Research Evaluation, 16(1), March 2007, pages 3–12
DOI: 10.3152/095820207X190674; http://www.ingentaconnect.com/content/beech/reev

The UK Research Assessment Exercise: the evolution of a national research evaluation system

Katharine Barker

The Research Assessment Exercise (RAE) represents one of the most institutionalised forms of research evaluation in the OECD economies. It has become a primary means of concentrating resources for research in a relatively small number of universities. Its main purpose is to inform funding decisions, and the indirect effects come from the public signalling of quality. The RAE seems to have reinforced the traditional 'high science' ideals of British universities, encouraging greater coordination of research around traditional disciplinary concerns and inhibiting applied research. Debates among universities and policy-makers have led to a reshaping of the exercise. Its evolution into a highly elaborate procedure combined with its failure as a strategic policy tool to steer university research towards socio-economic impacts have led to its abandonment in its present form.

THE RESEARCH ASSESSMENT EXERCISE (RAE) is the United Kingdom's national research evaluation system, as defined by Whitley (2007).[1] It is a systematic and comprehensive evaluation of all research disciplines with explicit references to grading scales (Campbell, 2003), and covers all higher education institutions, which are mostly but not exclusively universities.[2] The universities are the largest performers in public sector research in the UK. The RAE has evolved through four cycles: 1986,[3] 1989, 1992 and 1996 with most recent exercise in 2001, and a system in place for 'RAE 2008'. It exists to allocate block grant funding according to a predominantly retrospective peer-based quality assessment.

Katharine Barker is at PREST, Manchester Business School, University of Manchester, Oxford Road, Manchester, M13 9PL; Email: kate.barker@mbs.ac.uk; Tel: +44 161 275 5932; Fax: +44 161 275 0923.

An earlier version of this paper was presented at the workshop Changing Knowledge Production through Evaluation, Bielefeld, 9–11 June 2005, funded by the *Sociology of the Sciences Yearbook* and the PRIME Network of Excellence (CIT1-CT-2003-506596), <www.prime-noe.org>. The following colleagues provided useful comments: Luke Georghiou, Philippe Larédo, Vivien Walsh and Richard Whitley.

This paper aims to examine the debates which have led first to an extensive modification for its forthcoming 2008 iteration and more recently a desire by the UK Government to replace it with a metrics-based system of assessment. The RAE has become a dominant phenomenon for UK universities, which devote huge resources to developing their strategies towards it and managing it, as well as donating the time of academics and administrators to serve on and service the panels. The debates about the form and evolution of the RAE demonstrate tensions around the nature of university research (traditional, disciplinary and autonomous versus applied, steered and multidisciplinary). Moreover, the recent increases in public expenditure on science in the UK have an explicitly economic rationale and expected economic returns. It seems that uncertainties about the effectiveness of peer review-based evaluation in allocating resources for economically useful university research have resulted in the desire by government and industry to move to a metrics-based system.

The operation of the RAE

In short, the RAE is a periodic national peer review organized by 'units of assessment', which broadly

relate to disciplines (such as physics) or subject areas (such as area studies). The units of assessment are evaluated by peer review panels mainly made up of UK academics with a few non-UK academic advisors and some representatives of research 'users'. In 2001 there were 69 subject panels of six to 20 members and for RAE 2008 67 panels are grouped under 15 umbrella panels. The peer review has been structured through the implementation of written guidelines for panels, the provision of panel secretaries by the Funding Councils and the requirement of universities to present evidence in a standard format. Up to now, not all academics from a department or subject group needed to be submitted but the final rating for each area in each institution has noted the proportion of staff submitted and subsequent funding has been proportionate to the numbers included.

The outcome of each RAE to date has been quality ratings awarded through this managed *ex-post* peer review exercise. The ratings are used to calculate block grants for research to universities, which for the English Funding Council is £1.3 billion per year. Over 90% of the Funding Councils' research grant to universities is determined by the RAE results. The Funding Councils' block grants are intended to support ongoing research costs, including salaries of permanent staff and infrastructure according to the priorities determined by the universities themselves. This line of funding forms one side of the UK's 'dual support system' for university research. The other component is competitively allocated project-based funding from the Research Councils, awarded through *ex-ante* peer review of grant applications.

In the face of much criticism (see later section) the Funding Councils have been robust in their defence of the RAE. They maintain that it is governed by a set of principles (HEFCE, 1999):

- Clarity: for example, defined criteria.
- Consistency: the assessments should be consistent across cognate areas and in the calibration of quality ratings against international standards.
- Continuity: the RAE develops through an evolutionary process.
- Credibility: to those being assessed (due to the exercise being a peer review of research outputs).
- Efficiency: previous RAEs have been extremely cost-effective given the value of public funds distributed through their ratings.
- Neutrality: the RAE should carry out its function without distorting that which it is measuring.
- Parity: the RAE is concerned only with assessing the quality of research of participating universities regardless of its type, form or place of output.
- Transparency: all decisions and decision-making processes are explained openly.

