

University of Birmingham

Response to the Joint Funding Bodies' Review of Research Assessment

Introduction

The University of Birmingham's response to the Joint Funding Bodies' Review of Research Assessment is given below, with an executive summary provided at the beginning.

Context of the Consultation

We would like to express at the outset our view that the nature of any future research assessment exercise is inextricably bound up with other ongoing discussions with respect to the funding of Universities, and in particular research, following the Transparency Review and other studies, the CSR, and pending the Government White Paper due in January¹. We are aware that there is pressure to change the RAE system, which is seen to have contributed to the current position of under-funded research, over-trading and undue strain on the dual support system². In addition, if the "core plus" funding model is to be applied, consideration presumably needs to be given to the development of a review system which takes an overview of all elements of institutional activity, especially the balance between them³. There is also the growing importance of the regional agenda, and the possibility of greater collaboration between institutions, perhaps to address regional needs or with a view to making better use of scarce resources. Taken together, these suggest perhaps that a more radical approach to assessing performance is needed than is proposed in the consultation paper. In other words, we wonder whether this consultation should have commenced by discussing aims and principles rather than the details of methodology.

The consultation paper, however, requires responses in relation to a number of different scenarios, and we have responded to these, while bearing the wider and perhaps more important issues in mind. The University welcomes the Review of Research Assessment and the opportunity this provides to re-consider methods of assessing research quality in the UK. We would stress very strongly however that we consider that whatever future method of research assessment is selected, the concerns expressed through this consultation by us and by others in the UK HEI community about the various shortcomings of the possible schemes should be taken seriously and fully addressed before the introduction of any new system.

Format of the Birmingham Response

In order to aid analysis, the response has been structured around the consultation paper, with the following internal headings:

- Executive Summary
- Concerns over RAE

¹ The Cross-cutting review of Science and Research: Final Report makes a number of useful points - see pages 50 – 57.

² A good example – if one is needed - of the effect of these issues and their inter-relatedness is provided in the Medical area. Following the RAE 2001 outcome with its increase in grades and inadequate funding, UoAs 1 and 3 suffered the greatest percentage fall difference between 1996 and 2001 than any other UoA (at about 20%). Simply put this meant that all Medical UoA's had to increase by at least one grade simply to attract a similar resource. Thus QR research income to most Medical Schools has barely changed despite increased quality at a time when research expansion is desperately required. Alongside this, the percentage of research supported by council funding has fallen between 1996 and 2001 (we are all aware of the issues re MRC income). This has meant that Medical research is increasing dependent upon Charity funding (CRUK, Wellcome etc) which is administered without overhead add on costs. This has further destabilized the funding of medical research. Alongside this, staff are also having to cope with the pressures of increasing medical student numbers to meet NHS doctor recruitment targets.

³ This should not be taken to mean that we endorse the suggestion in the consultation paper that research and teaching should be jointly assessed. Careful and detailed consideration to what such a review system might be is needed.

- **Key assumptions**
- **Three Contextual Factors**
- **Possible Models**
 - *Expert Review*
 - *Algorithm*
 - *Self-Assessment*
 - *Historical Review*
- **Cross-cutting Themes**
- **Any Other Points**

Extracts from the consultation paper are italicised to place our response in context; our comments are provided in bold.

Executive Summary

- Four possible processes of review are postulated. It is our view that any one of these used on its own would be unacceptable. As an institution we favour expert review, perhaps with limited, clearly defined algorithmic elements, and possibly with some element of self-assessment/historical review which might enable a “lighter touch” process to be implemented.
- A number of modifications are proposed to the existing system, addressing concerns with respect to the transparency of the process, the review cycle, the nature of the evidence base, equal treatment for all staff, the appropriate configuration of panels etc
- The need to review Third Task/Knowledge Transfer activity is accepted, noting the importance of first defining this activity in terms which are acceptable across the sector. It is suggested that submissions are assessed for research and KT quality according to separate criteria, and scored separately, although the same evidence base should be used.
- We do not consider that any research assessment process should determine the distribution of funds between subject areas, or whether additional funding is needed for new/emerging subject areas, although it might be appropriate for panels to flag concerns with respect to the latter scenario for consideration elsewhere.
- We would like to see the ratings scale rationalised but not expanded.
- We consider that the following characteristics are essential for the process: rigor, fairness and transparency. The following are desirable: administrative efficiency, operational ease.

