

**Preliminary analysis of the responses to  
the 'Invitation to Contribute'**

**Second meeting of the Review of  
research assessment**

**17 December 2002**

**Issue**

1. Preliminary analysis of the responses to the 'Invitation to Contribute'

**Recommendation**

2. The group is asked to note the paper. A separate presentation will be made on the implications for the work of the review.
3. The group is asked to consider whether funds from the review budget should be used to commission a more rigorous analysis than we have been able to attempt in the time available.

**Timing**

4. A decision on whether to commission additional analysis must be taken at this meeting if that analysis is to ready for inclusion in the final report.

**Further information**

5. From Tom Sastry (0117 931 7458)

## Background

### The Invitation to Contribute

1. The invitation was published on 27 September and closed on 29 November. Despite the short response period we have received 398 responses<sup>1</sup>.

### The analysis

2. Respondents were divided into four categories:

- Higher Education Institutions (we would have included FE institutions had we received any responses from them)
- Subject associations, departments, faculties and learned societies
- Individuals responding on their own behalf
- Stakeholders including sub-sectoral groupings such as the Russell Group and bodies outside the HE sector

3. A sample of each was read in detail.

	Responses	Approximate number read
HEIs	117	71
Subject associations	138	40
Individuals	89	30
Stakeholders	54	40

4. The analysis presented in this paper is based upon our reading of a sample of responses. Five analysts each read 20-40 responses and reconciled their conclusions. The group should note that this is not a scientific analysis of what is, after all, a very large qualitative dataset.

5. Except in cases where permission has been refused, we will be publishing all responses received from institutions, subject associations and external stakeholders on the review website. We do not currently plan to publish responses from individuals. This is because of the resource implications of checking such responses for potentially libellous content. We do plan to hold the responses so that they will be accessible to anyone with a professional interest in the dataset.

6. It is very much our hope that others will take up the challenge of using the dataset to investigate the views of the research communities and their stakeholders concerning research assessment. The group is asked to consider whether funds from the review budget should be used to commission a more rigorous analysis than we have been able to attempt in the time available.

### Recommendation

The group is asked to consider whether funds from the review budget should be used to commission a more rigorous analysis than we have been able to attempt in the time available.

---

<sup>1</sup> as of 11/12/02

## The results

### How should research be assessed?

7. The Invitation to Contribute offered four approaches to research assessment:

- expert review
- algorithm based entirely on metrics
- self-assessment
- historical ratings

8. It was emphasised that whilst these approaches are not mutually exclusive (for example, in previous RAEs expert panels have considered self assessment and metrics but have reserved to themselves the final decision on rating submissions).

#### Expert review

9. The consultation revealed overwhelming support for the continued use of expert review. Most respondents envisaged that assessors would wish to consider data metrics and/or self-assessment (or should be obliged to do so). Nevertheless it was the strongly held view of most respondents that assessment decisions should rest ultimately with a group of experts who should have a mandate to use subjective judgement to interpret the evidence placed before them.

10. There were two distinct schools of thought as to the kind of experts competent to assess research. Some argued strongly for orthodox peer review in which researchers are assessed by experts in the same field; others suggested that non-researchers, research users and even lay people ought to be given a much greater role in order to ensure that research aimed at non-academic audiences was properly recognised. Stakeholder groups in particular tended to be of the latter opinion.

#### Metrics

11. Whilst there was little or no support for the use of an algorithm to determine research quality, respondents devoted a good deal of attention to the use of metrics within an expert review system.

