Joint funding bodies' review of research assessment

Invitation to contribute

Executive summary

Purpose

1. This document invites initial contributions to the joint funding bodies' review of research assessment in higher education. As a basis for discussion, it identifies a number of key issues and possible approaches to research assessment. It also outlines the purpose of the review and its timescale.

Key points

2. The review will be owned by the UK funding councils and the Department for Employment and Learning in Northern Ireland (DEL NI), and be overseen by a steering group chaired by Sir Gareth Roberts. There will also be a wider consultative group.

3. This is only the first part of plans for consulting stakeholders in the course of the review, which will include meetings, focus groups and a dedicated web-site.

Action requested

4. At this stage we invite interested parties to debate the issues, including the ones identified in this document, and contribute to the scope of the review. Please send completed responses **by e-mail** to Vanessa Conte at responses@ra-review.ac.uk The closing date is **29 November 2002**.

Why are we reviewing research assessment policy?

5. On current measures of performance, UK research is in excellent health. Citation evidence confirms the strength of UK research in higher education institutions (HEIs) illustrated so dramatically by the results of the 2001 Research Assessment Exercise (RAE).

6. There are, however, good reasons to re-examine the continued fitness for purpose of the RAE. Concerns include:

- effect of the RAE upon the financial sustainability of research
- an increased risk that as HEIs' understanding of the system becomes more sophisticated, games-playing will undermine the exercise
- administrative burden
- the need to properly recognise collaborations and partnerships across institutions and with organisations outside HE
- the need to fully recognise all aspects of excellence in research (such as pure intellectual quality, value added to professional practice, applicability, and impact within and beyond the research community)
- ability to recognise, or at least not discourage, enterprise activities
- concern over the disciplinary basis of the RAE and its effects upon interdisciplinarity and multidisciplinarity
- lack of discrimination in the current rating system, especially at the top end.

7. For these reasons, we have set up a review of research assessment led by Sir Gareth Roberts, President of Wolfson College, Oxford. The review is expected to complete its work by March 2003, although there will, of course, be a need for further detailed work by the administrators of any new system.

Assessment of research quality: the context

8. In conceptual terms, the issue of research assessment might be seen as straightforward. However, to arrive at the best possible assessment process, we must first answer the philosophical question, 'what is meant by quality in research'. Is quality simply another term for intellectual excellence, or does it have other dimensions reflecting its likely impact within and beyond the research community? As we are ultimately using the information to calculate funding entitlements for institutions, this question has to include an assessment of their contribution to the development of researchers as well as their research output. Once we have the answer – or answers – to the philosophical question, we are left with the technical but non-trivial problem of designing a system, which provides a fair and accurate assessment of quality while minimising burden on all concerned.

9. These two themes, the philosophical and the practical, will run throughout this review. This document, coming as it does at the beginning of the process, tends to emphasise the former, although by no means exclusively so. We have also made a small number of assumptions which the review will not challenge:

a. The dual support system will continue. There will thus be an ongoing need for a method of allocating funds selectively. Research assessment of some description will continue to be used for this purpose.

b. The quality of research will continue to be considered in a global context. It will therefore need to be assessed at a national and international level.

10. We also wish to introduce as context three relevant factors:

a. There is, quite properly, an increasing emphasis upon the ‘people dimension’ – that is, the contribution made by institutions to the supply and development of researchers.

b. There are now public funds available to universities and colleges for knowledge transfer activities. Work is continuing to develop measures of excellence in those activities, many of which involve research services to external partners.

c. With the competition for research funding being increasingly fierce and the costs of research in many subjects increasing, there is a need to consider whether targeted help is required to enable new subjects and new fields to develop. It may (or may not) fall to the research assessment process to identify suitable candidates for any such assistance.

Our plans for the review

11. The review will be owned by the UK funding councils and DEL NI and be overseen by a steering group chaired by Sir Gareth Roberts. There will also be a wider consultative group, with access to steering group papers and remote discussions.