A. Concerns over RAE

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 multidisciplinary*
- *lack of discrimination in the current rating system, especially at the top end.*

We agree with the majority of the above points, with the following provisos:

Game-Playing (bullet two) We are concerned at the implication that rules are being changed not because there is anything intrinsically wrong with them, but because the sector has developed a good understanding of them and makes use of that understanding to maximise the outcome positively. It is difficult to envisage any system where this would not, ultimately, be the case. One could just as easily classify this process as “responding to market forces” as “game-playing”. This statement also implies that the research assessment process is a game rather than a serious attempt to seek a true measure of quality. If it is the latter, then the most appropriate method should be used and safeguards built into it, rather than that a different method should be introduced which might be less susceptible to “game-playing” but which does not carry out the required task.

Interdisciplinarity (bullet seven)

We would like to remind the Review Group that that this concern differs in nature for different subject areas. Within Arts and Humanities, it is necessary to encourage and give full credit, in the assessment process, to work that crosses boundaries and synthesises different disciplinary methods and content. It is also necessary to take account of criteria developed and pursued within organically evolving but recognisable and institutional disciplines. The standards of coherent and organic disciplines must take a central role, however flexibly and diversely, in assessing outputs in the arts and humanities.

B. Key Assumptions

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.*

Dual Support

We welcome this re-statement of the continuation of the dual support system, which we regard as being particularly essential for the Arts, where QR is and should probably remain a high percentage of total research support, given the nature of research in this area.

Global standards

We also welcome the re-statement of the importance of considering research in a global context. It was generally felt that the introduction of international advisers was a positive step in this direction in the last RAE and some such scheme should continue in future, although no doubt with refinements built on the experience gained in 2001.

C. Three Contextual Factors

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.*

The People Dimension

There is some measurement of the output of trained researchers in the current assessment process which measures the numbers of postgraduate students involved. This would be a more effective measure if it measured completion rather than volume.

The Third Task

Perhaps the first and most important point to make is that in order to measure Third Task/Knowledge transfer activities, there must first be sector wide agreement of definitions of appropriate activity: we need to be clear *what* we are trying to measure in the way of outcomes.

We entirely agree that Third Task activities (not just Knowledge Transfer) do need to be both effectively measured and appropriately reviewed, especially if Third Task Funding is to be rolled into the block grant, and we acknowledge that the boundary between research and certain Third Task activities, including KT, is difficult to draw. However, we would point out the anomaly of (potentially) funding KT activities from HEROBaC/HEIF funds, then judging them as part of a Research Assessment exercise and in effect then funding them through the QR grant. This anomaly would obviously be removed if the core plus funding model were implemented and KT was funded as an activity within that.

We would suggest that an appropriate way to approach the assessment of research and knowledge transfer would be for assessors to look at the totality of activity in any area - research council-sponsored, industry-sponsored, etc. - from different directions and rate the quality of activity related to each direction (research, knowledge transfer) separately on would then be a two parameter scale. This would avoid pigeon-holing any particular activity such as TCS as either research or knowledge transfer, and recognize that in many cases an individual will be undertaking both research and KT within the same project or piece of work and at the same time through academic scholarship. We consider a two parameter scale would be beneficial to the sector in allowing much finer identification of specific strengths in different kinds of institutions. This could be accompanied by a textual commentary of some kind making clear what in particular the assessors were commending.

We would also point to the need to be sensitive to disciplinary differences with respect to both this point and point (a) above. Within Arts and Humanities, undue weighting of these two factors could build into the exercise a presumption in favour of established stars or those who can afford to run training programmes etc out of existing funding. This risks militating against the development of the new: new ideas often come from the edges and it is important to remain alert to innovation and to the need to sustain the potential for research throughout the system.

Funds for engaging with the community – an activity that may also draw on or stimulate research activities – are also now available and perhaps also should be considered.

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.

New Subjects

Whilst we recognise the need to support the development of new subjects and new fields of study – where they are of sufficient volume and significance, and where they are in areas where there are significant difficulties in gaining funding from existing sources,⁴ we would argue that the role of any assessment process should be to identify these areas and no more. Any decisions about action then needed should be placed in the hands of some other review body, which will need to validate this judgement and make recommendations (to the funding councils and research councils?) about action needed. It is important to avoid confusing the process of assessing research quality with judgements about funding levels necessary to promote a subject area, to avoid self-interest being exercised by those conducting the assessment. This assumes a continued input from the subject community to the assessment process through peer review or self-assessment. It is difficult to see how a purely mechanistic approach could result in this kind of judgement and an approach based on historical evidence would arguably militate against the development of new areas.