As we will see, all of the above assertions have been challenged.

The framework of the peer assessment — the 2001 RAE

As stated above, the Funding Councils have asserted that the RAEs have been clearly documented, consistent across academic areas and with continuity between exercises as far as possible. The process required that each 'unit of assessment' in each institution (subject/disciplinary area, often corresponding to a department or school) presented seven sets of data. The core of the assessment was performed on the four items of research output per named member of staff produced during the allotted time period. Other data were provided on all academic staff and support staff, those staff included in the return as 'research active', and data on research students, external research funding and studentships, research environment and 'additional information' including esteem indicators and evidence of impact of research.

A key decision for institutions, and for 2001 dealt with at very senior levels in most institutions, was which members of staff to class as 'research-active',[4] according to the perceived quality of their research outputs. The submission might also have had to present a collection of research staff which do not necessarily work together in one group, but are allocated to the subject area where they are most likely to help achieve a high score. By the 2001 exercise, universities were putting enormous efforts into the wording of these submissions.

The ratings were defined as shown in Table 1. As McNay (1999) points out, these definitions are highly open to interpretation. He provides an account of the lack of consistency in the 1996 RAE between panels' criteria and processes, ranging from marked disparities in how different levels of quality are defined to very different practices in the actual scrutiny of the research outputs. For example, were research outputs read in full, or was the reputation of the journal taken as a proxy of the quality of the article, and what sampling technique was used for the review of the research outputs? There seems to have been considerable variation. According to McNay (1999) the international panellists did not take part in the detailed scrutiny of the research outputs, thus begging the question of how international quality was assessed.

The core of the assessment was performed on the four items of research output per named member of staff produced during the allotted time period

Table 1. Increasing the thresholds of quality definitions: RAE 2001 and RAE 2008 compared

RAE 2001 rating	RAE 2001 definition	RAE 2008 rating	RAE 2008 definition
		4*	Quality that is world-leading in terms of originality, significance and rigour.
5*	Quality that equates to attainable levels of international excellence in more than half of the research activity submitted and attainable levels of national excellence in the remainder.	3*	Quality that is internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence.
5	Quality that equates to attainable levels of international excellence in up to half of the research activity submitted and to attainable levels of national excellence in virtually all of the remainder.	2*	Quality that is recognized internationally in terms of originality, significance and rigour.
4	Quality that equates to attainable levels of national excellence in virtually all of the research activity submitted, showing some evidence of international excellence.	1*	Quality that is recognized nationally in terms of originality, significance and rigour.
3a	Quality that equates to attainable levels of national excellence in over two thirds of the research activity submitted, possibly showing evidence of international excellence.		
3b	Quality that equates to attainable levels of national excellence in more than half of the research activity submitted.		
2	Quality that equates to attainable levels of national excellence in up to half of the research activity submitted.		
1	Quality that equates to attainable levels of national excellence in none, or virtually none, of the research activity submitted.	unclassified	Quality that falls below the standard of nationally recognized work. Or work which does not meet the published definition of research for the purposes of this assessment.

Note: In the 2001 RAE a single rating was given to each submission under a unit of assessment. In RAE 2008, the panels apply ratings to the three elements of the quality profile: research outputs, research environment and esteem indicators. The overall profile combines these to indicate the percentage of the activity reaching each rating.

The funding consequences of the 2001 RAE

While the consequences of the grade awarded were important for reputation, attracting external research income and overseas students,[5] it is the consequences for the research block grant which make the RAE game so crucial for universities. According to the RAE results, 98% of post-1996 Funding Council research budgets were allocated as block grants to universities. This was calculated according to the product of the volume (head count), the subject quantum (to distinguish between the costs of different types of research) and the research rating factor. This formula was applied consistently across all universities. Figure 1 shows the weighting of the RAE results.

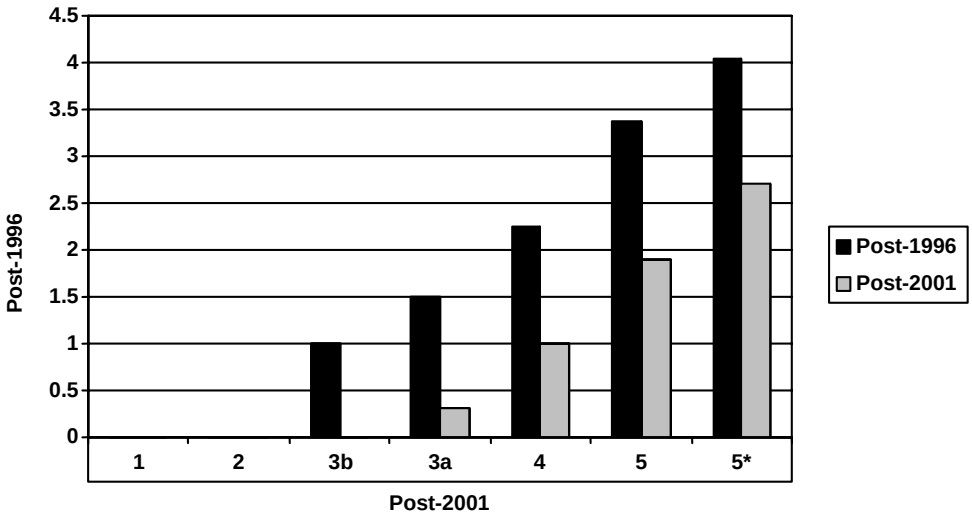

Figure 1. Research rating weights for the quality-related funding formulae (for the Funding Council for England)

