doubt about the wisdom of allocating research funds for the next eight years on the basis of an evaluation of research conducted up to 10 years ago. For the sake of the validity of the exercise it is important to find a means of conducting more up-to-date evaluations. If that cannot be done on the basis of citations, then that alone provides grounds for reconsidering HEFCE’s proposals.

23. Another criticism of the RAE is that it is expensive – a claim that has been found to be untrue or at best a gross over-simplification. The Government reviewed separately the RAE and the Research Councils’ grant allocation processes, and found that the RAE cost less than 1 per cent of the grant allocated compared to the Research Councils whose processes accounted for around 20 per cent of the grants they gave. In both cases the great majority of the cost (around 90 per cent) was not direct cost, but the cost to the universities and individual academics – in the case of Research Council grants arising largely from the fact that more than 70 per cent of applications were unsuccessful, and in the case of the RAE from the preparations and tactical analysis they did prior to submission.

24. It is too early to say what the cost of the citation analysis component of the REF will be, but given that much of the data will come from published sources, it is reasonable to assume that the direct costs will be less⁸. However, as stated, the majority of the cost of the RAE does not arise from the assessment itself. And there is no reason to suppose that universities and their leaders, as well as individual academics will not put as much effort into trying to ensure they perform as well in the new competition as the old.

Assessment of the humanities

25. In its consultation paper HEFCE announces its intention to continue with peer review, albeit with a “light touch”, as the means of assessing research in

⁷ Although as a matter of fact, the majority of the research submitted for assessment was produced towards the end of the assessment period.

⁸ But there must be a worry that HEFCE will effectively be handing control of a key feature of its funding method to a private – effectively monopoly – concern, and so the main direct cost of the exercise will in future be out of its hands.

the humanities. That is greatly to be welcomed. It was never plausible to think that humanities research would be susceptible to assessment by reference to purely quantitative measures. So in future we will have a system whereby funding for the non-science subjects is allocated according to the quality of the research that has been carried out, while funding for the sciences is allocated according to the extent to which research is referenced by other academics (for whatever reason) and the research income that has been generated.

26. And if a means can be found to lighten the touch and reduce the burden imposed by the assessment process in the non-science subjects then that is for the good. It remains to be seen what is meant by "light touch", but according to the consultation paper one element in this "light touch" will be a smaller number of assessment panels than in the past. Care needs to be taken over this proposal.

27. First, the number of panels in the RAE has largely been dictated by the need to ensure that academics are assessed by people whom they regard as their peers and whom they trust as competent to assess work in their discipline. The desire to reduce the number of panels is understandable – and it is not the first time that this has been attempted – but it runs the risk of undermining confidence in the peer review that will be carried out in the non-science subjects. But in any case the extent to which implementation of this proposal will reduce the assessment burden is very limited indeed. It will do nothing to reduce the burden on academics and their institutions; and on the Government's own figures, that amounts to 90 per cent of the cost of the RAE. The saving involved will be extremely small, and the risk to confidence in the exercise large.

Light touch assessment

28. The notion of "light touch" is worth considering further. Arguably, the 1989, and possibly even the 1992, Research Assessment Exercises were "light touch" – certainly they were relative to the 1996 and 2001 Exercises – and it is worth reflecting on why the touch became heavier. On the one hand, it was to counter "games playing" that threatened to undermine the integrity of the exercise, and on the other hand it was to make the process more "fair" by allowing universities to provide contextual information. In particular there were concerns about equal opportunities. So panels were encouraged to take account of staff returning after a career break, newly appointed staff who were new to the profession and the fact that people may have moved universities shortly before the exercise.

29. Metrics alone, without the intervention of peer judgement, cannot do this, as is recognised in the Leiden report, which insists that in the interests of equal opportunities citation information must be interpreted by a peer review panel. The Leiden report says on pages 36-37:

“In order to substantially improve evaluation procedures, an advanced bibliometric method needs to be used in parallel with a peer-based evaluation procedure. Bibliometric analysis will make evaluation procedures more transparent and objective, and thus contribute to the enhancement of equal opportunities.”

Evidence Ltd, in its report to Universities UK, comments on page 27:

“Equal opportunities considerations point to a number of issues where peer review is entirely capable of adjusting perceptions for particular cases but metric algorithms can make no such adjustment.”

30. It would have been possible to make the RAE far more “light touch” by reversing these changes, and it remains to be seen what measures prove possible to make assessment in the non-science subjects “light touch” but nevertheless rigorous and fair in the future. But the ambition is admirable. The risk to the Research Evaluation Framework process is that when the science subjects see their non-science colleagues being assessed by a peer-review process that is light touch they will ask why the same process could not have been applied to them. The reverse is also possible, but unlikely.

Other purposes of research assessment

31. As is recognised in the Leiden report the Research Assessment Exercise has multiple purposes, of which the most important is to ensure the allocation of resources to support the highest quality research, but others of which include the provision of management information to universities about their strengths and weaknesses at a quite detailed level. The proposal to summarize performance in just six broad units greatly limits the management information that will be available from the Research Evaluation Framework process. Although information will, apparently, be available at lower levels of aggregation, HEFCE believes that such analyses are not sufficiently robust for its purposes, and so are unlikely to be useful to institutional managers either.

