



THE EVOLUTION OF THE UK'S RESEARCH ASSESSMENT EXERCISE:

PUBLICATIONS, PERFORMANCE AND PERCEPTIONS

Journal of Educational Administration and History, 2005, 37 (2), 137-155.

Valerie Bence

and

Charles Oppenheim*

Department of Information Science,

Loughborough University,

Loughborough, Leics LE11 3TU

* To whom correspondence should be addressed

ABSTRACT

This article examines how the UK Research Assessment Exercise (RAE), along with the development of performance indicators, fits into the production of research within a changing policy context. The historical evolution of the RAE from 1985-2001 is considered, before looking specifically at how research output (in the form of journal publications) was, and is, assessed. Such changes in output measurements necessitated a move from quantitative to quality evaluation, and the role of quality – perceptions and performance- is examined.

INTRODUCTION

In 1985, the UK University Grants Committee (UGC) took a step towards establishing a national system for the evaluation of research in universities by announcing the first Research Assessment Exercise (hereafter the RAE), which would take place in 1986. This evaluation would facilitate selective research funding across university departments. The subsequent development of the RAE as a discriminating funding allocation system has been one part of complex policy changes through which a new set of relationships between higher education and the state would be negotiated during the 1980s and 1990s. This article explores the relationship between the RAE, the production and output of research and one of the main vehicles for that output, academic journals.

The perception of quality has long played an important, if subliminal role, in journal evaluation and submission decisions (Fry et al, 1985; Barma et al, 1991; Campbell et al, 1999; Brinn and Jones 1996) The use of peer review and refereeing papers submitted for publication has evolved to become a self-policing mechanism, arguably for the community, by the community, which attempts to maintain quality standards and to an extent guard the reputation of individual journals. Peer review is also the basis for research assessment via the UK's RAE whereby, "the 2001 RAE will follow broadly the same approach as previous exercises....submissions will be made to a number of subject-based Units of Assessment (UoAs) and information supplied by HEIs will provide the basis for *peer review assessment* of research quality by specialist panels"(Higher Education Funding Council for England, 1999). In this article, the evolution of the RAE to become a peer reviewed assessment system will be tracked, and the position of journal publications within that system reviewed.

The RAE and the Research Production Model

Whilst performance indicators and other forms of evaluation can be implemented at all stages of the research production process, the focus of this work is on how the RAE assesses research outputs¹, which in turn determines the funding needed by universities for further commissioning of the process itself. Figure 1 below, has been adapted to show where the RAE (and such output) fits into the research production model and how the results feed income back into the process.

Figure 1 The Research Production Model

In blunt terms, the extent to which research is published and further cited in recognised journals has been accepted a major basis for developing a metric or measurement of the value of the research in question via the RAE (Cave et al, 1997). The publications of a department have always been important informal indicators of research activity, and it was never disputed that they would form one of the accepted performance indicators (hereafter PIs) for university research assessment. The move to include such outputs as a quality or performance indicator, meant that many decisions had to be made; the types of publications to be included, weightings (if any) to be given to various types of publication; sources of information about publications; whether to count total publications for a department or for each member of staff etc. Thus, from 1986, academic departments of British universities would be primarily assessed in terms of the value of this research output for the first time and this assessment would provide the basis for the distribution of a significant part of their future income.

Performance Indicators in Higher Education

In recent years, the UK higher education system has developed a range of statistics in an attempt to measure its performance, account for the public funding it receives, inform management

¹ The assessment of research inputs is acknowledged as part of the process (i.e., received research income etc) but is not under consideration here.

decisions, satisfy students, employers and other 'customers' of higher education, of the quality of the service it provides, and give assurance that academic standards are being maintained.

The development of such statistics, or PIs, use in higher education has been a long process and is well documented (Sizer, 1979; Cuenin, 1986; Cave et al, 1997) but despite the growing attention paid to PIs in this area, it is difficult to find a single authoritative definition of them. Cuenin (1986) draws a distinction between simple indicators, performance indicators and general indicators. Simple indicators, he suggests, are usually expressed in the form of absolute figures and are intended to provide a relatively unbiased description of a situation or process. This could correspond to the UGC's early use of the term 'management statistics' (University Grants Committee, 1987). General indicators, are mainly derived from outside the institution and can be as diverse as opinions, survey findings or general statistics. Cave et al (1988) concluded that "the peer review exercise carried out by the UGC in 1986 into the quality of research in UK universities was an attempt, to convert a general indicator into a PI.". PIs however, differ from simple indicators in that they imply a point of reference, a standard, an assessment or comparator, and are therefore relative rather than absolute in character, "although a simple indicator is the more neutral of the two, it may become a PI, if a value judgement is involved"(Cave et al, 1988, p. 17).

In 1986 a joint CVCP/UGC² Working Group defined PIs as "statements, usually quantified, of resources employed and achievements secured in areas relevant to the particular objective of the enterprise" and suggested they should be used as signals or guides rather than absolute measures. Their 1986 statement (Committee of Vice Chancellors and Principals, 1986) also suggested that the percentage of staff in a cost centre, who had NOT published during the preceding 3 years could be a PI. Such decisions became controversial when PIs became a potential element in resource allocation decisions or when they affected individual or departmental reputations. For

² Committee of Vice Chancellors and Principals/Universities Grants Committee

example, Crewe's (1987) study of British university publication rates (for Politics departments) was not published in the journal for which it was originally accepted, for fear of litigation.