12. Some were opposed to the use of metrics per se. Others took particular exception to particular measures

“The use of impact factor and/or citation index, as in earlier RAE's, are a metric more of topicality, and possibly volume, than quality. Further they give a disproportionately high ranking to review journals, which while important, do not reflect quality or originality”

13. Others noted the high correlations (in some subjects) between RAE ratings and other measures.

14. The following broad generalisations can be made about the response:

- There was a general feeling that there was room to make the process more transparent. Respondents considered that panels could be more explicit about the metrics they would use and the weightings they would be given. However, many who argued for this also emphasised the importance of subjective judgement so it hard to get a sense of priorities.
- Those who commented upon individual or group assessment were almost unanimous that it was incompatible with the use of metrics. At the individual level, metrics could give rise to great unfairness and so long as they form an important part of the process it would remain vital to preserve the confidentiality of assessments of individuals<sup>2</sup>.
- All the available metrics came in for criticism<sup>3</sup>. Many respondents disputed the use of quantity measures such as PhD numbers. Even completions were held by many to be a quantity measure. Reputational assessment was almost universally rejected whilst citations were accused by many of having a conservative bias.
- Most respondents considered grant income to be an input measure- some even suggested that, if the funding councils are interested in value for money, less credit ought to be given to those who win large amounts of funding unless they manage to translate this into a greater quantity of high-quality research.
- Some respondents, however, considered metrics a necessary balance to the professional opinion of the panels. It would be fair to conclude that those who were the most sceptical about the judgement of panels were the greatest enthusiasts for the role of metrics.
- In general, metrics were not considered appropriate for the arts and humanities.

#### Self-assessment

15. There was some limited support for a system based around self-assessment. Many respondents discussed self-assessment as a viable option though few preferred it to expert review.

16. Many doubted that self-assessment would prove less burdensome than expert review 'if done properly'.

17. The strongest support for self-assessment came from a small minority who wished to see 'research council' type assessments- effectively the abolition of units of assessment to enable interdisciplinarity.

18. However, whilst respondents preferred expert review to self assessment there was some support for a greater element of self-assessment within the context of an expert review system, especially amongst the stakeholder group.

#### Historical ratings

---

<sup>2</sup> it would follow from this that it is safer, as well as more necessary, to employ metrics in larger units of assessment where errors at the individual level will have less of an impact upon gradings.

<sup>3</sup> we were perhaps naive to ask about metrics rather than other forms of "evidence" of research competence. The RAE already collects evidence of esteem (the RA6 form) but could conceivably admit a broader range of evidence. It would have been useful to know whether respondents who were unhappy about available metrics or who wished to assess other aspects of excellence had any views on this question.

19. There was little support for the use of historical data except as a means to establish the extent to which strategic objectives had been met over the assessment period (by comparing achievements with old strategy statements).

20. A minority of institutions however, did argue that as the 'balance of power' in research changes only slowly, it is hard to justify a 'blank sheet of paper' exercise every five years and that some use of historical data might be a pragmatic response to this reality.

" Changes in research strength are on the whole slow and an approach which took greater account than at present of historical research performance would provide a degree of stability and would make meaningful planning possible"

21. This view was not shared by subject associations or academic staff. In fact, some respondents were particularly vehement on the subject:

"The recipe for complacency...If this approach is adopted, you would need to send sackfuls of laurel leaves to the favoured researchers so that they could rest on them."

### **What is quality in research?**

22. A few respondents asserted that quality exists independently of fitness for purpose or utility. More often this was implicit in the arguments of those making the case for 'orthodox' peer review:

"Any effort to find so called objective criteria... should be strongly resisted, as such methods can lead to the overlooking of real quality in terms of creativity and originality. Hence the idea of employing so called experts in lieu of peers is a travesty of academic process, since such experts will inevitably use some 'objective' criteria which cannot access these distinctive qualities in academic writing."

23. For others the notion of quality was itself deeply problematic if divorced from the impacts of the research.

24. It would be wrong to say that there was a consensus around this question, and dangerous to attempt to present a 'middle position'.

25. However, many responses accepted both that assessors should be prepared to consider the quality of research in relation to its purpose and that whether research is 'blue skies' or directly applicable it should be subject to rigorous assessment.