Other elements of the REF

32. The proposals to allocate some of QR by reference to the number of postgraduate students and the amount of charity income are not new, and maintain current practice. These need to be watched closely – when there was no limit on the number of postgraduate research students eligible for funding, during the mid-1990s, this provided a strong incentive for universities to recruit increasing numbers of postgraduate research students, and led to a rapid rise in their numbers. HEFCE was obliged to counter this by capping the number of such students that would be funded.

33. It is HEFCE’s continued intention to include research income as a key element in the mechanism for allocating QR funding that is potentially the

most damaging of the Research Evaluation Framework proposals, although it is of note that HEFCE has not said what income will be taken into account in the REF. It is true that HEFCE inherited this from the original Government proposals and it could be that political considerations override others, but the effects of this aspect of the proposal risk badly undermining the health of the UK's research base. Different considerations apply to income from Research Councils and from industry.

Research Council income

34. Research Council income is the most burdensome and costly for academics to secure. In December 2006 RCUK – the umbrella body covering all the UK Research Councils – produced a report that concluded that their grant allocation processes accounted for as much as 20 per cent of the value of the grants they allocated. In HEPI report 28 “A Dangerous Economy”, produced in December 2006, HEPI argued that that did not necessarily provide grounds for taking resources from the costly Research Council processes and distributing them through the far more cost-effective QR process (based on the RAE). But this did provide grounds for the Research Councils to ensure that their grant was provided for those things that were not and could not be funded through QR. Research Council funding had to be distinct to justify the huge overhead.

35. The fact remains though that Research Council processes are very costly, partly because of the nugatory effort that is expended on applying for grants – over 70 per cent of grant applications are unsuccessful. Anything that causes an increase in grant applications will reduce the success rate and increase the cost – more than pro rata.

36. Even though the new HEFCE proposals water down the influence of Research Council funding on the new arrangements, they by no means eliminate it. Indeed, because one of the strengths of the citations approach is that it may be hard for academics and their managers to identify how to modify behaviour in order to succeed in that part of the allocation process, that is likely to have the effect of focusing attention on those parts of the process whose effects are understood and can be influenced by their behaviour. In our previous report we warned that “linking QR to research grant income could cost over £700 million over seven years even once the savings from scrapping the RAE are taken into account.” Since then RCUK have reported on the costs of their processes, and there is no reason to revise this figure. It is a huge figure, and of course it will be largely invisible, since most of the cost will fall to individual academics and their institutions, not to the Research Councils themselves.

37. Even if metrics based on research income are to be part of a ‘basket’ of measures for allocating QR, there seems no good reason for including grants won from Research Councils, and every reason for excluding them. Whereas

counting income from industry may give rise to behaviour that the Government wishes to encourage (discussed below), there is no equivalent behavioural benefit to be achieved from including grants won from Research Councils. On the contrary, given that the only behaviour likely to follow from this is to encourage more and more grant applications, the behavioural impact will be wholly negative. Even if Research Council income is used only to inform a quality profile without influencing funding, while that would be less damaging, it would still be likely to send messages to universities and academics that would not be in the interests of the research base.

Income from Industry

38. Unlike grants from Research Councils a case can be made for including income from industry in any 'metrics basket'. In our report of June 2006 we warned of the implications and risks of including income from industry as a metric. These included importantly that it would encourage universities to offer cut price research to industry (because universities would know that every research contract won would bring with it a significant amount of QR grant) at a time when the Government was trying to encourage universities to be more rigorous in their cost recovery. And so it would mean public money being used to subsidize private business.

39. However, this disadvantage may be offset by the policy gain of encouraging universities to form closer links with industry. That is a legitimate matter for political and policy decision, as long as the decision is taken with full regard to the consequences. It is certainly the case that the behavioural impact will be less negative than including Research Council grant as a metric.

Conclusion

40. The present proposals from HEFCE represent an advance on the original proposals for a metrics-based approach to assessment and funding based on research income. HEFCE is now seeking to identify a metrics-based approach to assessing quality, and proposes citation analysis as the means of doing so. This report has assessed the HEFCE proposals, and concluded:

- According to the experts in citation analysis that HEFCE has consulted, a citation-based approach does not measure research quality
- That may not matter if the results of citation analysis align well with the assessments of quality revealed by the RAE results
- If citation analysis does not closely mirror the results of the 2008 RAE, then the role of the expert panels will need to be extended to provide the peer review suggested by the Leiden report

- Such an extension of the role of the expert panels would have the added advantage of bringing the methods for evaluating the humanities and the sciences closer together; it would also enable equal opportunities issues to be handled more sensitively than could be achieved with a purely citation-based approach
- HEFCE proposes to carry out the analyses in six broad subject groupings, because citation analysis does not provide stable and valid analysis at lower levels of aggregation. That may produce results that suffice for the purpose of distributing funds at institution level, but that is all: it will not provide some of the more detailed comparative and management information provided by the RAE
- There are other changes needed to make the proposed system work without doing damage to the health of the research base – for example
 - Research Council income should not appear among the REF metrics
 - The proposal to reduce the number of panels in the assessment of non-science subjects needs to be reviewed
 - The evaluation period needs to be reduced – 8-10 years is far too long a retrospective period on which to base funding for the next 5-10 years.