Sir Alan Wilson FRS
Director General, Higher Education
Department for Education and Skills
Sanctuary Buildings
London SW1P 3BT

*tel* +44 020 7451 2507
*fax* +44 020 7451 2691

www.royalsoc.ac.uk

*From the President Lord Rees of Ludlow*
26 July 2006
Our ref: PRS/ASP/CCG

Dear Alan,

I enclose a copy of the Society's response to the *Next Steps* consultation on the reform of higher education research assessment and funding. I hope our submission is helpful, and look forward to continuing discussions on the subject. I have also sent a copy of the submission to John Kingman at HM Treasury.

The Society has repeatedly called for a review of research assessment post 2008,[2, 3, 4] and we agree with Government that a review is required. However, our overall conclusion is that the proposed income metric based models are not appropriate, and we feel there is merit in exploring ways forward that are not covered by the options presented in the consultation. There is scope to simplify and better target the collection of quantitative data to underpin the work of peer review panels, but we do not believe that it is sensible to dispense with the panels. The Society is keen to support the Government in reaching an effective solution.

The Society has always had an active interest in research assessment and resource allocation, submitting evidence to major consultations, and publishing proactive position pieces. Our key principles are included in work we undertook in 2002-04:

- Any assessment method must include an element of peer judgement.[1, 2, 4]
- Quantitative indicators can inform the judgement of these peer panels.[1, 2, 3, 4]
- Different subjects may have different relevant indicators.[1, 2, 3, 4]

We agree with Government that dual support is a valuable system that rewards excellence and nurtures promise. Part of the value stems from the way the dual support system allows greater plurality of decision-making in funding allocations and hence greater scope for creativity to flourish. The peer review judgements facing Research Assessment Exercise (RAE) panels are different from those facing a grant panel looking at the promise of a specific project proposal. If decisions about the quality related (QR) component of the Higher Education Funding Councils' block grant were made solely on the basis of decisions already made by Research Councils and

other funding agencies, the effective outcome of an income-based model, some of the value of dual support would be lost.

The RAE was introduced in 1986 in part to instil a culture of greater public accountability into the university system. In this it has certainly succeeded, and for some universities at least it has also been a stimulus for improved management information systems so that some of the data collection activity would continue even if the RAE were abolished. We agree with Government that research assessment in the UK now needs to be reconsidered. We endorse the stated aims of properly recognising and rewarding both interdisciplinary and user-focused research and reducing the administrative burden for panels and institutions.

We believe that changes in individual and institutional behaviour are a necessary response to any reward system. To discourage negative behaviour by institutions, any replacement system must minimise both the data requirements and dissuade institutions from expending disproportionate resources in over-preparing for the assessment. The assessment system must not encourage individuals to focus on a particular aspect of research at the expense of other valuable parts of academic endeavour.

The way in which research assessment results are used to inform resource allocation is of prime importance. The stark differences between the funding levels associated with different RAE ratings in previous rounds have led to institutions and individual departments investing significant time and effort to ensure that their submissions will be as impressive as possible. In particular, decisions about how departments and groups work with each other, within and outside the institution, may be affected. It is therefore important that the manner in which results are used to determine funding is designed in a way that is transparent, and encourages institutions to make decisions about research groupings purely on the basis of how to build and sustain excellent research.

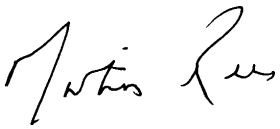
We remain pleased with the significant improvements made to the 2008 RAE, in particular the move toward the production of a research profile for each assessed area. This provides the potential for funding without great discontinuities between departments awarded adjacent ratings. This was a key recommendation to arise from the work we have undertaken[1,2].

The Society's second 2003 paper[3] highlights the close correlation between institutions' HEFCE research funding and their income from Research Council grants, as part of a discussion of problems associated with research assessment and resource allocation. The graph shown was updated in the *Next Steps* document.[5] The correlation is not unexpected, because both variables are related to institution size – indeed a plot of HEFCE research funding against numbers of research active staff shows a similar correlation. In addition, the pattern and volume of grants awarded is strongly subject-dependent. We therefore believe that grant income does not give a robust proxy measure of quality, which should be the key variant for a policy of distributing funds selectively.