As a PI of research, publications can provide useful information to policy makers. But during the late 1980s, debate centered on quantity versus quality. Some publication rankings do take account of quality (for example by counting articles in 'top' journals only, by using different weightings, by counting books reviewed in 'top' journals or by the using citation impact of journals), but many believe they are not a real measure of the quality of the research being carried out – always assuming that agreement has been reached on what constitutes the 'top' journals in the field.

Another danger is that concentrating simply on numbers of publications produced could lead to an over-production of poor quality articles. The most pressing technical problem is that not all journals are of equal worth; therefore, any concentration on top or core journals within a discipline is purely subjective. Nonetheless, Gillette (1986) voiced the opinion after the first RAE, that publications provided "the most valid, fair and direct way to compare the research performance of departments".

Traditional 'reputational rankings' (via peer review) can encompass department, institutional or journal ranking judgements. Cave et al. (1997, p. 203) suggested they are "easier and more appropriate to convert into quantitative PIs than other forms of peer review but there are various problems with them". These problems include the fact that they may lack objectivity and can be over-influenced by tradition and departmental size. Cave et al (1997, p. 205) further state that unless such reputational rankings are strongly informed by quantitative data, there is an inevitable tendency for there to be a bias in favour of larger and therefore better known departments. In addition, reputational rankings may do less justice to departments straddling

disciplines or with specialist sub-divisions, unless care is taken to include explicit judgements on them all.

These and other issues are important caveats about the use of PIs (Johnes and Taylor, 1990; Cave et al, 1997). It is doubtful whether, or how, indicators can give an accurate idea of the degree of 'value added' provided by an institution, author or publication. Many feel that the very existence and nature of the RAE, particularly because of the resource allocation aspect, may be determining rather than measuring the way research is conducted in universities (Lee and Harley, 1998). However, by 1986 universities knew that the UGC was ranking research when making decisions about resource funding, but the ways in which these rankings were produced were not explicit and this led to calls for more transparent, objective PIs (Rogers and Scratcherd, 1986).

So, individuals have had to make subjective choices regarding **where** to place research output; **which** publications/journals to choose and **how** they are rated and evaluated – little, if any, of which is made explicit, open or known. Reputation and perception are therefore crucial factors. With publications, there is a large range of publication types, and decisions have to be made as to how to rank the various types – and then more specifically how to rank different journals within that category alone. In many instances, definitions of quality rest largely on their reputation and perception – which are very difficult to quantify. Different methods of assessing quality, primarily peer review and impact measures using citation counts for individual journals, do seem to arrive at broadly comparable ratings (Baird and Oppenheim, 1994; Hemlin, 1996). However, these are assessments of the research outlets and not the research itself. Comparisons are made all the more difficult by the fact that the nature and quality of individual articles within journals will vary, just as the nature of individual journals varies. With no objective process, system or measurement in view, individuals and departments are left with subjectivity only. This has led

Campbell et al (1999) to state “it is *perceptions* of quality that influence departmental decision-making in connection with the RAE”.

There is also danger of a ‘halo effect’ (Jones et al, 1992), whereby departments may benefit unduly from the reputation of the institution as a whole, or an article benefit either by the reputation of the publication vehicle (i.e., the journal) or by the reputation of the author or the author’s institution. A journal’s reputation is founded on the reputation of its authors, editors and referees, and Wells (1999) concludes that this also affects the perceived quality of the research itself.

In some disciplines, journals have been ranked into ‘league tables’. This can be done several ways, including by peer assessment or impact measures. Such studies are undertaken for a variety of reasons (Extejt and Smith, 1990; Campbell et al, 1999). Some have been prompted by the US academic community’s need for evaluating publication records for tenure and promotion decisions and others simply to assist academics best target their research output (Judge, 1994).

The changing policy context and the development of the RAE

The evolution of the RAE took place during the debate surrounding PIs. All institutions were increasingly being asked to justify their performance and account for their use of resources to external funding bodies. The evaluation of research performance in the UK during the 1980s and 1990s should be viewed alongside changes in the structure of university funding, the increasing importance attached to assessing quality and the development of performance indicators to facilitate this. The move to quality assurance and assessment was to assume an increasingly important role in (the then) government’s determination to impose some of the disciplines of the market on higher education, including competition, an increase in the power of consumer demand and the concept of universities as well-managed corporate enterprises. The broader

policy context was that government also expected that resources should be managed with maximum efficiency and accountability, if necessary creating structures for effective decision-making, transparent measurement of performance and ultimately for resource allocation with reference to that performance.

At the same time, rapid rises in costs and overheads meant that it was increasingly important for universities to maximise their research income. Universities traditionally carry out the dual functions of teaching and research, which feed into and from the other. The capacity to earn resources for and from the research activities of their faculty was becoming a critical issue.

A Department of Education and Science (1991) White Paper set out changes in policy to ensure that teaching and research funding should be separately identified and that research funding should be “allocated selectively to encourage institutions to concentrate on their strengths”. The subsequent 1992 Further and Higher Education Act (Department of Education and Science, 1992) introduced the following major changes:

- The ‘binary line’ distinguishing universities from polytechnics and colleges was abolished, allowing the latter to be included in the RAE,
- The establishment of four UK country funding councils³
- Research funding should be allocated entirely on a selective basis,
- Quality audit would be the responsibility of individual institutions and quality assessment that of the funding councils,
- This assessment would inform funding decisions and would be based on the development and implementation of PIs.