The research rating weights thus increasingly diverted funding towards the higher ratings. After the 2001 RAE, a subsequent government paper insisted that the funding was not sufficiently concentrating funding to the very best performers. Those departments which achieved 5* in 1996 and 2001 were rewarded with the *post-hoc* status of 6* and given extra funding, a move which met with some criticism, since it was not subject to any further peer review. While the formulae are not revealed in advance of RAE results (since the Funding Councils have a fixed budget to allocate), after 1996 the drive was to get the highest research rating possible rather than include as many academics as possible. The formulae gave much more significance to research rating than to volume (a point returned to below), and so including 'weaker'[6] staff and dropping a grade had serious funding consequences. After 2001, research having "attainable levels of national excellence in virtually all of the research activity submitted, showing some evidence of international excellence" (the definition of "4") had much reduced funding compared to the 5* grade, which received over two and a half times as much.

Impacts of the RAE — academic research and UK universities

What evidence exists that the RAE has changed academic practice and knowledge production? *A priori*, one can make arguments for and against the assertion that the RAE would certainly change the nature of academic research in universities. To start with arguments against, since the RAE is an *ex-post* scrutiny of research outputs which is carefully managed by the Funding Councils to achieve the principles outlined above, it should not change academic practice, since research is continuously subject to peer review in any case. The RAE has been supposed to treat all types of research output equally and to take user views into account. The main reason why one would expect the RAE to have a major effect on academic practice is the strong funding steer which is applied according to the grade achieved by each subject area in each university. The incentives to perform are very strong and the consequences of failure are severe. Thus, universities and academics have been drawn into a strategic game, and invest tremendous efforts in trying to win.

However, does game-playing for the RAE change academic practice significantly? Are the changes in university management, the organisation of research and the development of university strategy in response to the RAE fundamentally affecting the choice of problem area, the tools of enquiry and the intellectual agenda? If such effects are there, can they be attributed to the RAE as opposed to ongoing changes in the research landscape? Finally, with such a highly stratified university sector in the UK where research funding from non-RAE sources (industry, charities and competitive project grants) is already highly concentrated into around 20 institutions, there will be very different effects on academic practice depending on the nature of the university.

Mary Henkel (1999), for instance, suggests that the RAE can be seen as an attempt to give academic control back to universities in a period of funding constraints through the harnessing of peer review and internal quality control. However, she describes a "profound disturbance in the academic system" (p. 106) in linking the peer assessment to strong funding concentration. However, Ball (1997) and Henkel (1999) point out that the RAE has highlighted the importance of research in universities and particularly in the post-1992 universities[7] where block funding for research was not previously available. While not starting from a level playing-field, at least these institutions could receive some funding in recognition of areas where they had developed research activity.

The 1996 and 2001 exercises provoked a flurry of academic articles and journal editorials giving a view from particular disciplines. Harley (2002) (reported in Hare, 2002) surveyed over 800 academic staff in business and social sciences and found varying views. Some academics felt their work was validated and others found difficult the increasing managerial control of research by institutions and departments. Applied disciplines such as planning have suggested that the RAE fails to take into account high-quality strategic and applied research. The emphasis of the panels is on refereed publications, particularly in top international journals, thus pushing apart research and practice with potentially negative impacts on the delivery of teaching, which requires individuals who are engaged with practice (Nadin, 1997). Similar arguments are made for law (Vick *et al*, 1998). The RAE, it was felt by such groups, pushes a uni-dimensional view of research which may not be relevant to the diverse areas of practice in a profession like planning and law.

Similarly, nursing researchers have provided reflections on the RAE in their subject, which was ranked last in terms of the grades awarded to units of assessment (Tierney, 1994; Traynor and Rafferty, 1999). Nursing was a late-comer into higher education, and mainly located in post-1992 universities which mostly did not emphasise research in their missions. Nursing and clinical medicine have faced competing pressures from the RAE and the National Health Service, and academic staff have the tasks of both training health care professionals and delivering health care, on top of which a research imperative has been added from the RAE. Clinical medicine generally fared badly in the RAE, which has created tensions between research, teaching and the practice of medicine in many university medical schools (Symonds, 1999). McNay (2003) concludes that with the exceptions of social work which was reviewed with social policy, and accounting and

Interdisciplinary areas also seem to have been discriminated against, although evidence for an actual retreat from such research is not easily found

finance which was "deviant" in its very high grades (McNay, 2003: 51), professional areas have fared worse than traditional academic disciplines. Interdisciplinary areas also seem to have been discriminated against (eg environmental sciences), although evidence for an actual retreat from such research is not easily found.