D. Possible Models

Four possible processes of review are postulated. It is our view that any one of these used on its own would be unacceptable. As an institution we favour expert review as the dominant process, perhaps with limited clearly defined algorithmic elements, and possibly with some element of self-assessment/historical review which might enable a more “light touch” process to be implemented. Our reasons for this perspective are outlined below.

⁴ For example, a new field of medical research is likely to be funded by existing medical research funders whilst nursing research is clearly not easily able to attract funds from medical or social sciences funders.

Expert Review

1. 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 groupings⁵, over the previous cycle, and/or their likely performance in the future.

2. 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.

Expert Review

Research is a professional activity, and must primarily be assessed by competent professionals in the appropriate fields. The international academic system of research is built around peer review. Attempts have been made by various groups (including the Royal Society) to identify alternative procedures – none of which, to our knowledge, have been successful. Various attempts to examine inter-assessor reproducibility in assessment of the quality of papers and other research outputs have highlighted that the more “expert” the assessor is, the more likely they are to identify both the significant factors of quality in an output and any deficiencies therein.

The selection of peers should continue to be firmly based in the appropriate research communities, with members of user groups being involved only in those subject areas where there is a demonstrable need, and with a defined role that does not involve them judging academic quality, but rather user relevance/dissemination etc, which could be given more weight in the assessment process. Non-UK advisors should continue to be used (although consideration could be given as to whether their involvement in the 2001 RAE was optimal).

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

Teaching and Research Assessment Combined

This would not be acceptable, given the divergence in methodologies likely to be needed to assess these two areas. A recognition, however, that the demands of each impinge on the other, and that research properly informs teaching, is appropriate. In particular the impact of teaching and teaching administration on the number of staff not returned in the assessment exercise could be allowed for more realistically. Any moves in the direction of combined teaching and research review would surely need to result in the development of some means of reviewing institutions by mission.

4. 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?

Expert Review

We would suggest that if a system based on expert review is considered then it should be not too dissimilar to the current process which has had many successes as well as some demerits. A number of important modifications/improvements are however suggested.

- An element of self-assessment could be built into the process to reduce the burden on institutions. There could be the option for institutions that judge the quality of research in a particular area to be unchanged from previous assessments or for 5 or 5* Units to opt for a “lighter touch” assessment, which might take the form of a digest of some of the data outlined below and a very brief commentary. In the case of 5 or 5*s, this could be seen as comparable to the Ofsted approach to assessing high-performing schools. Panels would judge whether a fuller assessment is required or would be beneficial to the institution. Care would have to be taken to ensure this did not result in complaisance.
- A rolling cycle of reviews could be undertaken but care would be needed at the institutional level that the tightness of staff accounting is not lost between successive levels of the review. (See Cross-cutting Themes).
- To address concerns within the sector with respect to transparency and fairness, the rules and procedures for the process and panel criteria should be published well in advance of any assessment round (ie years not months!) and should be as clear and comprehensive as

⁵ A grouping might be (for example) a research group, network, department, faculty, institution or consortium.

possible. They should enable the sector to understand clearly on what basis judgements are being made.

- The introduction of feedback to units was a welcome development in 2001, but if any such system is to operate in future, the quality and consistency of what is provided has to be improved. There was a huge variation in what was received at Birmingham, ranging from a single brief paragraph from one panel to a couple of pages of detailed and helpful comments from another.
- Discussions should take place within a number of discipline areas to identify the appropriate configuration of panels (see (d) below). This might lead to the creation of sub-panels or the formation of a small number of new panels, although broadly we feel the current number of panels is about right. If moves were made to reduce this number, it is our view that in order to assess activity fairly, it would still be necessary to draw on expertise from the whole range of disciplines represented by the current set of panels. Details of all involved in the review process should be made known well in advance.
- If elements such as KT are to be included in the review process, one option rather than having user representatives on panels would be to convene a separate experts group to assess these aspects of a unit's return, where a unit indicates it so wishes. This would obviously be more viable with a rolling programme of reviews as such a panel would probably otherwise have an unmanageably high workload.
- As suggested above, a two parameter grading system would enable research and knowledge transfer strengths to be identified. Detailed descriptors could also be used, indicating particular areas of strength. KT is one area of assessment that might be carried out to a large extent algorithmically if an agreement of the measurable outcomes could be achieved.
- We believe attention needs to be given to auditing the returns more carefully. We are not convinced that the current level of audit is adequate in comparison to the sums of money the system is used to distribute.
- Other more detailed suggestions are included below.