Every succeeding RAE has seen extensive consultation and considerable changes, with consequences becoming progressively more severe. The first exercise determined the distribution

³ HEFCE (Higher Education Funding Council for England); SHEFC (Scottish Higher Education Funding Council); HEFCW (Higher Education Funding Council for Wales); DENI (Department of Education Northern Ireland).

of almost half the UGC's research funding (Kogan and Hanney). In 1992, no funds were provided for departments given the lowest of the five grades. In 1996, seven grading categories were used, but those ranked 1 or 2 received no funding at all. Following the 2001 RAE, funding was also withdrawn from those departments receiving the then lowest (funded) grade of 3b, see Table 1. (HEFCE Website; Database Resources Research Group Website)

TABLE 1 Changes in the RAE rating scales over time.

Within this climate of selectivity and assessment, academics have been increasingly concerned to examine the nature of their own work from a critical point of view. Willmott (1995) is one who sees the influence of the RAE over funding, as central to the commodification of academic labour. The potential impact of the RAE on the way academics work and think about themselves is well described by Parker and Jary (1995), who fear that the very identification of the new academic may come to be formalised in terms of what is needed to gain a high rating rather than in terms of independence and creativity of thought. Miller (1995) also argues that academics become constrained, monitored and documented via various PIs and that individual goals of scholarship and enquiry are displaced by economic considerations.

Five RAEs have been carried out and it now forms a regular event in the academic calendar. Since 1986, the results of this exercise have had a profound effect on university research funding and the relationship between quality and quantitative indicators of research performance has been a constant source of debate. Whilst this article is concerned with details of a specific type of publication (namely journal articles) submitted for research assessment, a summary of previous RAEs will be given in order to show the development of procedural and policy changes, especially regarding the assessment of publications as research output.

CHRONOLOGY

RAEs (originally called the Research Selectivity Exercise) were carried out firstly by the Universities Grants Committee (UGC) in 1986, followed in 1989 by the Universities Funding Council (UFC) and subsequently in 1992, 1996 and 2001 by the four country funding councils, including HEFCE. It now encompasses all higher education institutions wishing to receive funding from these councils. The first exercise was a relatively low key affair involving only 'traditional' universities. It resulted in only a small proportion of funding being apportioned as a result of the ratings. The second exercise had a larger proportion of funding dependent on the ratings of the subject panels. By 1992, virtually all university research funding would be determined by RAE rating. Methodologies developed as each subsequent exercise saw changes in the data it was thought appropriate to collect. There are approximately 70 subject categories defined and assessed by the funding councils, these individual university subject departments, (called Units of Assessment or UoAs) are ranked by specialist panels within (initially) a five-point scale.

1986-1989

Early attempts at research assessment did not go smoothly. The 1986 Research Selectivity Exercise consisted of the UGC asking British universities to complete a four part questionnaire covering various aspects of their research income and expenditure, research planning priorities and output. The responses received were considered by the UGC's subject sub-committees and rated against a variety of scales and standards (Phillimore, 1989). This exercise was probably the first attempt in any country to make a comprehensive assessment of the quality of university research. It is therefore not surprising that it was imperfect and was criticised (University Grants Committee, 1988) over the following issues:-

- the criteria for assessing research quality had not been made clear to universities,
 - the identity of assessors whose advice had been sought was withheld,
-

- evaluation of research on the basis of UGC cost centres/university departments had not allowed proper assessment of the work of interdisciplinary research groups and of joint departments,
- information sought from universities had biased judgements in favour of larger departments,
- the descriptive terminology in announcing the ratings was confusing; 'below average' had been understood to imply a low absolute standard,
- different assessment standards had been used for different subjects,
- being retrospective, the exercise had taken little or no account of work in progress and research potential,
- there was no appeals mechanism against individual ratings,
- there had been insufficient consultation with subject and professional groups.

Compilations of such criticisms were reported by Smyth and Anderson (1987) and Phillimore (1989). They concluded that the RAE lacked consistency, anonymity and comprised incomplete and inaccurate data collection. As early as 1987, a poll conducted for the *Times Higher Education Supplement* showed that the first RAE was opposed "by the majority of academics" (Times Higher Education Supplement, 1987). The debate became especially heated because PIs (such as those utilised in the RAE), whilst forming a potential element in resource allocation, also affected public reputations and perceptions (either of a department, university or individual).

A request for consultation towards the 1989 RAE received almost 300 responses and led to one of the main criteria being "in assessing publications and other publicly available output, the general approach will be that of informed peer review" (University Grants Council, 1988). A study by Gillette (1989) of various PIs also concluded that only those based on journal peer

review would constitute PIs capable of yielding a reasonably valid measure of departmental performance.

The second RAE in 1989 included a number of changes following the criticisms of the first. The most significant were:-

- extensive consultation with universities, funding agencies, subject associations and professional bodies prior to defining the exercise,
- 152 subject units of assessment, rather than cost centres,
- the UGC's sub-committees were reconstituted and expanded. Almost 70 peer review panels were instituted and panel membership was made public after the exercise,
- details of up to two publications per member of staff were requested, in addition to information on research students, external research income and future research plans,
- the information sought was greater than in 1986 and focussed on individual units rather than on university wide data,
- a 5 point rating scale, with international and national criterion referencing was used for all subjects.