Other concerns are the irreversibilities of the outcomes: reinforcing the differences between those departments which do well and obtain the block grant funding and those which do less well and are effectively denied the funding, thus making re-entry unlikely. At institutional level, universities are free to use the block grant strategically: they could invest in poorly performing departments, but would risk failing to maintain their top areas. In practice, there are several examples of departments achieving a poor RAE grade in research-led universities being closed or re-organised. Successive RAEs have had remarkably consistent results. McNay (2003) reports that after RAE 2001 174 institutions submitted returns and 75% of the funding went to 24 of them: almost no change from RAE 1996 when 26 institutions out of 192 taking part took 75%.

The funding consequences of the RAE act, as stated above, on all universities, without any 'policy factor'. There is no attempt to distinguish between units of assessment (subjects/disciplines) on the basis of the overall quality of the UK's standing in that area. Thus, there is no attempt to apply selectivity criteria; it is about concentration within units of assessment and not strategic promotion of different research themes. Industry and other influential users of research including some government ministries have remained hostile, mainly for undervaluing applied research.

Game-playing, grade inflation and dissatisfaction

It can be seen that the design of the ratings and the funding model gave opportunities for game-playing. The volume (people entered) is arithmetic, while the rating factor is geometric. As described above, there was a strong incentive to leave out weaker researchers to gain higher rating. Many universities in the 2001 RAE tried to maintain their ratings but include as many researchers as possible, but were shocked to find that the funding for "4"s and below was much reduced. This problem was reinforced by the growing indirect benefits of rating — other research funders including industry tend to use rating in search process for partners — external research funding for departments rated 4 or higher increased by 61% since 1996.

Between the 1996 and 2001 RAEs, there appeared to be a remarkable improvement in the quality of research, in terms of the outcomes for the units of assessment and the numbers of staff receiving funding and located in 5 or 5* departments.

The general consensus is that the improvement was probably due to three causes:

- Grade inflation, which is well known in peer review when resources are constrained (Gibbons and Georghiou, 1988);
- Strategic behaviour and 'game-playing', that is weaker researchers not included;
- Actual improvement in research due to the incentive of the evaluation system and research output management in universities.

The 'improvement' in research took place in a modest proportion of universities.

What about the benefits from successive RAEs? Universities and academic staff see some positive effects. Selectively directing funding to the highest-rated departments has some justice, but the degree of selectivity has caused concern. It has been accepted as the price to pay for the block grant system in the UK's dual support for university research, and it has raised the profile of research with some evidence of improved quality. The top-performing academics have seen individual benefits in terms of increased salaries and attractive packages for relocation as universities compete for the best researchers. For post-1992 universities it has given some a new source of income and allowed pockets of research to emerge. For middle-ranking universities the sums involved are too great to ignore.

In recognition that the previous RAE model had outlived its usefulness the UK funding bodies commissioned a review of the RAE headed by Sir Gareth Roberts (2003). The report was published in May 2003 and the review was based on consultation,

Table 2. RAE outcomes in 1996 and 2001

	RAE 1996	RAE 2001
Number of submissions	2,874	2,728
% gaining a funded grade	76	83
% rated 5 or 5*	20	39
Total staff submitted	48,072	48,022
% staff funded	88	92
% staff rated 5 or 5*	31	55

Source: adapted from McNay (2003: 51)

meetings with stakeholders, workshops, operational review and international review. The perceived difficulties with the RAE which the review addressed were:

1. Very small numbers of staff on the borderline could affect the financial outcome for a research area, thus encouraging institutions to leave out staff to raise the overall rating ("cliff-edge").
2. The ratings conceal great unevenness in quality within an area (hiding "pockets of excellence").
3. The composition of ratings was not transparent (publications/environment/esteem).
4. The consistency across panels doubted.

The review suggested some solutions to the problems above and was highly influential in shaping the form of the next RAE.

Raising the stakes, changing the impacts? RAE 2008

First, a revised panel structure grouped cognate areas together, so that 67 units of assessment gather under 15 main panels made up of the chairs of the sub-panels. This was intended to lead to greater consistency of panel operation and fairer assessment of interdisciplinary areas. Second, a graded quality profile replaces the single grade. The main panels elaborate criteria and definitions for the peer review, which are all made publicly available, and they decide on how the quality profile is constituted from the elements of research outputs, research environment and esteem indicators. Submissions no longer have to state what proportion of staff they submit. According to the Funding Councils, the panels include international peers and user representatives, although it seems that some international input is via advisors.