5. 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?

Prospective or retrospective assessment

Assessment should consist of a combination of prospective and retrospective, and we feel that previous assessment rounds have got this balance about right. Too much weight should not be based on prospective assessment given the difficulty of predicting the future with accuracy! We feel panels could take more account of targets set in previous years and whether they have been achieved.

b. What objective data should assessors consider?

Data to be assessed

We would suggest the following sets of data, based around previous requirements but with a number of important modifications:

Research Outputs: Consideration should be given to the number of outputs to be cited, where we would strongly suggest that the submission of fewer outputs for young/new⁶, staff part-time staff or those who have had a career break should be acceptable, within clear boundaries. Judgements should be made about quality, not quantity, in these cases. In addition, the provision of contextual information to explain the research content of certain kinds of output (particularly with respect to performance based research, but also potentially with regard to applied research) should also be required. There is support in some areas (eg Medicine) where there are relatively high rates of publication for the submission of more than four outputs to be the norm, or for there to be some consideration of quantity as well as quality. Clarification should also be provided over whether in the case of multiple-authored articles emphasis is being placed on senior authorship, which may be appropriate in some disciplines.

Expenditure. Care needs to be taken to understand the relative importance of external funding in different subject areas: a uniform approach should not be encouraged as this would clearly be detrimental to some areas. (A close reading of the 2001 Panel Criteria will demonstrate the different

⁶ Including staff making the change from practitioner to academic eg in Education or Engineering.

approaches taken). We would note however the self-reinforcing nature of the current system: high RAE grades encourage other funders to support particular areas; a high volume of peer reviewed funding then encourages panels to award high grades. More information on the way in which funding leads to research outputs and research context could be provided, to give assessors a better insight into the conditions of production and consequently a base from which to evaluate whether improved funding might support existing or developing excellence. In other words, assessors should be able to judge whether and in what way the levels of funding received contribute to the quality of outputs generated⁷.

Research Student data: Completion data is generally regarded as a better guide of the quality of a research culture than volume, and the way in which this data has previously been reported requires review to improve its usefulness. Data about competitive studentships only should be collected. In addition, first destination data⁸ could be useful as an indication both of an institution's contribution to the body of new researchers and/or as a KT indicator.

Knowledge Transfer Data: If Knowledge Transfer and associated activities are to be included in the process, measures of a range of additional areas will also need to be developed. We would draw the Working Groups attention to the relevant national discussions emanating from the recent Business Interaction Surveys undertaken for HEFCE and the DTI and contributions to the debate made by the Engineering Council amongst others.

Consideration of data

It might be possible formally to develop an algorithmic approach to some of this data for some subject areas, such as Engineering, where traditionally panels have in any case done this. However, we would strongly emphasise that this would be highly inappropriate in some subject areas (such as the Arts and Humanities) and that very careful consideration be given to the possible outcomes of such a process (see the section on Algorithmic Review below). Essentially, we believe part of the value of expert review lies in the contextual knowledge and understanding experts are able to bring to understanding the data submitted.

c. At what level should assessments be made – individuals, groups, departments, research institutes, or higher education institutions?

Level of Assessment

If a judgement is to be made about the quality of the environment within which research is undertaken, as well as the quality of the research outputs, then published grading has to be at the level of grouping of staff rather than at the level of individuals. This obviously requires submissions to be made at this level as well. Given that all HEIs organise themselves differently, and given the development of cross-boundary research, we would suggest that HEIs be given the freedom to group staff into units as they see fit but that part of the requirement of any return should be to demonstrate the coherence of such groupings. In practice, most institutions would continue to return at the level of disciplinary grouping (possibly School/department).

d. Is there an alternative to organising the assessment around subjects or thematic areas? If this is unavoidable, roughly how many should there be?