We support proposals that the 2008 RAE goes ahead as planned, and that a shadow exercise using a modified approach runs in parallel. That gives the Government, working with universities, learned societies, customers of research, and the wider community enough time to get the modified approach right. It is important that any replacement actually delivers overall savings of time and effort, and we are keen to help in considering the way forward and evaluating the options.

We will publish this letter and the submission on our website in due course.

Yours sincerely,



Martin Rees

References


1 Royal Society 2002 Roberts review of the RAE: Submission from the Royal Society

2 Royal Society 2003a The Royal Society response to the Roberts review of the research assessment exercise.

3 Royal Society 2003b Supporting basic research in science and engineering: a call for a radical review of university research and funding in the UK

4 Royal Society 2004t the Royal Society's submission to the House of Commons Science and Technology Committee follow-up enquiry into the RAE

5 HMT, DTI, DfES and DoH, 2006 *Science & Innovation Investment Framework: The Next Steps*.

THE ROYAL SOCIETY

Policy Document 24/06

# Response to the DfES consultation on *the reform of higher education research assessment and funding*

The Royal Society is pleased to respond to the Department for Education and Skills' consultation on the *Reform of higher education research assessment and funding* (DfES 2006). This submission has been formed with the advice of a small group (membership at the end of the document), and approved by the Council of the Society. We make some broader points before responding to the specific consultation questions.

**Key Points**

- We are not persuaded that an income-based formula is a viable option for assessing research performance at the level of university departments, and remain strongly committed to the need for subject based review panels. These should be, as now, informed by a series of qualitative and quantitative indicators.

- We agree with Government (HM Government 2006) that dual support is an effective mechanism to sustain excellent research. The vital plurality of judgement, which is a central feature of dual support, is lost if either funding stream is directly dependent upon the other.

- We agree that the assessment process needs to be more efficient and streamlined for institutions and assessors, and that user-focused and interdisciplinary research should be better recognised and rewarded.

- We believe that any reward-linked assessment will influence individual and institutional behaviour: that is its aim. Behavioural responses to any system will need to be monitored to identify negative effects.

- There is now the opportunity to consider carefully how best to meet these objectives of increasing efficiency while avoiding negative consequences. The solution must involve an approach based on peer review and supported by a simpler, more aptly targeted set of indicators. This needs to be in place for a trial run alongside the 2008 RAE, so there is time to get it right. We are keen to assist in the development of the new assessment process.

**The current situation and the way forward**

Research in the UK receives public investment selectively, via the dual support mechanism which sustains high quality research and nurtures promising projects and individuals. Research Council (RC) funds are distributed on the basis of specific grant applications, judged on promise, while Higher Education Funding Council (HEFC) Quality Related (QR) funds are allocated on the basis of past achievements, as assessed by the Research Assessment Exercise (RAE). RC funds must be spent on the project for which they were awarded, whereas HEFC QR funds can be used at the discretion of the institution. The two streams are judged on different criteria, so to assume that the RC peer review process could be an appropriate substitute for the RAE panel peer review is a mistake.

The unit of assessment for HEFC QR funds is (approximately) the university department, and funds are then allocated at the level of the institution as part of the block grant for strategic decision-making by institutional leaders. Subject-based peer review panels use quantitative indicators to inform their judgements, although the degree to which different panels use them varies substantially. The RAE ratings are not only used to

inform HEFC QR allocations, but also as a useful strategic tool for university managers, and as benchmarks by the wider international community, for example by companies seeking research contracts.

Some of the past dissatisfaction about the RAE arises from the way in which the ratings have been used in funding formulas, in particular the large discontinuities in funding between adjacent ratings. There is also significant dissatisfaction over the amount of effort required for the full assessment process. For panel members, there is an overly-high volume of material to consider. Panels will need to ensure they collect a modest amount of information, and that all the information they collect is used in reaching their decisions. In the first incarnation of the RAE, departments submitted five publications in total, allowing panels to focus on assessing the top-quality outputs from each department. A return to peer-assessment of the peak of research quality, with reference to the remainder of research undertaken via quantitative indicators, may be a more efficient use of panel members' time and effort.