The idea of a different assessment track for the least research-intensive institutions was not adopted. There is no obligation for institutions to take part if they judge that it is not worth the overhead. There will not be a separate assessment of institutional research strategies and dissemination of research beyond academia. For RAE 2008, universities are left with a new system to 'second guess'. The generic definitions of the ratings for RAE 2008 (see below) indicate a further strong steer towards international excellence, probably in reaction to the grade inflation of the 2001 round.

RAE 2008 has a five-point scale, but the scoring system is in three parts. Panels will judge what proportion of the activity in each heading meets the quality definitions. A highly elaborate rating system results, with the aim of giving a more accurate and holistic picture of quality and removing the cliff-edge cut-off points. The profiles are produced by combining three separate sub-profiles for research outputs, environment and esteem. Research outputs are mostly journal articles and conference papers but may be outputs embodying research such as devices and intellectual property. Environment includes data on research students, research income and organisational structures conducive to research, while esteem relates to indicators of national and international recognition such as prizes, fellowships and membership of advisory bodies.

The sub-panels and main panels have defined the weightings given to the three elements and within the elements what kinds of indicators to include. For example, the panel for economics and econometrics gives a 10% weighting to esteem indicators and 70% for outputs, while physics weights esteem indicators at 20% and outputs at 60%.[8] How the decisions will actually be implemented and what differences these distinctions will make is unclear. For example, while it is easy to see how a distribution of publications can be generated it is far less obvious how esteem and environment will translate into profiles. The rather broader scope of what is taken into account in RAE 2008 will make academic managers think beyond lists of publications and force consideration of the conditions for undertaking research as well as attracting stars with esteem indicators and 4* publications.

The quality levels are much more ambitious than those defined for the previous RAE, when the top grade of 5* could be obtained for having international excellence in more than half the research activity submitted and national in the remainder. This is replaced with needing to be "world-leading in terms of originality, significance and rigour". Table 1 compares the rating scales, and tries to equate the 2001 and 2008 definitions. Roughly speaking, 2008 removes the distinctions at the lower end of the scale, and adds a new higher level at the top end. The push appears to be for even greater attention to international excellence, and it is widely expected that the consequences of RAE 2008 will be to channel funding towards the highest grades.

Panels are charged with a complex task of breaking down the elements of research output, environment and esteem indicators and constructing a profile which will give institutions a set of ratings in a grid by unit of assessment. Figure 2 is the worked example given by the HEFCE. The quality profile in the example shows that 10% of the activity is unclassified and 15% is world-leading. The panels work to guidance on rounding the profiles.

Similarly to the 2001 RAE, the submissions include data on staff, research students, research income, research outputs and the outputs themselves and the supporting statement on research environment and esteem indicators. It is clearly an elaborate scheme. While one intention is to discourage simple ratings and rankings at institutional level, naturally rankings could be produced from RAE 2008 results and certainly will be (at least by the press).

The strong presumption is that the funding will be concentrated at least as strongly if not even more