Assessment Areas

We do not see any alternative to organising assessment around subject or thematic areas, but would offer the following comments:

- The number/scope of panels must reflect how research in some areas currently focuses in thematic areas rather than traditional disciplinary groupings (eg Medical Sciences) and must also recognise the breadth of certain subject areas, such as Geography, (which is everything

⁷ The University of Birmingham would be willing to undertake some sponsored research in this area, building on its experience of the Research Exploitation Audit Process (REAP) sponsored by EPSRC in 1997.

⁸ First Destination Data is currently collected for HESA purposes for FT H/EU students only; with effect from next year data for PT students will be collected as well. However, data for International students is not required by HESA, so presumably is not routinely collected by the sector.

from 'history' and 'cultural studies' to 'hydrology' and 'environmental sciences'), Art and Design etc.

- Consideration should be given to the number of assessors appointed to panels (more may well be needed) and the formation of appropriate sub-panels to cover particular specialisms. The aim should be to reduce the need to consult outside the panel itself whether through cross-referral or through specialist advisors, and where either of these is needed, far more transparent processes should be adopted. In the case of interdisciplinary work, if cross-referral is essential the aim should be to ensure that appropriate assessors are fully brought into the discussions of the "home" panel so as to ensure that judgements are made in accordance with the culture of that particular discipline and on the basis of the rules of the home panel. (Whilst it was intended that this would happen in 2001, we understand that in practice the cross-referral system was unable to cope with the volume of cross-referrals generated and few panels had sufficient time to engage with each other at the level intended).
- If the current panel system is to continue, we would like consideration to be given to the establishment of a Public Policy Panel, and would be happy to provide further evidence in justification of this.
- We have particular concerns about the way applied research seems to have been considered by social sciences panels. We understand that certain panels regarded Government-sponsored work *de facto* as not being cutting edge. We would stress the need for assessment to focus on the quality of outputs, and not judge these solely by the inputs which gave rise to them. We believe this is an area where there is a need for much clearer thinking by panels in future. One option might be to develop a specific set of criteria to define excellence in applied activity.

Although there may be a need to create a number of sub-panels better to cope with particular complex disciplines, or possibly a small number of additional panels, broadly we believe the number of panels to be about right, with the addition of a panel to cover Public Policy. If moves were made to reduce this number, it is our view that in order to assess activity fairly, it would still be necessary to draw on expertise from the whole range of disciplines represented by the current set of panels.

e. *What are the major strengths and weaknesses of this approach?*

Strengths:

- The whole current system of research activity is based around peer review: it is well understood, and, with appropriate safeguards, credible.
- This is probably the only way in which a genuine judgement of research quality – as opposed to an assessment of proxies of quality (through algorithms) – can be made.
- Proposals suggested above would help address concerns about interdisciplinary/thematic research, and about the assessment of new/young staff, PT staff and staff returning to work after a career break.
- A team of peer reviewers, who should ideally be at the forefront of their subject, should have a collective and consistent approach to issues of quality and standards when assessment takes place.

Weaknesses:

- The range of criticisms which have been levelled against this process, many of which we have attempted to address above or under the cross-cutting themes section below.
- There seems to be a concern about the subjective nature of this process. In our view it is entirely appropriate that the assessment of quality should include both objective and subjective elements.

Algorithm

We do not support the use of a solely algorithmic approach to research assessment. Comments below outline our concerns.

6. *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*

Reputation

We believe it would be difficult to develop an objective quantitative measure of reputation which is after all a subjective concept. We would contend that this approach would be biased against younger staff, and would reflect a time lag in terms of appreciation of an individual's work. We would suggest therefore that reputation is not contained in any algorithm developed.

- *external research income*

Research Income

We believe that this measure would be most subject to distortion as a result of an algorithmic review of research. Those who had, would gain more; those who had little would lose that. We agree there is a current close correlation between external income and QR, but this is rather a chicken and egg situation, and is arguably the result of historical differences in the resources available to institutions. At individual subject level this correlation breaks down, especially within the Arts and Social Sciences. Here, because funds fall far short of demand, external income is an extremely incomplete and inequitable measure of performance. We would also note that certain forms of external income (eg charitable) are not necessarily open to all candidates and could not be considered in such a scheme.