Institutions use significant resources while preparing their submissions, because the results are so significant. It is desirable that the exercise does not place a large burden on institutions, which can be done by minimising the amount of information that is collected. It is as important to structure the assessment to reduce the need for institutions to expend resources 'repackaging' their research and undertaking dry-runs. The most effective way to do this is to structure the assessment to ensure the information collected is objective and organized as simply as possible.

There has already been a significant amount of effort by panel members, institutions and individual researchers towards the 2008 RAE round. For reasons of stability and morale, it is important to allow the 2008 RAE to go ahead as planned. Bill Rammell MP (2006) and others have suggested that the results of the RAE round would inform funding allocations for not longer than a single academic year. This is a dangerous proposition, which may lead to the exercise not being taken seriously by panels, institutions and individual researchers. There is now the opportunity to consider carefully the aims of a new arrangement and to fine tune the methodology to ensure it is effective, efficient and streamlined, both for institutions and assessment panels. This needs to be in place for a trial run alongside the 2008 RAE, so there is time to get it right. We are keen to assist in the development of a new assessment process. Once the assessment methodology is determined, the way in which the results inform resource allocation needs to be considered carefully, and in a transparent manner.

**Responses to the consultation questions**

1 *Which, if any, of the RAE 2008 panels might adopt a greater or wholly metrics-based approach?*

The primary assessment mechanism for all subjects must be peer judgement. The extent to which indicators can be used varies enormously between subjects, and members of each subject community know the relevant indicators and the degree to which those indicators can be useful when making judgements about their own discipline. There may also be significant differences in the suitability and the degree to which indicators can be used at the sub-discipline level, as well as significant differences in the values of any quantitative indicators used. The use of a similar set of indicators across subjects may provide an opportunity to examine any differences in assessed quality between subjects.

2 *Have we identified all the important metrics? Bearing in mind the need to avoid increasing the overall burden of data collection on institutions, are there other indicators that we should consider?*

No single indicator is a sufficiently close proxy for research quality to do more than inform peer review panels. All relevant indicators need to be considered together and the results synthesised by a panel, rather than simply being plugged into a formula.

Research indicators include both input and output measures as well as measures of volume and esteem, but these indicators should do no more than inform peer review panels. No single measure should be used out of the context of other indicators, and panels should consider both qualitative and quantitative indicators. It is also important to note the difference between outputs (eg articles in peer-reviewed journals), their influence in the scientific community (citations), outcomes (specific changes caused by the research outputs) and impact (longer-term effects on, for example, health and prosperity).

Some research indicators could include:

- Research Council funding
- Other research funding
- Research active staff
- Research students
- Bibliometric data
- Patents
- Invited conference presentations
- Major prizes
- International recognition
- International exchanges
- Fellowship of the Royal Society and other Academies
- Linkages with other universities
- Linkages with industrial, charitable and other organisations
- Opinions of businesses (local, national and international)
- Engagement with policy-makers
- Academic community service (RC panel service, advisory boards, Academy/Learned Society councils and committees)

Assessment based solely on research income will not take account of any research done by staff without explicit financial support. This would have a particularly strong effect on mathematics and other theoretical disciplines (including mathematical physics, theoretical chemistry, bioinformatics etc), which are often undertaken by individuals and groups that have minimal costs above salary.

Studies of bibliometrics suggest that using relative citation rates (within disciplinary areas) should at least be considered as a possible quality measure. This would also capture research that is not supported through specific grants, but which has staff and infrastructure costs. Raw citation data are not robust when applied as a snapshot at the level of the individual, while trends in aggregate are informative.

Hirsch (2005) has proposed the index *h*, defined as the number of papers which have been cited $\geq h$ times, as a useful index to characterize the scientific output of a researcher. It is calculated by determining the highest number of papers that have attracted that same number of citations or more. For example, an individual with an *h*-index of 50 would have fifty papers which have each been cited fifty times or more. The *h*-index measures the broad impact of an individual's work and can be found easily by ordering papers by 'times cited' in the ISI Web of Knowledge database (http://portal.isiknowledge.com). It avoids the considerable disadvantages of using total citation numbers, total numbers of papers, citations per paper and number of significant papers which have previously been proposed (Hirsch 2005). It can be applied to individuals as well as groups, but care would be needed in the selection of appropriate time periods and consideration given to the potentially difficult issue of multi-author papers. Different authors may have contributed at very different levels to a particular paper and appropriate weightings may have to be applied. Appropriate time periods and index values are likely to vary significantly by subject.