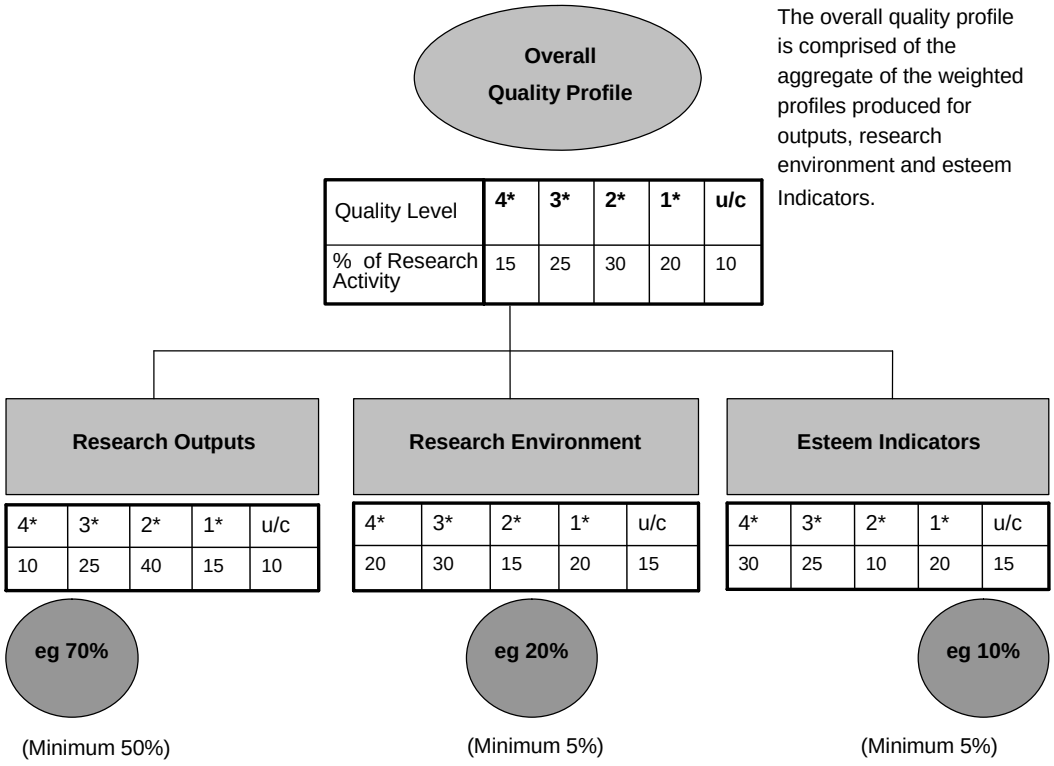

Figure 2. Building a Quality Profile
Source: HEFCE 2006

strongly in the top-rated institutions and the reputational effects may be even stronger than in RAE 2001 because of the raised thresholds. Recruiting and retaining 'stars' has become even more important. However, the profile system means that including weaker staff has less impact on the direct funding outcome than in the previous single grade system.

The battle of metrics and peer review: contesting the future of the RAE

After the Roberts Review explicitly rejected the use of performance indicators except as an aid to expert judgment, the higher education community was taken by surprise in March 2006 when the Treasury announced a policy update on the Ten-Year Science and Innovation Investment Framework (Stationery Office 2004). The Ten-Year Framework, published in July 2004 alongside the government's spending review, promised additional public spending on science with the expectation that it would be a key contributor to the UK's international economic competitiveness and quality of public services. It was positive about the RAE, stating that the reforms for RAE 2008 would "provide greater reward, and thus stronger incentives, for academics to work on both research relevant to users and work which crosses disciplinary boundaries".

However, the update, called *Next Steps*,[9] stated:

> in theory, the RAE is supposed to reward excellent user-focused research in the same way that it rewards excellent curiosity-driven research, but it is not at all clear that this has occurred in practice. (para 4.7)

Similar doubts were expressed about the treatment of interdisciplinary research, with the peer review process here (and in the Research Councils) seen as a bastion of conservatism; the adverse behavioural effects of the RAE on publishing and recruitment giving cyclical effects, which inhibit planning in universities; and its spiraling cost, estimated (by the HEFCE) to be at least £45 million for RAE 2008.

The Treasury presented its preferred solution that the RAE would be reformulated to a predominantly metrics-based system, and that RAE 2008 could be abandoned if a majority of universities were in agreement. Far from pleasing most of the research-intensive universities (with the exception of a couple of high-profile vice chancellors), despite many complaints about the elaborate nature of RAE 2008, the idea of abandoning it caused general consternation and confusion in the attempts to define what a metrics systems might look like and what the outcomes

However, the profile system means that including weaker staff has less impact on the direct funding outcome than in the previous single grade system

Reflection on the evaluation process shows strong stakeholder engagement, through the participation of academics on peer review panels and frequent reviews of the RAE

might be. Post-1992 universities which may have applied research rather than Research Council grants were generally in favour of rewarding all external research income.

The Department for Education and Skills launched a four-month consultation in June 2006.[10] Five funding models were proposed and all were based on research income and volume measures. Research income taken into account was from research councils, government and the National Health Service, charities, domestic and international business. Research volume measures include research-active academic staff, research staff, research students and research assistants. Only outcomes for science, technology, engineering, mathematics and medicine were modeled, with the provision for developing metrics for the arts and humanities. The main arguments debated are shown in Table 3, including points about research income and bibliometric indicators.

While the details of developing and implementing a metrics-based system for the UK will produce fierce ongoing arguments, more important is the policy debate underlying the proposed changes. It appears to demonstrate disillusionment on the part of policy-makers with peer review. It is acknowledged in the *Next Steps* document that successive RAEs have improved the international quality of UK university research (as measured by publications and citations) and have led to better planning and management of research within institutions. But the policy rationale behind the Ten-Year Investment Framework is clearly an economic one. The performance targets for the Ten-Year spending relate to investment in R&D, socio-economic impact and the UK's performance at the top of world science. UK industrial expenditure on R&D has been falling, and the research councils are under criticism for failing to engage with knowledge transfer agendas.[11] Peer review is the culprit for promoting conservatism, risk avoidance and 'silo' research.

The academic research community (that is the 19 research-intensive universities of the Russell group) will not relinquish peer review easily: the RAE is just as important as a label of quality as it is a funding distribution mechanism. The Treasury may wish to retain the QR funds as a source of strategic investment and independence for universities to pursue excellent research. However, it wishes to broaden what 'excellence' means beyond only the satisfaction of criteria internal to science.