12. The steering group will draw up a list of broad approaches to research assessment. A team of people – led by Siân Thomas, at the HEFCE – will support the steering group and undertake development work on each approach, showing how it might work and what behavioural impacts it might have.

13. The output of the review will be a number of detailed models of research assessment and a covering report. This will be presented to the chairs and chief executives of the funding councils (and their equivalents at DEL NI) before formal consultation with stakeholders.

14. In its report, the steering group will either identify a single preferred option or suggest the circumstances in which particular models would be most appropriate.

15. The membership and terms of reference for the group are at Annex A.

Timetable

16. This is only the first part of the plans for consulting stakeholders in the course of the review. A series of public meetings and focus groups will be held from October (dates to be confirmed). In addition, the review team will be meeting individual stakeholder groups, and records of these meetings will be published. Working documents produced by the review will be published on a dedicated web-site, as will responses to this invitation (unless confidentiality has been specifically requested).

17. A formal consultation lasting 12 weeks will be launched once the review has completed its work. A provisional timetable is given below (some dates to be confirmed):

Invitation to contribute opens	4th October 2002
Website launched	11th October 2002
First steering group meeting	October 2002
Invitation to contribute ends	29th November 2002
Public meetings	November-December 2002
Focus groups	October-November 2002
Second steering group meeting	December 2002
Third steering group meeting	March 2003
Completed report	April 2003
Publication of report: formal consultation	May 2003

Informing the review

18. We invite interested parties to debate the issues, including the ones we have identified, and contribute to our review. Respondents may wish to convene focus groups or workshops and submit the formal record of their discussions. We have provided what we hope are helpful notes to stimulate debate (see Annex B).

Approaches to assessment

19. We can envisage four distinct approaches to assessment:

- expert review (including peer review)
- algorithm based entirely upon quantitative metrics
- self-assessment
- historical ratings.

20. This document invites respondents to explore each of these approaches in turn. This does **not** mean that we believe that these methods have to be used in isolation, or that respondents should feel constrained in proposing systems that employ elements of two or more of them. It is, for example, clearly possible to use metrics to inform a self-assessment or an expert review. The 2001 RAE did just that: assessment panels were obliged to consider some objective data (such as grant income), and allowed to stipulate that they would consider others (such as bibliometric data). Nevertheless, the RAE was ultimately an expert review system, because the final decision rested with the panel.

21. Each approach is discussed in detail in Annex B.

Cross-cutting themes

22. There are also fundamental issues which need to be addressed regardless of the approach taken to assessment:

a. What should/could an assessment of the research base be used for?
b. How often should research be assessed? Should it be on a rolling basis?
c. What is excellence in research?
d. Should research assessment determine the proportion of the available funding directed towards each subject?
e. Should each institution be assessed in the same way?
f. Should each subject or group of cognate subjects be assessed in the same way?
g. How much discretion should institutions have in putting together their submissions?
h. How can a research assessment process be designed to support equality of treatment for all groups of staff in Higher Education?
i. Priorities: what are the most important features of an assessment process?

23. Notes elaborating upon each of these issues are also provided in Annex B.

How to respond

24. We will assume that all respondents consent to the publication of their response. If you wish your response, or any part of it, to remain confidential, this must be clearly indicated both in the covering e-mail and the front page of the response itself.

25. What we seek at this juncture is the clearest possible sense of what matters to interested parties, so that we can place those concerns at the heart of the review and ensure that they inform the development of proposals.

26. The purpose of this invitation, therefore, is to generate ideas and insights rather than to discriminate between them.

27. The notes in Annex B set out some of the issues around each question. They are intended to stimulate discussion. We encourage respondents to challenge any underlying assumptions they discern in our presentation of the issues.

28. Please make clear at the top of the response on whose behalf it has been submitted. In particular, please indicate whether it represents the corporate view of an institution, organisation or grouping, or the private view of an individual or group of individuals.

29. Responses should include the details (name, telephone number and e-mail address) of someone we can contact if we have any queries about the response.

30. Please send completed responses **by e-mail** to Vanessa Conte at responses@ra-review.ac.uk. The closing date is **29 November 2002**.

