

## Response to the Funding Bodies Review of Research Assessment from the Research Committee of the University of York.

1. GENERAL VIEW. The view of the University of York is that, on balance, the RAE, like the TQA, has been a good thing. It has focussed on quality and has stimulated us to improve our performance. It has allowed a comparatively young university like York to see where it stands relative to other, more established institutions, and it has provided an independent evaluation of the quality of what we do. It has grounded funding in actual performance rather than historic reputation and influence, which can only be a good thing. Hence we support the stated intention to retain the RAE in some form or other and to use it to inform future funding.

But the RAE is about distributing QR. If the funding councils decided to do away with QR, then there would be no further need for the RAE. If the RAE is to continue, and is to be used to inform the distribution of QR, then the funding councils need to face up to some difficult questions. What does it cost to maintain an internationally-competitive department of Biology, or Politics, or English? How much of that cost should be met by the funding councils and how much should departments be expected to find from other sources? Therefore, how many internationally-competitive departments of what size could each funding council hope to sustain in each discipline? In the absence of answers to such hard questions, there is a real danger that grade inflation will result in funding being spread too thinly for *any* departments to operate in serious competition with their international rivals.

If the RAE is to inform the distribution of QR, then the measurement tool needs to be fit for the purpose. Form and function: you cannot design a sensible assessment process without evaluating its ability to perform the job for which it is needed.

2. PHASING AND FREQUENCY. A rolling programme of RAE assessments would have the advantage that institutions would not have to manage such major and unpredictable changes in funding. It would also prevent the pattern of pre-RAE investment and transfer deals followed by a post-RAE hiatus.

The disadvantages of a rolling programme include the fact that, as with TQA, universities would need to have administrators and academic committees that would be devoted only to preparing RAE submissions. At least with the current system the disruption comes just once every 5 or 6 years. With a rolling programme the aforementioned administrators and academic committees would rapidly become more adept at manipulating the system. That would lead to faster grade inflation of the sort that undermined the TQA.

Whether phased or big bang, the frequency of assessment needs to be no higher than necessary yet often enough to be responsive to changes that result from investment in departments or staff turnover. If there is to be no appeal against the judgement of the examiners, then it also needs to be often enough that judgement errors on the part of panels do not permanently blight the lives of staff in hard-done-to departments. On balance we believe that once every 5 or 6 years is about right.

3. WHAT SHOULD THE SCALE BE? At the moment the RAE has 7 grades. The bottom two are seldom used and funding for 3b grades is nothing or very little. Grade inflation across successive RAEs has resulted in a bunching at the top end which in turn has led the funding councils to introduce major differentials between funding at the different levels. For a large submission the consequences of just missing out on a 5\* or a 5 are huge yet unpredictable (because the funding consequences are only announced retrospectively, so institutions are left trying to guess the "correct" trade-off between size of submission, likely grade and resultant funding, and end up being penalised for taking decisions based on inadequate and incomplete information). This illustrates the point made in section 1 above about the need for the assessment scale to be designed for the purpose to which it will be put.

This problem could be reduced if there were more gradations to the scale. One possibility might be to allow scores of 4.9 and 5.1 on something like the current scale. Another would be to have Panels rate submissions on several sub-scales (relating, for example, to quality of publications, research income, health of graduate school etc) then aggregate the score (as with TQA, but taking steps to avoid the ceiling effect of TQA). A ranked output from the assessment exercise rather than a categorical output would allow much fairer distribution of QR.

4. WHO SHOULD DO THE RATING AND HOW? It goes without saying that evaluation needs to be as objective as possible. In some disciplines a combination of factors like amount of peer-reviewed research funding per staff member and impact factors of journals or numbers of citations per paper would predict most of the variance in panel scores. In other disciplines there seems little alternative to the subjective assessments of a panel of experts. Some useful work could be done to discover how well a metrical approach applied to the results of the 2001 RAE could have predicted the actual results. Is there a weighting

of funding, impact factors and citations that predicts rating? Does that weighting vary from UoA to UoA? If there are areas of good agreement, the panels in those areas might be encouraged (or required) to use such judgements to inform their ratings. Objective metrics are not flawless, but neither are the judgements of peers.

5. RETROSPECTIVE OR PROSPECTIVE? We would strongly oppose any attempt to ask panels to make prospective evaluations of how departments will fare in future. Panel members are no more skilled clairvoyants than the rest of us. They cannot know which departments will lose all their best staff in the transfer market and which will gain rapid strength through institutional investment. We would dispute the assumption in Section 16a that the distribution of research strengths is likely to change very slowly. When aggregated across institutions there is no doubt a high correlation between scores of successive RAEs, but we have a UoA that went from 3b in 1996 to 5 in 2001, so at the micro-level (the level at which individual staff members operate), change can be rapid. Evaluations should be based on the *recent past*, not the future or the distant past.