Subsequently, the 1989 RAE was better received than in 1986, but criticism (Universities Funding Council, 1989) continued; it was centred on:-

- difficulties for universities in preparing their submission strategies, since the 5 point rating scale was not published until very late on in the process,
 - the full list of UoAs was not settled in advance,
 - some of the forms were unnecessarily complex and there was need for more precise definitions of publications,
 - no facility for systematic verification of the accuracy of the submissions was built into the exercise, and there was some evidence of deliberate 'misreporting',
-

- the exercise favoured large departments, particularly in the physical sciences,
- science ratings favoured excellence in basic and strategic research to the disadvantage of applied research,
- by assessing all staff in post for any part of the 5 year review period, the exercise was unduly retrospective.

The post-exercise report concluded that “publication data was found to be unreliable, and where it was reliable, it said nothing about the quality of the output”(Universities Funding Council, 1989, para 23). These early RAEs demonstrated that methodologies for assessing research output were the main areas of difficulty for both academics and assessors alike. Problems centred on what to collect and what measurements to use. Much discussion took place on the relative merits of bibliographic profiles, various measurements of publications and forms of citation analysis. As early as 1989, an unpublished bibliometric study by SEPSU⁴ concluded that there was a need for “comprehensive publication lists to be obtained directly from university departments because at least half of the research output from many science departments was missed by the Science Citation Index” (Anderson, 1989).

1990-1992

Up to this time, university departments were government funded by formulae in proportion to student numbers, to undertake both teaching and research. Departments in the ‘new’ universities (former polytechnics) were government funded (also in proportion to student numbers) to carry out teaching – with no formula funding for research. Staff in the latter, who undertook research, did so in time free from teaching and, if funded at all, such research was largely through industrial sponsorship or consultancy. The Further and Higher Education Acts 1992 bought

⁴ Science and Engineering Policy Studies Unit

about the official ending of this divide between universities and polytechnics, by enabling the latter to apply for university status, and also thus compete for funding via the 1992 RAE.

The Universities Funding Council (UFC) had intended to carry out the next exercise in 1993, but the creation of the four funding councils meant that it was brought forward to 1992, so that the results could help determine the grants for 1993-4. To compensate for this change in timescale, work accepted for publication could be included. The following other major changes were made for the 1992 RAE :-

- Higher Education Institutions (HEIs) could select which 'research active' members of staff to put forward,
- research was graded on a five-point scale,
- the amount given depended on the quality rating, the number of staff declared 'research active', the amount of research income and an element on future research activity,
- the exercise was made less retrospective by seeking detailed information relating to staff in post on 30th June 1992 – the 'snapshot' approach,
- separate assessments were provided for in applied and basic/strategic research in science and engineering,
- in recognition of the longer timescale for research in the arts and humanities the assessment period for these UoAs was extended by one year to four and a half years,
- assessment was divided into 72 academic UoAs,
- in addition to each academic nominating two publications, quantitative information on **all** publications was required under a range of headings,
- a formal audit process was introduced to verify the accuracy of the submissions.

By the closing date, over 2,800 submissions had been received from 192 HEIs, covering the research of over 43,000 full-time equivalent researchers in post at that date. Following this

exercise, older universities received 91% of available research funding, new universities 7% and colleges 2% (Higher Education Funding Council for England, 1994). HEFCE produced a consultative paper in 1993 inviting comments on how future RAEs should be conducted. It included a report on the 1992 exercise, explaining that a consensus had been reached that future exercises should again be based on “informed peer-review” and invited wide consultation (Higher Education Funding Council for England, 1993). Although extensive amounts of quantitative publication output data had been collected for the 1992 RAE, HEFCE’s report explained that “limited use” was made of the lists of publications for various reasons, including “that it cannot be stressed too much that the exercise was conducted with quality rather than quantity as the main criteria.” (Higher Education Funding Council for England, 1993, para 23).

Doyle and Arthur’s (1996) statistical work on the 1992 RAE led to a fairly damning summary of the exercises up to 1996, and specifically that (for UoA 43) “in its implicit policies the panel has fallen prey to just about all the different variants of the home team bias. English universities are favoured, old established universities are favoured, panelist’s own universities are favoured and British journals are favoured, although marginally so. Large institutions are also favoured”. Such judgements led to major changes before the 1996 RAE in terms of the need for outside assessors (outside the university sector and outside of the UK) and for Panels to consider rankings in the absence of affiliated/interested panel members.

1996 RAE

By 1996, the RAE had moved away from quantity measures, i.e., collecting full lists of publications. After extensive comments on the 1992 format, and some consequent revisions, procedures for the 1996 RAE were announced in June 1994 (RAE 96 1/94). The required framework was that a list of ‘research active’ staff in a department or unit was to be put forward,

with details of up to four publications per selected member of staff, but NOT the total number of publications for departmental staff (as in 1992).

Outputs must have been published, or otherwise placed in the public domain by the census date (31/3/1996) “HEFCE wish to signal clearly that the RAE is concerned with research quality, and that numbers of publications and other forms of assessable output is not considered necessarily to be an indicator of research quality” (RAE 96 1/94, Annex C 24). Some months later, HEFCE qualified these straightforward but restrictive reporting requirements to include “indications of peer esteem” in the form of editorships of prestigious journals or papers in key conferences (RAE 96 3/95).⁵

Seven possible rating outcomes from 1-5* were established for 1996. The former band 3 was subdivided into 3a and 3b, and a highest grade of 5* was introduced, see Table 1. Furthermore, following a legal case bought by the Institute of Dental Surgery over the grade it received in 1992, each panel had to produce a statement of criteria that it would use to assess research for each UoA.