- *bibliometric measures (publications or citations)*

Bibliometry

Measuring the volume of outputs has been tried in the past (in RAE 1992) with well-known results. We would draw attention to the fact that citation indices exist only for certain kinds of output and therefore discriminate against the many areas where the standard outputs are not academic journals, do not work for specialist areas and new/emerging areas where there are only a small number of researchers working, and do not necessarily reflect the quality of a piece of research – inaccurate research can be widely cited as its conclusions are rectified. In addition, there is the problem of US scientists not regularly accessing European Journals and hence not citing European work in some areas, which is a cultural rather than a quality issue. In many cases, it would require specialist input to make sense of any weightings applied. We would also note that within complex subject areas very complex weightings would be needed to take into account the different activities being undertaken (eg clinical research v basic science).

- *research student numbers (or completions)*

Research Students

Research student numbers certainly give an indication of levels of activity – particularly in Science and Engineering subjects – but are not necessarily an indicator of quality. It would be preferable in any case to use completion rates, which could be seen as a reflection of the quality of research environment.

- *measures of financial sustainability.*

Financial Sustainability

We were interested to see a proposal to measure the financial sustainability of research in an environment in which it has been demonstrated through the Transparency Review that research does not pay. Financial sustainability in research at the national level is dependent upon additional resources being made available to the sector. A proposal such as this would certainly not foster the development of new/cutting edge research.

7. *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?*

Metrics

We are of the view that an approach that made use solely of metrics in assessing quality would be unacceptable. Building a limited algorithmic element into the peer review process, however, with a greater or lesser dependance upon this approach according to subject area, might be acceptable, not least if it can reduce the burden on assessors. If Knowledge Transfer is to be incorporated in to the process, it might be susceptible to algorithmic analysis, but more work needs to be done in this area before conclusions can be drawn.

8. *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?*

No.

b. *What metrics are available?*

N/A.

c. *Can the available metrics be combined to provide an accurate picture of the location of research strength?*

No.

d. *If funding were tied to the available metrics, what effects would this have upon behaviour? Would the metrics themselves continue to be reliable?*

No. This would simply reinforce existing strengths and give new/emerging areas no opportunity for development. We contend that we would rapidly see those who have much being given more, and those who have little losing even that.

e. *What are the major strengths and weaknesses of this approach?*

Strengths:

- Not subject to individual bias

Weaknesses:

- Such an approach would be entirely retrospective
- The algorithms would potentially become so complex to account for all balances that they would become meaningless, and certainly would not be transparent.
- The process could potentially be so complex that it would not be any less burdensome.
- There would be endless arguments about the nature of the algorithm.
- Such a process would not be entirely objective because any decisions about the nature of the algorithm (what goes in it and what weightings are applied) would be subjective and highly political.
- The publication of the algorithms would enable “gameplaying” to take place.

Self-Assessment

9. *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.*

We note that submissions are already challenged to provide a measure of self-assessment through the current RAE system in RA5 and RA6 where reflection on achievements against plans is encouraged.

10. *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?*

Self Assessment

As suggested above, we propose a modification of the existing system to reduce the burden on institutions which would involve a lighter touch with respect to those areas institutions consider to be unchanged since the previous assessment, or for units previously graded 5 or 5*.

We do not regard self-assessment per se to be an acceptable proposal.

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

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

Weaknesses:

- **We find it difficult to see how a self-assessment process could be developed which would be credible outside the community, and, crucially, to Government. If one of the key functions of an assessment process continues to be the distribution of the QR element of the block grant, then the process has to be fair, transparent and credible to all external as well as internal parties. Given that the outcomes of the current peer review process have been challenged by Government and others, then surely a self-assessment process will be subject to even more criticism and be seen as even less robust.**

- We also find it difficult to see how a system could be developed which would enable consistent, reliable and fair judgements to be made across the board. Consider the standards which need to be set to enable such judgements to be made:
 - The precise definition of what constitutes the quality of a research output – which presumably needs to be differentiated quite carefully to identify different levels of quality (ie in current terminology international, national and sub-national quality).
 - The precise definition of what constitutes a strong or weak research culture, including an indication of the level of activity (income generation, research student numbers) that equates to a particular level of quality,

Who would make the judgements of the standards to be applied, and when? If before the self-assessment process, where would the data come from for such judgements to be made? If data have to be separately collected and then judgements of quality thresholds made, one might as well run a peer review process (this is one element of it after all) and this would certainly not lessen the burden on institutions. If standards were not defined, it would be impossible for consistent decisions to be reached across the sector.