6. GRADE INFLATION AND THE QUESTION OF WHO SHOULD BE REVIEWED. Across all subjects, 39% of submissions were graded 5 or 5\* in 2001 compared with 20% in 2001. One UoA went from 20% 5/5\* grades in 1996 to 63% in 2001, another from 22% to 70%, and a third from 37% to 70%. In a fourth UoA, all submissions were rated 4 in 1996 while in 2001 they were all 5's. This kind of grade inflation should never have been allowed to happen. It threatens to undermine the credibility of the RAE (see the Report of the Science and Technology Select Committee).

Much of the observed 'improvement' in RAE grades was due to game-playing on the part of universities in the area of selective submission of staff. If successive RAEs were genuinely improving research in the UK, then we would expect to see more staff submitted as research-active with each passing RAE. Yet the opposite has happened: supposedly 'research-led' universities omitted 25% or more of their academic staff in 2001 in order to secure the best ratings they could from those they did submit.

The ability to leave staff out of submissions has allowed institutions to make a strategic pitch for grades based on their own evaluations of whether staff were International, National or sub-National. We are aware of examples around the UK of staff who were doing work of clearly National quality being sacrificed from a submission (or moved across to another submission) in order to make the numbers stack up for a 5 or 5\* rating. That should not be allowed to happen again.

The 2001 RAE results included the absurdity (from our point of view) of 5\*D and 5F ratings. How can a department claim on its web site or in its prospectus to be of international standing when more than half of their staff have been declared research-inactive? The requirement on RAE panels to evaluate only those people submitted irrespective of what proportion of researchers in a given area they comprised led panels to focus on RA2 outputs and income per submitted staff member. Sections in RAs 5 and 6 describing the research culture become uninterpretable and therefore unusable when significant proportions of staff are being omitted.

If the funding councils are serious about wanting to assess all research in the UK then they should require institutions to report on the research contributions of all staff on academic contracts and should assess the totality. There should be no declaring research-active staff as inactive, no moving staff sideways from one UoA to another, and no hiding staff in non-declared UoAs. A single Commentary section should replace the current RA5 and RA6, which are hard to differentiate, providing a commentary on the hard data of the earlier sections.

7. CONSISTENCY OF OPERATION ACROSS DISCIPLINES. Our attempts to discover how individual panels operated suggested alarming differences between panels in how they went about their business. Panels dealing with radically different areas will inevitably want to proceed in somewhat different ways, but if the exercise is to retain credibility then as much should be done as possible to ensure that different panels operate the same rules and procedures in the same ways. This should include more regular meetings of Panel chairs before and during the assessment and more authority given to the secretariat to ensure that agreed procedures are adhered to. Greater transparency and consistency can only be to the benefit of the RAE.

8. INTERDISCIPLINARITY. RAE panels always tend to be drawn from the cores of their disciplines. Cross-referral of interdisciplinary submissions only sends the submission from a panel focussed at the core of one discipline to a panel focussed at the core of another discipline. Neither will rate the interdisciplinary work as highly as core research. In our view, the RAE will always exercise a centripetal influence, pulling research back towards the gravitational centres of disciplines. If the funding councils wish to retain the RAE, that is a price they will have to be prepared to pay.

9. RIGHTING WRONGS. Any mechanism relying on human judgement will be imperfect. A submission that lies on the cusp of, say, 4 and 5 might be tipped one way by an RAE Panel and the other way by a Panel with a different composition. The funding implications of such borderline decisions can be colossal. We believe that the assessments of one of two of our submission were unduly harsh.

We appreciate that the funding councils do not want to get involved with large numbers of Appeals, but if something like the current scale is to be retained, and if the funding differentials are to remain so steep, then we would encourage the review group to consider whether some limited form of appeal that acknowledged the inevitable fallibility of panels might not be allowed. One could, for example, allow a maximum of one appeal per institution which would only be considered if the original panel had placed the submission close to a borderline. The effort involved in creating an appeal mechanism should be weighed against the cost to institutions, departments and individual staff of an unfair evaluation.

10. COUNTER INTELLIGENCE. The credibility of the current RAE has been seriously undermined by game playing. We recommend that the review group assembles a shadow group of wily Heads of Department and Pro-Vice Chancellors whose job it would be to devise ways of subverting and twisting to their own ends any proposed set of rules.

26 November 2002