

## SOCIAL SCIENCES DIVISION, UNIVERSITY OF OXFORD

### HEFCE Review of the Research Assessment Exercise 2001: submission from the Social Sciences Division

#### Introduction

This paper summarises the results of consultation with the constituent departments and faculties of the Social Sciences Division, and of discussions held by the divisional board and its Research Committee on the methodology employed in the 2001 Research Assessment Exercise and the future of research assessment.

A thorough and detailed analysis of research assessment, on which the division's response to the consultation has drawn, has been submitted separately to the review by Professor David Hendry, Head of the Department of Economics. The division fully endorses his conclusions and recommendations. A copy of his paper is attached to this submission.

The consultation document proposes four possible approaches to assessment: expert review, algorithm, self-assessment, and historical ratings. The view of this division is that some form of 'expert review' system should be employed, but with substantial changes to the ways in which UoAs are assessed and funding allocated. Detailed comments on the problems with the current system, and suggestions for alternative approaches are set out below. These comments address a number of the crosscutting themes identified in the consultation document.

#### **An appropriate model for research assessment: expert review**

The strongly held view of the division is that a system of expert review should constitute the main method of research assessment, and should be based on the research contributions of individual scholars. A qualitative evaluation of the quality of research should be a central plank of any system. This might be supplemented with other methods, such as the use of algorithms. If metrics were to be used in this way, clear and unambiguous criteria would need to be established. Historical ratings could also be used, which would reflect the long gestation of fundamental research. Detailed methodologies might vary between disciplines, and it should be the responsibility of the expert panel for each UoA to define research excellence as appropriate for that discipline or group of disciplines. It cannot be assumed that excellence is one-dimensional or that the same sense of excellence is appropriate in the humanities as it is in the natural sciences, for example. A combination of methods to supplement the main method of expert review would be a way of capturing different dimensions of research excellence. It is vital that research should be assessed on the basis of social as well as economic value.

*Membership of the RAE Panels:* The selection of appropriate panel members is vital to the effectiveness of expert review, and it is the division's view that the way Panels are selected needs reform. In disciplines which are not consensual, it is not clear that the same gradings would have been obtained in the last exercise had different panels of expert advisers been enrolled. In any case, the selection procedure for Panel membership needs to be more transparent.

More generally, there is concern that the Panels may not have the time or breadth of expertise to ensure that fair ratings are given. The fact that panels' gradings are now submitted to some overseas scholars for confirmation is a welcome development. However, there is an argument for membership to consist primarily or solely of overseas scholars of recognised international standing, given the importance attached to the criterion of 'international excellence' in grading departments.

*Panel consideration of general submissions:* In the 2001 RAE it was not clear how much consideration Panels gave to the narrative parts of RA5s and RA6s, which is worrying, given that these elements of submissions are the only opportunities units have to fully explain the data presented elsewhere. RA6 might need some kind of external adjudication to give an independent view of evidence of esteem. Panels should consider all submissions from departments, and the relative weight given to these should be made known in good time.

#### **Assessing the research contribution of individuals**

That assessment should be based on the work of individual researchers is considered essential, as this is the only way to ensure all research effort is properly appraised. Assessment should be based entirely on outputs, and *not* inputs such as the ability to attract external research grants, which disadvantages some areas of research.

*Research outputs:* The number and type of outputs required for submission are viewed as problematic. There is a need for a balance between an adequate number of items to be evaluated, and an inadvertent emphasis on volume. Up to four items per researcher should not mean that Panels *expect* four. A smaller number should not be interpreted as indicating a lack of research productivity. Account needs to be taken of differences between disciplines and subject areas, and it is suggested that different subjects might set their own criteria, mediated by expert views from learned societies and other appropriate professional bodies. There is also an argument for giving greater weight to the best items of those submitted by a particular individual. Consideration should be given to the inclusion in assessment of other activities that enhance the quality of research output, such as editing journals and refereeing, although it is recognised that there are serious problems of measurement involved.

It is suggested that a broad graduated scale be used to define the quality of individual researchers' work, which would make finer distinctions of quality than the current system. This might, for example, take the form of: outstanding (5); excellent (4); international (3); high national (2); national (1); and sub-national (0). These scores would then be used as the basis of a department's rating (see below).

### **Rating of departments**

The current rating system is problematic. Some of the criteria used by Panels in their assessment are considered to contain ambiguities. For example, in the 2001 RAE it appeared that units were marked down on the basis of a small number of individuals with weak research records, despite more than 50% of submitted staff having work judged to be of international excellence. The current criteria are complex and do not have rules to cover cases which meet some but not all criteria. Thus, a requires all to be national and up to half international, but it is not clear how 'up to half' is interpreted, nor what happens to units where up to half the staff are 'international' but not everyone counts as 'national' (i.e. those which exceed the criteria for 5, but fall short of those for 5\*). The scope for tactical decisions as to who should, or should not, be submitted needs to be reduced, and there should be clearer criteria as to eligibility and requirements.