However, a definition of ‘research’⁶ (RAE 96 3/95) which could not be challenged, was provided by the funding councils. The task of the panels was thus to interpret this definition or ‘fine-tune’ it to the specific requirement of the discipline (Broadhead and Howard, 1998).

⁵ In addition, the number of research students and studentships, amounts of external research income, and statements of research strategy and future plans had also to be submitted. Departments choosing not to participate would be disqualified from receiving governmental research funding.

⁶ “Research for the purpose of the RAE is to be understood as original investigation undertaken in order to gain knowledge and understanding. It includes work of direct relevance to the needs of commerce and industry, as well as to the public and voluntary sectors; scholarship, the invention and generation of ideas, images, performances and artefacts including design, where these lead to new or substantially improved insights; and the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction. It excludes routine testing and analysis of materials, components and processes, as distinct from the development of new analytical techniques.”

If a member of faculty moved institutions during the four year assessment period, credit for the work would go to the institution they were affiliated with on the census date. Problems associated with what became known as the 'transfer market' for top researchers led to a rethink on this for 2001. Other problems arose over who to submit. A department could choose to withhold an academic's work from the department's submission, i.e., designate them 'non research active'. The eventual funding awarded would be based on the overall quality grade multiplied by the number of research active staff in the department (RAE 96 1/94, Annex C 17). This, too, resulted in some manipulation and gamesmanship and led to changes for 2001.

The 1996 exercise endeavoured to judge research culture rather than merely measure research activity. Assessment was by informed peer review, with 60 subject panels considering submissions for assessment in 69 UOAs. Panel Chairs were appointed by the funding bodies, on the advice of the previous Panel Chair. A limit of 33% was set on panel personnel continuity (Broadhead and Howard, 1998) and the remaining Panel Members were selected following advertisements for nominations from over 1300 learned societies, subject associations and other interested bodies. Even given this substantial effort to broaden the range of Panel Members, there was still criticism. Roberts (1999) found statistical evidence that the outcome of the 1996 RAE was biased towards departments with members on assessment panels and Piercy (2000) noted that "the game is that we are judged primarily by other academics, on the basis of publications read only by other academics and research grants awarded by academics to academics". Even the House of Commons Science & Technology Select Committee questioned whether the Panels operated in a 'clubby atmosphere' (Science and Technology Select Committee, 2002).

No funding was given to departments graded 1 or 2. The amount provided to departments in each discipline was designed to reflect both the volume and quality of a department's work and also the relative cost of research in the field (Higher Education Funding Council for England, 1997).

2001 RAE

Extensive consultation again took place following the 1996 exercise and as a result, further changes were made. New measures built in to the 2001 exercise were :-

- Processes were put in place to help panels achieve consistency between UoAs, especially for interdisciplinary work. As the benchmark for the RAE is international excellence in each subject, every panel consulted a number – usually five – non-UK based experts in the field. These advisors reviewed submissions already provisionally awarded top grades.
 - Sub-panels and Umbrella Groups were set up. Sub-panels looked mainly at thematic areas specifically in Medicine. Panel Chairs met in their Umbrella Groups and discussed approaches taken during grading to ensure that any differences in the grade profiles for their Panels reflected genuine differences in overall quality, rather than variable approaches to marking. Groupings were:
 - Umbrella Group I - Medical and Biological Sciences
 - Umbrella Group II - Physical Sciences and Engineering
 - Umbrella Group III - Social Sciences
 - Umbrella Group IV - Area Studies and Languages
 - Umbrella Group V - Humanities and Arts
 - More recognition was taken of staff circumstances, i.e., stage of research career, career breaks, long-term projects and secondments. In addition, a new category of staff was introduced – Category A* staff. This was designed to address problems caused by the so-
-

called academic transfer market, where one institution could poach staff from another institution close to the RAE census date. Previously the 'losing' institution would forfeit the right to use the researcher's publications as part of its own submission. In the BMS UoA 43, for example, 11% of staff fell into the A* category (RAE Panel Overview Report BMS 2002). In such cases, up to two outputs could be submitted by both the old and the new departments, compared with a maximum of four publications for other Category A (and C) staff. This allowed A* staff to 'count' for both the losing and gaining institution. However, the institution employing them at the census date received all resulting funding!