- One of the strengths of having a team of peer reviewers, ideally at the forefront of their subject, is that there is – or should be - a collective and consistent approach to issues of quality and standards when assessment takes place. This would be impossible to achieve if every subject grouping in every institution were to make its own decisions on how to apply rules unilaterally.
- A self-assessment process will be hugely divisive within an institution and within the sub-groups of that institution, unless there was a requirement for every member of staff to be considered. It is difficult to envisage many heads of subject areas queuing up to undertake the task of judging their colleagues.
- It is difficult to see how any institution – or the constituent elements thereof - which regards itself as placing research at the forefront of its activities would be prepared to submit assessments which do not show the highest possible quality of research being attained. We are reminded of early experiences with Teaching Quality Assessment, which were also based around self-assessment. Did any unit ever label itself as unsatisfactory or even satisfactory? And what happened when a local judgement was challenged by the validators? This particular system evolved over time to a very intrusive form of peer review, which surely says something about the robustness of self-assessment processes. The sector should certainly take the opportunity to learn from this particular process and not make the same mistakes again.
- We find it difficult to see how such a system could be less onerous than the current one. Not only would institutions have to collect a range of data to be judged – as they do at present – they would also have to undertake the judging process itself (not currently the case!), and then presumably experience some form of validation and/or audit of their decision. (The latter would be an absolutely crucial element if such a scheme were put in place).

Historical

12. 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.

13. 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).

We do not consider an historical approach to be acceptable per se, but have proposed above a variant on this which might allow for a “light touch” approach for certain units.

14. 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?

No.

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?

This would encourage complacency and discourage innovation. Research and researchers would follow funds to “historically” strong institutions, with catastrophic separation of research and teaching in a large proportion of the HE sector.

e. What are the major strengths and weaknesses of this approach?

Strengths:

- **Less burdensome to institutions**
- **Would support subject areas in temporary/short-term transition**

Weaknesses:

- **It assumes research strength only changes very slowly and is not responsive to a variety of shorter term influencing factors.**
- **It would encourage maintenance of the status quo and be divisive in terms of equal opportunities.**
- **It would provide little stimulation for improvement to the weaker institutions.**
- **It would discourage innovation**
- **It would not be responsive to new developments in institutions or significant shifts of research teams from one institution to another etc**

E. Cross-Cutting Themes

(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?

Purpose

We accept that the original purpose of the introduction of a research assessment process was to enable the distribution of the QR element of the block grant to institutions, and this will clearly remain an important driver for the process. It is self-evident, however, that the outcomes of the process will be used by research funders to make judgements about the allocation of further grants, by students (especially international students and postgraduate students) to inform their decisions about where to study, by staff when making decisions about career progression, by institutions to publicise their strengths and distinctive characteristics and by the media, especially for league tables. It is difficult to see how any statement issued by the funding councils or others could stop any of these activities, but better information could be provided about the process (its purpose and methodology) and what the outcomes mean to the various stakeholders in higher education

Information Content

We agree there might be some value in designing a process which provides useful management information, but this should not duplicate other data collection exercises (ie HESA) and should be conducted in as technically efficient manner as possible to reduce the administrative burden.

Interaction with Research Councils

We feel that the extent to which it is appropriate to work with the Research Councils needs to be considered very carefully, not least because there are a wide range of other bodies which fund research within the UK, but also because RC assessments should focus on the person/project/programme(s) they have funded, and not other activity, whilst the RAE is aimed at assessing the overall picture. We are concerned that too close an association between the RCs and the Funding Councils with respect to research assessment would undermine the principle of the dual support system.

(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?

Frequency of Review

We would tentatively suggest a rolling review over a period of 5 – 7 years might be possible, with clusters of cognate subjects assessed together. We would note, however, that these groupings would have to be carefully constructed to allow interdisciplinary work to be given appropriate consideration. We would reiterate the need for a longer lead in time for arts subjects (as is currently provided). One possibility might be to review Arts and Social Science subjects at one point and all remaining areas separately (ie a few years before). We recognise, however, the possible significant logistical problems of undertaking a rolling review both locally and nationally.

(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?

Research Excellence

We developed the following definitions internally following the publication of the Panel Criteria.

International quality

International quality research may be defined as including all or some of the following (some criteria are perhaps more discipline specific than others).