In the 2001 exercise, it appears that there was considerable grade inflation at the top end of the scale, which did not seem to be explicable by a substantial increase in performance. There is a general failure to distinguish the 'outstanding' from the 'good', and there is a need to establish a mechanism for drawing such distinctions, in order to make the RAE worthwhile.

A simpler scoring system is needed, with funding on the basis of a graduated scale, which would avoid significant funding thresholds, and reduce panel debates on boundary cases. Scores for individuals should be cumulated across each department. If the cumulated score was used as the basis for funding, however the total was reached, all staff could be submitted without cost. This would have the advantage of providing incentives to raise the quality of output of all active researchers in a department, rather than omit those staff where it was feared that they would have a negative effect on the overall rating. The scale could be weighted to offer a premium to the very best researchers if so desired.

### **Ensuring comparability across Units of Assessment**

It appears that there was inconsistency across different subject areas in the 2001 RAE such that some panels marked more stringently than others (Social Policy departments, for example, scored low in relation to other social science disciplines, and it seems unlikely that this reflects a 'real' difference in research ability). Concern has also been expressed that in Sociology, younger academic staff may not have got fair treatment in terms of rating their research as of international excellence. This may have been due to a mechanistic approach to the ascription of staff to categories, allowance not being made for young/new staff when judging publications listed in RA2. Grade criteria need to be better specified to avoid inconsistency of interpretation (for example, analysis by BERA and SPA showed that some panels took international to mean international recognition in fact, whereas others interpreted it as work that ought to have international recognition as a reference point in its field).

Greater interaction and comparability between panels is needed to ensure parity of quality assessment and to reduce incentives to move staff into units where competition is perceived to be less intense (e.g. there is evidence that some weaker Economics Departments have moved into Management and Business). It has been suggested by Economics that an alternative objective system of cross-subject comparison needs to be found, perhaps involving comparison of top-ranking UK departments with non-UK departments, or utilising a list of top international researchers in each discipline. If bibliometric measures could be standardised within disciplines, they could be used between disciplines.

There is a general problem for interdisciplinary research that cuts across UoAs. Development economists were cross-referenced to Development Studies from Economics, for example, but because submissions were not entered as a group and with development studies researchers from other disciplines (who were not cross-referenced from their disciplines), they were disadvantaged. A similar problem occurred in relation to Economic History.

### **Allocation of research funding**

As much attention needs to be given to the funding allocation mechanism as to the assessment process. Within UoAs, it is considered desirable that the same score should attract the same funding, however the total is reached (using a scoring system like the one outlined above), which would reward the best research. This would be a vast improvement on the current system which can lead to very divergent funding levels because an 'international' rating is based on the median quality researcher, and a slight mistake by a panel can create a major rating change.

Efficient allocation between UoAs is a difficult problem. Any formula should in principle be based on the social and economic value of research outputs, and not on the cost of research, as it cannot be assumed the highest cost activities are the most productive. The use of cost bands is highly problematic and needs reconsideration, as the system locks in historical anomalies in funding, since UoAs better funded in the past have since undertaken more expensive research. Greater interaction and comparability between panels is essential (see above). In the last exercise, because disciplines in the same cost band with different average ratings received the same basic funding unit, a high-rated department in a UoA with a high average rating could receive less funding than a lower-rated department in another UoA.

Finding more effective funding formulae which create incentives to sustain and improve research performance is vital. The current system, with its sharp thresholds rather than graduated scale encourages strategic behaviour. Allocation rules might be changed to create positive incentives. For example, introducing a factor for length of appointment at the submitting institution prior to a particular assessment exercise would discourage strategic recruitment prior to the RAE.

### **Other issues**

#### *Timetable for RAEs*

It is considered desirable that the period between successive exercises should be at least five years, and possibly longer for some UoAs to reflect the circumstances of particular subject areas (a rolling system would be too onerous for institutions and departments). This would take account of lengthy publication periods, and also reduce the current disadvantage to those engaged in long-gestation fundamental research. The problem of a more 'historical' allocation could be offset by the use of forward-looking indicators, such as confirmation that a particular item had been accepted for future publication.

#### *Preparations for the RAE*

The general view of this division is that the infrastructure and administrative aspects of preparing for the RAE in its current form place excessive burdens and costs on units and the University as a whole. Any new system must seek to keep its requirements within reasonable limits, and to utilise the submissions from departments and HEIs as effectively and transparently as possible.

JED/JED

25.11.02