- More feedback would be given on the ratings achieved via individual reports to institutional heads and UoA Panel overview reports were published on the Internet. In addition all submission details (except those deemed commercially sensitive) would be available on the Internet, after July 2002.
 - E-publications were stated to be acceptable submissions in the same way as equivalent peer-reviewed print publications. Panels stated that they would judge all publications and other research outputs on their merits, **not** on their medium of publication. Fulford (2000) (Chair of Review Panel for Archaeology 2001) wrote: "Colleagues are urged to exploit the exciting possibilities of electronic publication. They can be assured that the RAE panel will be concerned with evaluating the content of publications. The medium of publication, whether electronic, or in traditional, printed form will be of no account". This produced a mixed response post-assessment, i.e., "e publishing has increased and some interesting journals have been established since 1996; work published through this medium was assessed on equal terms with that in hard copy (although some proved difficult to access)" (RAE Panel Overview Report English, 2002); "Other formats of research output, for example, electronic publications, were little in evidence" (RAE Panel Overview Report History of Art, 2002).
-

Official RAE working documents and assessment criteria were published in advance of the exercise, which stressed the even-handedness with respect to the medium of output, whereby the Panels' only concern would be "assessing the quality of the research submitted regardless of its type, form or place of output" (Higher Education Funding Council for England, 1999). However, after the event, several Panels pointed out that no hierarchy or ranking of types of output or journals in the field would be used (Psychology UoA 13, Computer Science UoA 25, Law UoA 36) and individual Panel approaches did vary as to how the quality of outputs were judged :-

Pre-clinical studies, Anatomy, Physiology and Pharmacology (UoA 5, 6, 7, 8)

Weight will be given to work that has undergone a vigorous peer review, editorial or assessment process. Where this is not evident the work will be assumed to be of lower quality.

Statistics UoA 24

The Panel will take the perceived editorial standards of journals into account when examining outputs.

Economics and Econometrics UoA 38

Whilst importance will be attached to the peer-refereeing process of publications, the Panel recognises that some research of high quality will be found in less prominent journals.

Philosophy UoA 62

The Panel will take into account editorial or refereeing processes on cited works but outputs not subject to these will not be automatically regarded as of lesser quality.

Sport-related subjects UoA 69

Outputs that have been through a rigorous editorial, refereeing or reviewing process, will normally be regarded as at least equal to national excellence.

(RAE Panel Overview Reports, 2002)

Rating distribution for the last three RAEs is shown below in Table 2. Following the 2001 exercise, no funding was awarded to departments receiving 1, 2 or 3b grades (17% of submissions). This Table also shows that in 1992, 67% of departments were ranked either 1, 2 or 3. By 1996, the number of 4, 5 and 5* departments had risen to 43%, and this rose again to 65% in 2001 (Higher Education Funding Council Website; House of Commons Science and Technology Committee, 2002)

Table 2 Comparison of RAE ratings 1992, 1996, 2001 in here

This Table also shows that the 23% of staff working in 5 rated departments in 1992 (Bassnett, 2002), rose to almost 31% in 1996 (in 5 and 5*) and rose again in 2001 to almost 55%. On the face of it, this shows a steady increase in the quality of UK research activity. However, it could also mean that institutions have learned who, how and what to submit to the best effect. It could also simply be due to 'grade inflation' with Panels becoming more generous in their assessments. Reactions to the 2001 results ranged from "the results of the 2001 RAE show progress towards national standards of excellence throughout the HE research community....65% of submitted research was of national or international excellence, compared to 43% in 1996." (HERO Website) to "what that 55% represents (researchers working in 5 or 5* rated departments in 2001) is a morass of fiddling, finagling and horse trading. Nobody who works in a university in the UK in 2002 seriously believes that research is improving" (Bassnett, 2002).

CONCLUSIONS

The RAE continues to engender strong feelings amongst those who administer it, are judged by it and those affected by the results. Henkel (1999) concludes that it has "created a profound disturbance within the academic profession" by removing what had been largely implicit

processes of peer judgement and decision-making and placing them within a comprehensive standardised national framework of assessment, falsely concentrated into a highly public event, linked by formulae to the allocation of funding every four or five years.

Fulton (1997) stated that it had “been a vehicle for professional and personal humiliations” in that it revealed for all to see that a large proportion of staff in major universities were failing (or were unable) to meet these nationally set standards. In its extreme, it could lead to enforced ‘early retirements’. For example, Queen’s University Belfast after the 1996 RAE, both enforced such retirements and introduced new ‘research active appointments’ as part of a strategic plan to improve its ranking in the 2001 RAE⁷. Vice-Chancellor Professor Andrew stated “people who are not performing significantly in research cost the university money” (Foley, 1998).

Individual disciplines also have their RAE critics. Within Medicine, Williams (1998) calls it “misleading, unscientific and unjust” and Piercy (2000) from management explores “Why it is fundamentally stupid for a business school to try to improve its RAE score”.

Even given criticisms of the process, problems and inconsistencies surrounding this form of assessment, “no better device than this objectified subjectivity has been elevated as a candidate for the function required in this terrain of research selectivity”(Velody, 1999). The House of Commons Science and Technology Committee (2002, Conclusion, para 5) concluded that “ the RAE has undoubtedly brought benefits but it has also caused collateral damage. It has damaged staff careers and it has distracted universities from their teaching, community and economic development roles. Higher education should encourage excellence in all these areas, not just in research. Universities should be assessed on a balanced scorecard”.

⁷ QUB recorded much improved results in 2001. The number of departments with international research quality (graded 4 or above) increased from 20 in 1996 to 33 in 2001 and the percentage of academic staff in units graded 4 or above increased from 50% to 89%. 5* departments rose from 7 in 1996 to 16 in 2001. (Queen’s University Belfast Website)

In recent months, a high-level Committee has been looking at possible future methods of assessing research output from UK Higher Education. Whilst the intention was that nothing would be ruled in or ruled out, it is reasonable to conclude that the recommendations will continue to lay great emphasis on the quality of research output. As long as publications continue to be used as a primary method of research output, and as long as many of them involve refereeing processes, they will continue to be judged in any future RAE. Despite the efforts of Harnad and his collaborators (2003) to institute drastic changes to the RAE process, for the time being it is unlikely that much can be done to break the link between perception and reward.