31. We regret that it will not be possible to acknowledge or provide feedback to all respondents. However, if there are points of particular importance in your response which you wish to draw to the attention of the review team, please contact Siân Thomas at or Tom Sastry at rareview@hefce.ac.uk

Annex A

Steering group: membership and terms of reference

Membership

Sir Gareth Roberts (Chairman)	Wolfson College, Oxford
Sir Leszek Boriszeiwicz	Imperial College
Professor Vicki Bruce	University of Edinburgh
Professor David Eastwood	University of East Anglia
Professor Georgina Follett	Dundee University
Dr John Kemp	Evotek Neurosciences GmbH
Professor Fabian Monds	Invest Northern Ireland
Professor Terri Rees	Cardiff University
Professor Phil Ruffles	Rolls Royce plc
Sir David Watson	University of Brighton

Others have been approached. Further names may be confirmed in the near future.

Terms of reference

1. The review will investigate different approaches to the definition and evaluation of research quality, drawing on the lessons both of the 2001 RAE and of other models of research assessment,[1] and will advise on the future of research quality evaluation.

2. The output will be a number of models of research assessment and a short covering report to be presented to the chairmen and chief executives of the funding councils (including DEL NI). The report will either identify one preferred option or indicate the circumstances under which particular models would be most appropriate.

[1] The term ‘assessment’ is used here in its broadest sense to refer to any activity undertaken with the aim of providing information, assurance or feedback on the quality of research and associated activities and processes.

Annex B

Notes for facilitators

1. These notes are intended to guide those responsible for producing responses.

2. We have divided the topics for discussion into six groups. We hope this will be helpful to those organising discussions within their organisations or groupings. Four of the groups relate to the approaches to assessment outlined in paragraph 19 of the main document, and the fifth relates to crosscutting issues which will have to be addressed whichever approach is pursued. Group 6 prompts discussion of any topics that we have missed.

Group 1: Expert review

3. We have used the term 'expert review' to describe a system in which experts (possibly but not necessarily peers) make a professional judgement on the performance of individuals or groupings[2], over the previous cycle, and/or their likely performance in the future.

4. In such a system, assessors may make use of metrics, but the ultimate responsibility for decisions rests with them. Assessment may be undertaken entirely by peers or may incorporate others (such as representatives of user groups, lay people, and financial experts). The 2001 RAE was an example of this type of assessment.

5. A variant of this system would be a combined assessment of teaching and research.

6. Suppose the funding councils have decided that they wish to retain the judgement of experts as the cornerstone of the research assessment. They are, however, willing to consider any system, however different from the 2001 RAE, so long as that condition is met. How would you advise them?

7. In providing your advice, you are asked to consider the following questions:

   a. Should the assessments be prospective, retrospective or a combination of the two?
   b. What objective data should assessors consider?
   c. At what level should assessments be made – individuals, groups, departments, research institutes, or higher education institutions?
   d. Is there an alternative to organising the assessment around subjects or thematic areas? If this is unavoidable, roughly how many should there be?
   e. What are the major strengths and weaknesses of this approach?

Group 2: Algorithm

---

[2] A grouping might be (for example) a research group, network, department, faculty, institution or consortium.

8. Suppose the funding councils have decided to use an algorithm to assess research quality. The assessment must be 'automatic', leaving no room for subjective assessment. Metrics might include:

- measures of reputation based on surveys
- external research income
- bibliometric measures (publications or citations)
- research student numbers (or completions)
- measures of financial sustainability.

9. Assume the councils have not, however, formed a view on what metrics should be used or how they could be combined most effectively in an algorithm. How would you advise them?

10. You have been asked in providing your advice to consider the following questions:

   a. Is it, in principle, acceptable to assess research entirely on the basis of metrics?
   b. What metrics are available?
   c. Can the available metrics be combined to provide an accurate picture of the location of research strength?
   d. If funding were tied to the available metrics, what effects would this have upon behaviour? Would the metrics themselves continue to be reliable?
   e. What are the major strengths and weaknesses of this approach?