Conclusions

The national evaluation of research in UK higher education establishments is well entrenched. The RAE, starting quietly back in the 1980s, has evolved into a highly elaborate, time-consuming and expensive exercise with high-profile results and significant effects on funding. Reflection on the evaluation process shows strong stakeholder engagement,

Table 3. The metrics debate

Arguments for a metrics-driven system	Arguments against a metrics-driven system
More efficient in terms of direct and indirect costs of evaluation to funding councils and universities	Developing, collecting and verifying indicators likely to be costly in itself
Ongoing collection of indicators will reduce cyclical patterns of publishing and recruiting and RAE 'planning blight'	Indicators will have to be taken over three to five years to even out unstable effects of research income
Good correlation between QR income and research council and other external income – very high correlation in some areas such as chemistry	There are outlying institutions, particularly smaller ones, specialist ones
Peer review-based system leads to cautious research agendas	Metrics alone will not have the trust of the academic community since quality of outputs is not scrutinised
Peer review-based system does not encourage applied, multidisciplinary or industry-relevant research	Will distort institutional behaviour to supporting high cost capital-intensive science
Different combinations of indicators could be used for different subject areas	Will make research council grants even more competitive
Different combinations of indicators could be used for different kinds of universities	Would reward industry-funded research which might be of low quality (consultancy/testing) (Laudel, 2005)
Could be combined with 'light touch' peer review	Unclear how metrics can be applied to social sciences, arts and humanities
The funding allocations could be dampened to prevent large year-by-year funding changes to institutions	There are known problems with citation analysis and journal impact factors (Weingart, 2005)

through the participation of academics on peer review panels and frequent reviews of the RAE. However, outcome is very much an elite system to reward research-based institutions. The uneven returns of the RAE match and reinforce the uneven performance in research. We can see behaviour typical of peer review: grade inflation, reinforcement of existing hierarchies, underrating of areas which are not strongly disciplinary or theoretical and much uncertainty about consistency and comparability of the peer judgments which make up the results. Substituting direct peer review with metrics would remove many of these problems but bring other ones.

What difference has the RAE made over time to UK university research? It has largely reinforced the established order. The block research grant has not been diluted by the inclusion of the former polytechnics, since they receive the tail of the funding. Has there been much change over time? While individual institutions have had good or bad RAEs, according to the bets they placed on the funding formula, the overall ranking of research-led universities has not changed a great deal. Neither have we seen the emergence of a former polytechnic as a leading research university: the cumulative effects are too strong.

This leads us then to ask the question, why do we still have the RAE? Is it still relevant to have a national self-referential exercise when the benchmark is international quality and other rankings such as Shanghai are so prominent? Is international excellence in research the key contributor to universities' roles in promoting economic growth and regeneration? Will the universities continue to engage with the RAE while the proportion of their funding from this source reduces?

The frequent reviews of the RAE represent repeated consultation with the stakeholders on the evaluation process. However, the principle of allocating the block grant according to past performance to create winners and losers has not been up for negotiation. This means that some form of national evaluation will be present for UK university research in the foreseeable future. For the teaching-led institutions, so far the RAE still yields some funding which otherwise would not be there. The reputational effects of the RAE are so strong, and become stronger in a more competitive and internationalised environment, that the research-led universities have to take part. Yet other forms of signalling, such as ranking exercises and self-organised reviews, may take away the dominance of the RAE results.

From a UK science policy perspective, the Treasury has noticed that the RAE will not deliver strategic policy outcomes beyond helping to keep some UK academic science at world-class level. Even here, it is not certain that giving the best-resourced universities even more funding gives the best scientific return on investment. The peer review based evaluation system of the RAE is not the right instrument to deliver short- to medium-term socio-economic benefits from research funding or to steer universities to engage with industry, business and their regional economies.

Update

The outcome of the consultation is that the post-2008 RAE will be more heavily metrics-based for all subjects. An overall framework for the future RAEs is set out which allows for disciplinary variation. All subjects will collect data on research income, postgraduate numbers and "quality". Science, engineering, technology and medicine will use a quality indicator based on bibliometrics, while other subjects, including social sciences, humanities and mathematics, will use a peer review process which will be "lighter touch". The metrics will be overseen by a small number of advisory groups, and discipline specific indicators may be used. Data collection will start in 2009 and the changes be phased in gradually between 2010 and 2014.

Notes

1. While the RAE can be described as the national evaluation, it is in fact a joint exercise for the separate bodies which provide funding to the universities for teaching and research: the Higher Education Funding Council for England (HEFCE), the Scottish Higher Education Funding Council, the Higher Education Funding Council for Wales, and the Department for Employment and Learning Northern Ireland. It is managed by a team based at HEFCE, which is by far the largest grant-giver of the four bodies. The HEFCE is a non-departmental public body which is charged by its parent department, the Department for Education and Skills, with administering public funding to universities for teaching and research. Their strategic aim with respect to research funding is "to develop and sustain a dynamic and internationally competitive research sector that makes a major contribution to economic prosperity and national wellbeing and to the expansion and dissemination of knowledge".
2. For the purposes of this paper, the term 'universities' will be used.
3. Called research selectivity exercises until 1992.
4. This formal labelling of academics as 'research-active' or, presumably, 'research-inactive' was highly contentious.
5. Some overseas bodies would only fund postgraduate research student in top-rated departments.
6. Using 'weaker' in terms of the expected contribution to a favourable peer review rating in RAE terms.
7. In 1992 the former polytechnics left local authority control and gained university status.
8. Research outputs must have at least 50% weighting and the other two elements at least 5% each.
9. *Science and Innovation Investment Framework 2004–2014: Next Steps*, March 2006 <http://www.hm-treasury.gov.uk/media/D2E/4B/bud06_science_332v1.pdf>, last accessed 19 March 2007.
10. Department for Education and Skills, *Reform of Higher Education Research Assessment and Funding* <http://www.dfes.gov.uk/consultations/conResults.cfm?consultationId=1404>, last accessed 19 March 2007.
11. Research Councils UK, *Knowledge Transfer in the Eight Research Councils: the External Challenge*, available at <http://www.rcuk.ac.uk/innovation/rolerc/exchallenge.htm>, last accessed 19 March 2007.