REFERENCES

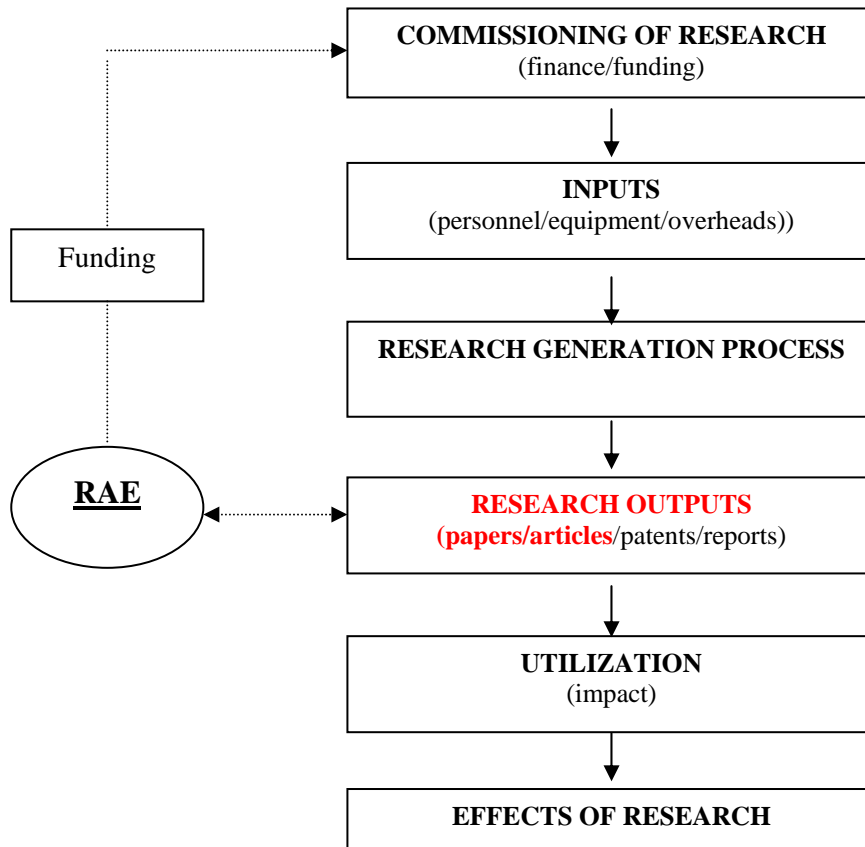
- Anderson, J. (1989), *New approaches to evaluation in UK research funding agencies*, (London, Science Policy Support Group).
- Baird, L.M. & C. Oppenheim.(1994), Do citations matter? *Journal of Information Science*, 20, pp. 2-15.
- Barma, S., Tersine R. & M. Buckley (1991), An empirical assessment of the perceived relevance and quality of POM related journals by academicians. *Journal of Operations Management*. April, pp.194-212.
- Bassnett, S. (2002), Fruitless Exercise, *The Guardian* 15/1 2002.
- Brinn, T. & J. Jones. (1996) UK accountants' perceptions of journal quality. *Accounting and Business Research*, 26(3), pp. 265-278.
- Broadhead, L., & S. Howard. (1998), The art of punishing: The RAE and the ritualisation of power in Higher Education. *Education Policy Analysis Archives*. 6(8).
- Campbell, K. et al.(1999), Journal publishing, journal reputation and the UK's Research Assessment Exercise. *Journal of Law and Society*. 26(4) pp.470-501.
- Cave, M., Hanney, S., and Kogan, M. (1988), *The use of performance indicators in HE: A critical analysis of developing practice*. (London, Jessica Kingsley) pp. 18.
- Cave, M., Hanney, S. & Henkel, M. (1997), *The use of performance indicators in HE*. Higher Education Policy Series 3, (London, Jessica Kingsley).
- Committee of Vice-Chancellors and Principals/University Grants Committee, (1986), *Performance Indicators in Universities: A first statement by a joint working group*, (London, CVCP.)
- Crewe, I. (1987), *Reputation, research and reality: the publication records of UK departments of Politics 1978-1984*. Essex papers in Politics and Government No 44 Dept of Government, University of Essex.
- Cuenin, S. (1986) International study of the development of performance indicators in higher education. *Paper given to OECD/IMHE Project, Special Topic Workshop*.
- Database Resources Research Group website*
URL <http://www.soi.city.ac.uk/informatics/is/drrg/rae97/rae97.html> [visited July 2002]
- Department of Education and Science, (1991), *Higher Education: A New Framework*, Cm 1541, (London, HMSO) para.38.
-