"Excellence must be seen as multi-dimensional and closely allied to fitness for purpose. It must recognise the importance of rigor and appropriateness of method to the problem posed no matter when and by whom this is posed. Attributes such as value to beneficiary, applicability and creativity are other dimensions which should be taken into account in the assessment."

26. Amongst stakeholders, there was an almost unanimous recognition that there is a need for research assessment to recognise and reward the diverse characteristics of a healthy research environment that enables high quality research. There was acceptance that the previous RAE was capable of recognising intellectual excellence but most respondents wished to see the criteria used to assess excellence broadened to include pure research, applied research, practise based research, impact, utility/relevance, research training, research management, collaboration, multidisciplinary, and knowledge transfer in its broadest sense.

27. There were some suggestions that researchers (and institutions) should not be expected to demonstrate excellence in all categories, thus recognising the diversity of the research base. Some responses developed this argument further to suggest separate scores for different types of criteria rather than using one overall grade or ranking.

28. There was general agreement that the characteristics of excellence would vary across subject areas. It was acknowledged that this would need to be reflected in the assessment rules. Stakeholders were inclined to favour a common framework for assessment with 'local' variations rather than complete devolution to subject panels.

## **Interdisciplinarity**

### Assessment units

29. There was considerable support for reducing the number of units of assessment, but this was by no means universal.

30. Reducing the number of units of assessment was acknowledged to be an effective means of reducing the number of cases in which trans-disciplinary working may be discouraged by the structure of the assessment (although it is not a mechanism for dealing with interface problems when they occur).

31. However, many individuals and subject communities stressed the importance of being assessed by genuine peers- which led many to favour the retention of the existing units of assessment, or even the creation of new ones.

32. Some responses presented both sides of the argument

“The consultation document raises the issue of subjects being grouped for assessment purposes. There might be the benefit of greater support for interdisciplinarity, if...research were to be grouped with a cluster of cognate disciplines... However, the attendant risks are that one particular view of research comes to dominate quality judgements, with even less accommodation of different research paradigms and emphases the result.”

### Cross-referral processes

33. Many identified the perceived efficacy of cross referral processes as crucial to the confidence of researchers in large panels. If researchers are confident that their work will be considered by experts in their field, even if those people are not members of the assessment panel, they are more inclined to accept the case for larger assessment units.

## **Planning and strategy**

34. A strong theme in many responses, particularly from stakeholders, was the need of researcher groups to be able to demonstrate appropriate forward looking research planning, both with respect to research training and management.

## **Institutional discretion**

35. There was a general welcome for the suggestion that institutions ought to be obliged to submit all staff. Individual academics and subject associations were strongly in favour, whilst institutions were split on the issue, with a majority opposed to any change.

36. The most significant objection to the inclusion of all staff was that it would penalise departments with significant numbers of teaching staff and would increase the pressure on committed teachers to undertake research for which they may have little or no vocation. Many of those who were generally supportive of the proposal maintained that a mechanism would have to be found to ensure that excellent research groups in teaching-led institutions could continue to win recognition.

37. An alternative suggestion was that scholarship could be assessed alongside research. This, it was argued, would leave no reason not to submit academic staff.

## **Equal treatment**

38. Perhaps surprisingly, there appears to be rather more concern in the sector about the impacts of the RAE upon new researchers than about its treatment of women and minorities.

39. However, the Equality Challenge Unit spoke for many in its contention that allowing institutions to choose not to submit staff was a threat to equal treatment:

“We are not in favour of continuing the present scheme whereby HEIs can select who is entered in the RAE. In particular, we note that exclusion from the RAE can have a permanently damaging effect on someone's career, even though the cause of their exclusion varies greatly and may be misinterpreted.”

40. The unit was also not alone in suggesting that panels ought to be more explicit about the way in which the circumstances of new researchers would be taken into account in making judgements about their research.