Group 3: Self-assessment

11. Suppose the funding councils have decided to pursue a self-assessment model in which institutions, departments or individuals assess themselves. A proportion of the assessments are reviewed in detail. In a self-assessment model, the assessment is made by the assessed, although its reliability may be challenged by the validators.

12. Assume the councils have not, however, formed a view on how the assessment should be structured and how self-assessments will be validated. How would you advise them?

13. In providing your advice, you are asked to consider the following questions:

   a. What data might we require institutions to include in their self-assessments?
   b. Should the assessments be prospective, retrospective or a combination of the two?
   c. What criteria should institutions be obliged to apply to their own work. Should these be the same in each institution or each subject?
   d. How might we credibly validate institutions' own assessment of their own work?
   e. Would self-assessment be more or less burdensome than expert review?
   f. What are the major strengths and weaknesses of this approach

Group 4: Historical ratings

14. Suppose the funding councils have decided to pursue a policy that gives each institution a rating on the basis of its historical performance and/or the value of its research infrastructure. Research would, in effect, be presumed to be strongest in those departments or institutions with the strongest track record.

15. The councils recognise that such an approach could only be used in conjunction with another system: there would need to be some way of identifying institutions whose performance was sharply improving or declining, even if the presumption was that the distribution of excellence would remain stable. It would also be possible to alter the share of the total pot provided for each institution on the basis of what had been achieved with the investment provided (a 'value for money' rating).

16. Assume you have been asked to advise on how such a system might work. In developing your advice, you have been asked to consider the following questions:

a. Is it acceptable to employ a system that effectively acknowledges that the distribution of research strength is likely to change very slowly?
b. What measures should be used to establish each institution's baseline ratings?
c. What mechanism might be used to identify failing institutions or institutions outperforming expectations? Could it involve a 'value for money' element?
d. What would be the likely effects upon behaviour?
e. What are the major strengths and weaknesses of this approach?

Group 5: Crosscutting themes

17. You have been asked to provide advice to the funding councils on the following fundamental issues:

a. What should/could an assessment of the research base be used for?
b. How often should research be assessed? Should it be on a rolling basis?
c. What is excellence in research?
d. Should research assessment determine the proportion of the available funding directed towards each subject?
e. Should each institution be assessed in the same way?
f. Should each subject or group of cognate subjects be assessed in the same way?
g. How much discretion should institutions have in putting together their submissions?
h. How can a research assessment process be designed to support equality of treatment for all groups of staff in Higher Education?
i. Priorities: what are the most important features of an assessment process?

18. We have elaborated on each of these questions below. Respondents may wish to use these notes as a basis for discussion.

a. What should/could an assessment of the research base be used for?

For the funding councils the immediate purpose of research assessment is to provide the information necessary to calculate funding levels. RAE ratings are, of course, used by others, including institutions themselves, for a variety of purposes.

What should research assessments be used for and by whom? Should the funding councils be more explicit about what the information produced by the exercise means, and what it ought to be used for? Should we look to design a research assessment process with the explicit aim of providing reliable management information for academic communities, institutions and other funding agencies? Is it the responsibility of others if they use ratings for purposes that may not be appropriate?

Is there scope for the funding councils to work with other funding agencies– particularly the research councils – to develop complementary assessment processes which minimise the total assessment burden? Could the funding councils and research councils make more use of data produced by their respective processes? If so, how?

b. How often should research be assessed?

How often should research assessment take place? Should all subjects and all institutions be assessed at the same time or with the same frequency? Should clusters of subjects be assessed separately?

c. What is excellence in research?

The purpose of research assessment is to provide information about the quality of research – but what is quality?

Another way of asking this question would be "what is it that distinguishes the best research"? Some might feel that this begs the question, "Is it helpful to speak of the 'best' research, in a way which implies that there is a magic ingredient that separates it from the rest"?

Are there different aspects of research activity (for example creativity and applicability) that each demand recognition? Did the 2001 RAE capture this?

d. Should research assessment determine the proportion of the available funding directed towards each subject?