References

Ball, D F 1997. Quality measurement as a basis for resource allocation: research assessment exercises in United Kingdom universities. *R&D Management*, **27**(3), 281–289.

Barnett, R 1996. The evaluation of the higher education system in the United Kingdom. In *The Evaluation of Higher Education Systems*, ed. R Cowen, pp. 144–158. London: Kogan Page.

Campbell, David F J 2003. The evaluation of university research in the United Kingdom and the Netherlands, Germany and Austria. In *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, eds. Philip Shapira and Stefan Kuhlman. Cheltenham: Edward Elgar.

Hare, Paul 2002. The UK's Research Assessment Exercise Impact on Institutions, Departments, Individuals. Paper prepared for OECD Conference on Higher Education on the theme of Incentives and Accountability – Instruments of Change in Higher Education, Paris, September 2002.

Harley, Sandra 2002. The impact of research selectivity on academic work and identity in UK universities. *Studies in Higher Eduction*, **27**(2), 187–205.

HEFCE, Higher Education Funding Council for England 1999. *Research Assessment Exercise 2001: Assessment Panels; Criteria and Working Methods*, Bristol: HEFCE.

HEFCE, Higher Education Funding Council for England 2006. *RAE 2008 Panel Criteria and Working Methods*, ref RAE 01/2006. Available from <www.rae.ac.uk/pubs/2006/01>, last accessed 19 March 2007.

Henkel, M 1999. The modernisation of research evaluation: the case of the UK. *Higher Education*, **38**(1), 105–122.

HM Treasury 2006. Science and innovation investment framework 2004-2014: next steps". March 2006. Available from http://www.hm-treasury.gov.uk/budget/budget_06/assoc_docs/bud_bud06_adscience.cfm>, last accessed 19 March 2007.

Laudel, Grit 2005. Is external research funding a valid indicator for research performance?, *Research Evaluation*, **14**(1), 27–34.

McNay, Ian 1999. The paradoxes of research assessment and funding. In *Changing Relationships between Higher Education and the State*, eds. B Little and M Henkel. London: Jessica Kingsley.

McNay, Ian 2003. Assessing the assessment: an analysis of the UK Research Assessment Exercise, 2001, and its outcomes, with special reference to research in education. *Science and Public Policy*, **30**(1), 47–54.

Nadin, Vincent 1997. Widening the gap between the haves and have nots in research funding: the UK Research Assessment Exercise. Editorial, *Planning Practice and Research*, **12**(2), 93–97.

Roberts, Gareth 2003. *Review of Research Assessment*, report to the funding bodies. Available at <http://www.ra-review.ac.uk/reports/roberts.asp>, last accessed 19 March 2007.

Stationery Office 2004. *Science and Innovation Investment Framework 2004-2014*. London: The Stationery Office.

Symonds, E Malcolm 1999. The impact of the Research Assessment Exercise on the funding of medical schools. Editorial, *Journal of Public Health Medicine*, **21**(2), 123–124.

Tierney, A J 1994. An analysis of nursing's performance in the 1992 assessment of research in British universities. *Journal of Advanced Nursing*, **19**(3), 593–602.

Traynor, Michael and Anne Marie Rafferty 1999. Nursing and the research assessment exercise: past, present and future. *Journal of Advanced Nursing*, **30**(1), 186–192.

UGC, University Grants Committee 1984. *A Strategy for Higher Education into the 1990s*. UGC.

Vick, D W, A D Murray, G F Little and K Campbell 1998. The perceptions of academic lawyers concerning the effects of the United Kingdom's research assessment exercise. *Journal of Law and Society*, **25**(4), 536–561.

Weingart, Peter 2005. Impact of bibliometrics upon the science system: inadvertent consequences? *Scientometrics*, **62**, 67–85.

Whitley, Richard 2007. The changing governance of the public sciences: the consequences of research evaluation systems. In *The Changing Governance of the Sciences: the Advent of Research Evaluation Systems*, eds. Richard Whitley, Jochen Gläser and Katharine Barker. Springer (in press).

Whitley, Richard, Jochen Gläser and Katharine Barker eds. 2007. *The Changing Governance of the Sciences: the Advent of Research Evaluation Systems.* Springer (in press).