“Another cause for exclusion (from submission as research active staff) to date has commonly been a career break in the assessment period or just before it, which may have a perceived or actual effect on research productivity. This tends to have greater impact on women than on men. It was theoretically addressed in the over-all regulations in the 2001 RAE, in that personal circumstances could be entered in a confidential section. But there was little confidence in the sector that consistent interpretations would be applied, and thus there were different behaviour patterns in

different UoAs and between different HEIs. An inclusive submission, in which the UoA as a whole was assessed, would be fairer and would encourage panels to take account of all contributions actually made.”

### **Frequency of assessment**

41. Respondents within the sector tended to favour a longer gap between assessments. This was generally considered a better solution than extending the assessment period because it would not entail some research outputs being eligible for more than one RAE.

42. Others suggested that, if four publications has (in effect, become a norm as well as a minimum, the limit could usefully be reduced:

“If the norm remains 4 pieces we believe there should be more time between reviews. We suggest that a period of 10 or 15 years represents a more genuine cycle of research than the previous cycles of 4 to 6 years. If the time between reviews remains 4-6 years the expectation - the norm - should be that two pieces of work should be submitted for each review.”

43. There were some voices in favour of rolling assessment but a majority were opposed.

### **Grading and scoring**

#### Greater discrimination in the grade scale

44. There was controversy both over the need for greater discrimination and over the means to achieve it.

45. Most respondents, whether supportive or not assumed that greater discrimination meant additional grades at the top end of the scale. Research intensive universities, concerned about ‘ceiling effects’ were understandably supportive of this proposal.

46. There was concern that merely adding additional points to the grade scale would place panels in an impossible position- obliging them to make fine judgements between leading departments which would be very difficult to justify.

“There should be no further refinement of the ranking hierarchy. The suggestion that there should be more discrimination in the rating system, particularly at the top end, is unacceptable as such a refinement would encourage sophisticated persuasion of the peer judges, and, in any case, it is quite difficult to differentiate between degrees of excellence at the top. The motive behind this suggestion is also suspect in that it seems to imply the much sharper targeting of research monies on grounds that are likely to be unsafe.”

47. Many respondents expressed concern over the comparability of ratings produced by different panels. Some considered that this was a problem which could be addressed through tighter controls on panels or a greater emphasis upon international benchmarking. Conversely,

many doubted that *absolute* ratings could be produced with sufficient reliability to be truly comparable, across subject areas. This led to calls for ranking, or normalisation of grades.

48. Ranking, would of course leave panels with the same difficulty as that described above: that of distinguishing between submission of a similar standard; it would, though, address the other prevalent concern about the grade scale- that institutions' behaviour is distorted by the need for institutions to gain or retain grades. Many respondents observed that if there were no grade thresholds, much of the stress and games-playing would disappear from the exercise.

49. Another solution which attracted a considerable degree of support might be described as 'profiling'. There were two quite different forms proposed. One, as noted above, recommended that assessors score different aspects of excellence separately; the other would simply produce a quality profile and take from the panels the responsibility of translating proportions of 'national' or 'international' research into grades.

### **Radical solutions**

50. A number of respondents suggested allowing researchers (or institutions) to archive their own work on a central database run by the funding councils. It would be possible to monitor the use made of such centrally archived work, providing a source of information which could inform any assessment.

### **Other points**

51. There was little support for the current system of "international/national/subnational" classifications. It was often noted that research in some fields is a far more globalised activity than in others and that to hint that geographical reach is a criterion of excellence may be at best confusing, and at worst unfair.

52. Most respondents accepted the need for an element of prospective assessment but considered that the emphasis of the exercise ought to be on past performance.

53. Many respondents argued that research assessment should not compromise an appropriate articulation between teaching and research either through research funding or assessment.

54. There were suggestions that the collection of data and evidence should be an ongoing process. SERA proposed that the "system should be developed as a single and administratively simple online database. It may involve periodic sectoral reviews through a panel of experts (with international membership) to trawl particular sets for related disciplines.