3 *Which of the alternative models described in this chapter do you consider to be the most suitable for STEM subjects? Are there alternative models or refinements of these models that you would want to propose?*

The idea of replacing peer judgement of outputs with quantitative measures of input is flawed. Therefore, we do not find any of the models satisfactory. We also note that section four of the consultation document presents a range of possible indicators, however the models in section five do not include measures other than research income. Rather than creating an algorithm that takes a series of numbers and calculates another number, subject based review panels are needed. These would be informed by a series of indicators, some of which will be quantitative, others qualitative (see answer to previous question for examples).

Any indicators are likely to be inter-correlated, and it is therefore desirable that numerical indicators are chosen such that their errors are weakly correlated. This statistical independence will improve the predictive power of the indicator set. If all of the indicators do not point in the same direction, the situation needs further study. Developing such a method of 'convergent partial indicators' would be a useful way of targeting the efforts of a peer review panel to where real judgement is needed.

Critical evaluation of various assessment methodologies is required, for example as undertaken for the medical sciences by the UK Evaluation Forum in 2006. Methodologies should not be judged by how closely they replicate the 2008 RAE, but how robustly they capture the quality of the UK's university research. Rather than seeking to replicate a particular assessment outcome, a new system should be an improvement on the last. In addition, any metrics-based model would not be independent of the 2008 RAE if the 2008 round makes increased use of numerical indicators.

It is very important for the assessment methodology to be clear about the scale of the unit of assessment, something about which the consultation document is ambiguous. The correlation in funding between research council grants and HEFC QR is only seen at the institutional level, and as both variables are a function of size, the correlation is not indicative of causality.

4 *What, in your view, would be an appropriate and workable basis for assessing and funding research in non-STEM subjects?*

As the Society is the UK's national academy for science, we will not comment on assessment mechanisms for non-STEM subjects. However, we are unconvinced about the benefits of having two completely separate assessment mechanisms, as many subjects bridge the boundary between 'STEM' and 'non-STEM', (eg cognitive science, computational linguistics). Cross disciplinary work is a significant, important and increasing part of UK research effort, and any new system needs to support the development of new fields and not discriminate against those that bridge current subject areas. The consultation document states that the recognition and reward of interdisciplinary research is a key driver for the reform of the RAE, so it is particularly important to ensure that any new system does not have subject divisions in assessment method.

5 *What are the possible undesirable behavioural consequences of the different models and how might the effects be mitigated?*

Any and all funding and assessment models will affect behaviour, and rather than seeking to mitigate effects, the methodology should be designed such that appropriate and sustainable behaviours are encouraged. Indeed, the aim of any assessment and reward system is to positively modify behaviour. In particular, any binary categorisation of individual researchers, or steep differences in rewards between ratings will have strong effects on behaviour as individuals and institutions seek to maximise their ratings.

The effects on behaviour of any assessment methodology should be continuously monitored, both to ensure the positive, desired effects and to identify the negative, less-desirable effects. It is also important that this

monitoring process can distinguish between behaviours that are beneficial in the short term and those that are sustainable in the longer term.

If QR funds were to be based on research grants and other contracts, a number of negative behavioural incentives would be created, including:

- For institutions and individuals to redouble their efforts to receive funds from the sources that are included in the QR calculation, leading to increased costs for those funders in processing the extra applications and a lower success rate for all applicants (ie decrease efficiency).
- For institutions and individuals to further downgrade teaching and other non-fund-raising activities, including engagement with the non-specialist public, individuals' Continuing Professional Development (CPD), and research training of postgraduate students.
- For institutions to focus on subjects that have higher research costs, and therefore for appointments to be made preferentially in subjects that have high research costs.

6 *In principle, do you believe that a metrics-based approach for assessment or funding can be used across all institutions?*

There is no single metrics-based formula that can be applied across all institutions or subjects; however, subject panels informed by indicators will be effective across subjects and institutions. Differences between disciplines are more significant than differences between institutions. The extent to which quantitative indicators can be used varies enormously between disciplines, and therefore the discipline-mix of an institution will be the determining factor.