- Equivalent to the best work in the discipline being produced anywhere in the world (eg in terms of levels of scholarship, innovation or originality)
- Sets the international research agenda or contributes significantly to its development (particularly relevant to science/medicine)
- Will be work with which any researcher in the field or sub-field should engage
- Will probably be disseminated internationally
- Demonstrates a substantive knowledge of developments in theory or practice internationally
- Most probably rigorously peer- or user-group reviewed
- Includes significant new information or ideas resulting in the modification of existing theories or experimental practice or understanding of mechanisms, or enabling significant new approaches or policies

- Outstanding with respect to originality/innovation, depth, range, accuracy, clarity, impact (the extent to which debate is informed by the work). May include significant empirical findings, conceptual contributions, innovative methodologies and techniques, theoretical developments

Staff producing such work would also be expected to be active participants in international conferences, and be obtaining research funding, supervising research and providing research leadership. Their work will influence the global research community and/or practitioner communities.

National quality

National excellence may be defined as including some or all of the following characteristics:

- Makes a valuable/intellectually substantial contribution to knowledge, understanding, theory or practice and/or original thought. Contains new information or ideas and may contain innovative developments (eg of interpretations, approaches, methods, techniques, deductions or practical applications).
- Worthwhile, timely and technically sound
- Assists in the development of existing theories or experimental practice
- Well-researched and substantive
- Relevant to policy and practice
- May be high quality work within existing research agendas or paradigms
- Research with which other researchers would be expected to be familiar

In addition to the above we would add the point that quality is about fitness for purpose, and that therefore understanding the purposes of research (which range from blue skies thinking to solving practical problems in innovative ways) are properly understood.

If we have any concerns with respect to RAE 2001, it would be with respect to applied research. We have suggested above ways in which these could be addressed.

(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.

Distribution of Funds

We do not believe that research assessment alone should be used to determine the distribution of funds between subjects.

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) Should each institution be assessed in the same way?

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.

Institutional Assessment

There may be some scope for clustering institutions into broad groups, but might this not re-create divides which no longer exist?

(f) Should each subject or group of cognate subjects be assessed in the same way?

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.

Uniformity of Assessment

We consider that there should be overarching principles which are the same across the whole range of subjects, but that as suggested above there may be scope for the establishment of a number of broad groupings which would then develop internally consistent and appropriate approaches.

(g) How much discretion should institutions have in putting together their submissions?

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.

Control

We would certainly not support a system that removed control wholly from institutions. We believe institutions should continue to have a reasonable degree of control over their submissions. We accept, however, that serious consideration should be given to the question of whether all staff should be returned for assessment in future exercises, noting however that this would significantly increase the burden and cost of the exercise for institutions, the assessment panels and the funding councils. A variant on the requirement that all staff be returned would be that a minimum of 70-75% should be returned in units wishing to be graded 5 or 5*. We would strongly urge the Funding Councils to require only those staff who are required to research as part of their contract to be eligible for assessment. We believe it is unfair for institutions to be penalised for staff whose contracts do not require them to undertake research and who are in effect doing precisely what they are employed to do.

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) How can a research assessment process be designed to support equality of treatment for all groups of staff in Higher Education?

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?

Equality of Treatment

The process almost certainly does discriminate against new/younger staff who have not yet built up an adequate body of research, against those returning to work after career breaks or serious illness or entering HE from another sector (practitioners becoming theorists)⁹, and against part-time staff, through the expectation that such individuals will have produced four research outputs and be performing on a similar level to colleagues. Of course, many panel criteria in 2001 specified that this would not be the case, without in most instances being explicit about how they would deal with such cases, and it would seem likely that many institutions elected to be cautious about including such staff as a result. A good way to address this would be to set different and absolutely explicit requirements for such staff (ie two publications not four; pro rata for PT etc), and/or to be explicit about how their output will be treated. A further approach would be to take a more rigorous line on expecting institutions to demonstrate how they manage the career development of staff.

(i) Priorities: what are the most important features of an assessment process?

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).

Important Characteristics

In our view, the following characteristics are essential for the process:

- **Rigor**
- **Fairness**
- **Transparency**

Whilst the following are desirable:

- **Administrative efficiency**
- **Operational ease**

Any Other Points

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.

⁹ A related issue is that of clinical lecturers in Medical units. These staff, when compared with their non-clinical counterparts, are often perceived as being weak on research grounds and many are not submitted as a result. However, these are vital training posts for academic medicine with individuals balancing ongoing clinical training with University commitments. A way to assess appropriate levels of contribution from such staff should also be developed.

Scale Complexity

We have no desire to see the ratings scale expanded, but would like to see it rationalised so that it runs from 1 – 7.