In devising a system of research assessment, it is important to know whether it will be required to inform the distribution of funds between subjects as well as between institutions.

There are a number of ways in which ‘subject pots’ might be determined. These include:

- the quality of UK research in the subject, benchmarked against international competition
- the volume of research in the subject that meets a given quality threshold
- a strategic judgement on the importance of the area to the UK
- a metric based upon external funding in the subject
- an overtly historical distribution which aims to retain the current balance
- a mixture of the above.

If the relative quality of research in different subjects is to be used as the basis for generating subject pots, how is this to be assessed?

e. <u>Should each institution be assessed in the same way?</u>

The 2001 RAE obliged all institutions to submit to the same assessment. The research outputs of a large multi-faculty institution with a strong research tradition were assessed in the same way as those of a small college with no tradition of large-scale investment in research.

Some would argue that this is an unfair competition; others that it is important for those with minimal resources to see where they stand in relation to leading units. A middle position would be that it is sensible not to compare institutions that are very different but that the system should provide a ladder of improvement so that all researchers and institutions have the opportunity to demonstrate potential.

f. <u>Should each subject or group of cognate subjects be assessed in the same way?</u>

How far should the nature of the assessment be allowed to vary between subjects? Should each subject community be free to define the sort of assessment most appropriate to it? Should the funding councils go further in standardising assessment practice? Or is the current balance about right?

This is not necessarily a simple choice between a greater or lesser degree of standardisation. One approach might be to define a small number of broad subject areas, and to make assessment methods within each area as similar as possible while allowing the broad groups to diverge from one another.

g. <u>How much discretion should institutions have in putting together their submissions?</u>

At present, institutions have a large degree of control over the content of their submissions, over who or what is assessed and by whom. This ensures that planning decisions do not make it impossible for the particular nature of an institution’s research to be appropriately assessed, but it also brings significant disadvantages.

There are two alternatives: a more rigid system, or a system in which submissions are made and controlled by individuals, research groups or networks rather than by the institutions. The former risks the disadvantages of any inflexible bureaucratic procedure; the latter would arguably be unfair to institutions, as their funding would be determined by an assessment into which they had minimal direct input.

Both, however, would provide more objective results: ratings, scores or shares of the funding pot would depend entirely upon the quality of research activity as measured by the exercise, rather than reflecting the willingness of the institution to trade funding for the prestige of a high rating. They would also close the question of alleged unfairness to individuals who perceive that the decision not to include their work in RAE submissions has damaged their careers.

h. <u>How can a research assessment process be designed to support equality of treatment for all groups of staff in Higher Education?</u>

The funding councils are committed to ensuring that their research assessment process is non-discriminatory. They are also committed to ensure that it does not reinforce a culture, wherever such a culture may exist, in which staff are disadvantaged on the grounds of sex, sexual orientation, race and ethnic origin, disability, age, religion or any other irrelevant characteristic.

Are there features of past research assessment processes which discriminate or which can be abused by those seeking to discriminate against any group? Are there subtler effects, adversely affecting the legitimate interests of groups of staff, to which the design of the process contributes? What are the essential design features of a research assessment process that encourages genuine equality of opportunity for all.

i. <u>Priorities: what are the most important features of an assessment process?</u>

Most people would agree that a successor to the 2001 RAE ought to strive to be all of the following (and many other things besides):

- not burdensome
- rigorous
- fair to individuals and institutions
- informative
- transparent
- resistant to games-playing
- administratively efficient
- flexible (so that changes in policy can be accommodated without redesigning the entire process)
- minimally expensive.

We invite respondents to identify the three most important characteristics of an assessment process. These need not be taken from the list above but should reflect characteristics of the process rather than the philosophy underpinning it (we have asked elsewhere what constitutes excellence in research).

Group 6: Have we missed anything?

19. We invite respondents to tell us whether there are other issues or options not considered here. In particular, we would be interested to hear of any approach to research assessment that could not be described as a variant of the approaches listed above.