7 *Should the funding bodies receive and consider institutions' research plans as part of the assessment process?*

We understand that institutions' research plans will be considered by subject panels for the 2008 RAE. These plans may remain a useful input into panel considerations, as a forward-looking complement to an otherwise backward-looking assessment. If previous plans have been effective then their results will be evident in the outputs/outcomes already under consideration. Equally, in future assessments, the effects of current plans will be evident.

8 *How important do you feel it is for there to continue to be an independent assessment of UK higher education research quality for benchmarking purposes? Are there other ways in which this could be accomplished?*

It is important that the UK Government and wider society have information about how effectively public money is invested. However, a benchmarking exercise has different aims from research assessment for resource allocation. As the aims are different, different methodologies are appropriate. The scale of the entity to be benchmarked is also important.

To make international comparisons between national research bases, trends in metrics are useful. At such a high level of aggregation, and as long as the direction of trends is taken into account, input cash metrics, volume metrics and output metrics (bibliometrics, patents etc) are all helpful. Like must always be compared with like, so when isolating and comparing components of national research systems, the function and context of each must be considered. Industrial experience of benchmarking exercises suggests that metrics work best when they are few in number, simple, sufficiently accurate to be credible, agreed with stakeholders, easy to develop (using existing data), robust and durable.

A system to benchmark individual institutions needs to have a different form. Given the increasing diversity in university missions, individual institutions should be encouraged to benchmark themselves against their best peers worldwide in the areas they prioritise – be this research, teaching, knowledge transfer or anything else. This information will be most of interest to the institutions themselves, who will be able to exchange best practice as well as developing their collaborative links with the institutions against which they assess themselves. The top universities in Korea have undertaken an exercise of this kind (for information, see Moon and Kim 2001), and are just beginning a new round.

The results of any institutional benchmarking or research assessment exercise will not only be used by Government and the institutions themselves, but also by potential international collaborators, potential industrial partners and all others who may seek to choose between research institutions. Any new system of assessment or benchmarking needs to be designed such that these purposes are taken into consideration.

**Group membership**

| | |
|---|---|
| Professor Ole Petersen FRS (chr) | Professor of Physiology, University of Liverpool |
| Professor John Allen | Professor of Biochemistry, Queen Mary, University of London |
| Professor Eleanor Dodson FRS | Professor of Chemistry, University of York |
| Professor Dianne Edwards CBE FRS | Professor of Palaeobotany, and Head of School of Earth, Ocean and Planetary Sciences, University of Cardiff |
| Sir John Enderby CBE FRS | Emeritus Professor of Physics, University of Bristol |
| Sir John Kingman FRS | Professor of Mathematics, and Director of the Isaac Newton Institute for Mathematical Sciences, University of Cambridge |
| Professor Georgina Mace OBE FRS | Director of Research, Zoological Society of London |
| Professor John Shepherd FRS | Professor of Marine Science, University of Southampton |
| Professor Christopher Snowden FREng FRS | Chief Executive, University of Surrey |
| Professor Philip Wadler | Professor of Informatics, University of Edinburgh |

**References**

DfES (2006) *Reform of higher education research assessment and funding.* DfES: London.
Available online at www.dfes.gov.uk/consultations/conDetails.cfm?consultationId=1404

Hirsch J E (2005) *An index to quantify an individual's scientific research output.* PNAS **102,**16569

HM Government (2006) *Science and innovation investment framework 2004-2014: next steps*. HMSO: London
Available online at www.ost.gov.uk/policy/science_consult.htm

Moon M and Kim K-S (2001) *A case of Korean Higher Education Reform: The Brain Korea 21 Project*. Asia Pacific Education Review **2,**96

Rammell, B (2006) in a presentation to the Higher Education Policy Institute Conference on research assessment.
Available online at www.hepi.ac.uk/downloads/BillRammell.doc

UK Evaluation Forum (2006) *Medical research: assessing the benefits to society*. Academy of Medical Sciences: London
Available online at www.mrc.ac.uk/index/public-interest/public-topical_issues/public-socio_eco_benefits.htm.

**Any enquiries about this submission should be addressed to**

Alice Sharp Pierson
Science Policy Section
The Royal Society
6-9 Carlton House Terrace
London
SW1Y 5AG

alice.sharppierson@royalsoc.ac.uk

+44 (0)20 7451 2589