- Department of Education and Science, (1992), *Further and Higher Education Act and Further Education (Scotland) Act*. (London: HMSO).
- Doyle, J., & A. Arthurs. (1996), The judge, the model of the judge and the model of the judged as judge: analysis of the UK 1992 RAE data for business and management studies. *Omega: International Journal of Management Science*. 24(1) pp. 21.
- Extejt, M., & J.E. Smith (1990), The behavioural sciences and management: An evaluation of the relevant journals, *Journal of Management*, 16, pp. 539-551.
- Foley, C. (1998), QUB asks 107 staff members to leave. *Irish Times*, 31/8/1998, pp.12.
- Fry, E., Walters G. & L. Scheuermann.(1985), The perceived quality of 50 selected journals. *Journal of the Academy of Marketing Science*, 13(1), pp.352-361.
- Fulford, Mike.(2000), *Email to 'Internet Archaeology'. 06 January 2000*
- Fulton, O. (1997), *Research assessment: present and future imperfect?* UWE and SRHE Conference Report. (Bristol, University of the West of England).
- Gillette, R. (1986), *Serious anomalies in the UGC comparative evaluation of performance of Psychology departments*. Department of Psychology, University of Leicester, pp. 4.
- Gillette, R. (1989), Research performance indicators based on peer review: A critical analysis, *Higher Education Quarterly*, 43, pp. 20-38.
- Harnad, S., Carr, L., Brody, T. and Oppenheim, C. (2003) Improving the UK Research Assessment Exercise whilst making it cheaper and easier, <http://www.ecs.soton.ac.uk/~harnad/Temp/Ariadne-RAE.doc>
- Hemlin, S. (1996), Research on research evaluation, *Social Epistemology*, 10 (2) pp.209-250.
- Henkel, M. (1999), The modernisation of research evaluation: The case of the UK. *Higher Education* 38, pp. 105-122.
- HERO website URL <http://www.hero.ac.uk/rae>
- Higher Education Funding Council for England website URL <http://www.hefce.ac.uk> [visited July 2002]
- Higher Education Funding Council for England, (1993), *Research Assessment: A Consultation Paper 6/93*, (Bristol, HEFCE)
- Higher Education Funding Council for England, (1994), *An Overview of recent developments in HE in the UK*, (Bristol, HEFCE).
- Higher Education Funding Council for England, (1997), *Circular 4/97* http://www.niss.ac.uk/education/hefce/pub97/c4_97.html.
- Higher Education Funding Council for England, (1999), RAE Circular 5/99 Assessment panels' criteria and working methods. Bristol, para 1.5.
- Higher Education Funding Council for England, (1999), *RAE Circular 5/99 Assessment Panels criteria and working methods*. URL www.hero.ac.uk/rae/Pubs/5_99 para. 1.4g.
- House of Commons Science and Technology Committee Report on the Research Assessment Exercise, (2002) *Results of the RAE 2001*, URL: http://www.parliament.the_stationary_office.co.uk/pa/cm200102/cmselect/cmsctech/507
- House of Commons Science and Technology Committee Report on the Research Assessment Exercise, (2002) *Second Report, Conclusion para 5*. URL: http://www.parliament.the_stationary_office.co.uk/pa/cm200102/cmselect/cmsctech/507
- Johnes, J., & J. Taylor.(1990), *Performance Indicators in Higher Education*, (Society for Research in Higher Education, Buckingham, SRHE and Open University Press) pp. 183.
- Jones, L. V., et al. (1982), *An assessment of research Doctorate programs in the USA*, National Academic Press, Washington DC) 5 vols.
- Judge Institute of Management Studies, University of Cambridge, (1994), *Quality rankings of journals in Management and Social Science*. (Cambridge, Research Papers in Management Studies).
- Kogan, M., & S. Hanney, *Reforming Higher Education. HE Research Policy Series No 50*, (London, Jessica Kingsley) pp. 98.
- Lee, F., & S. Harley.(1998), *Economics divided: The limitations of peer-review in a paradigm bound social science*. In Jary, D. & M. Packer.(Eds) (1998) *The New Higher Education*, (Staffordshire University Press).
- Miller, H. (1995), *The management of change in universities*, (Buckingham, Open University Press).
- Parker, M., & D. Jary, (1995), *The McUniversity: Organisations, management and academic subjectivity. Organization* 195 (2), pp. 319-338.
- Phillimore, A. (1989), University research performance indicators in practice: The UGC evaluation of British Universities 1985-1986, *Research Policy*, 18, pp. 255-271.

Piercy, N. (2000), Why it is fundamentally stupid for a business school to try to improve its RAE score. *European Journal of Marketing*, 34 (1/2) pp. 27-35.

- Queen's University, Belfast website <http://www.qub.ac.uk/info/rae/ataglance.htm> visited February 2003.
- RAE 96 1/94 http://www.niss.ac.uk/education/hefc/rae96/cl_94.html
- RAE 96 1/94 http://www.niss.ac.uk/education/hefc/rae96/cl_94.html Annex C 24
- RAE 96 1/94 http://www.niss.ac.uk/education/hefc/rae96/cl_94.html Annex C 17
- RAE 96 3/95 *Criteria for Assessment – working methods of panels*
- RAE Panel Overview Report (2002), *BMS* <http://www.hero.ac.uk/rae/overview/>
- RAE Panel Overview Report (2002), *History of Art, Architecture and Design* <http://www.hero.ac.uk/rae/overview/>
- RAE Panel Overview report, (2002), *English Language and English Literature* <http://www.hero.ac.uk/rae/overview/>
- RAE Panel Overview Reports, (2002), <http://www.hero.ac.uk/rae/overview/>
- Roberts, Chris. (1999), *Possible bias due to panel membership in the 1996 RAE*, Paper give at the Royal Statistical Society International Conference, Warwick University July 12th-15th 1999.
- Rogers, A. & T. Scratcherd. (1986), *Letter to THES* 7/11/86.
- Science and Technology Select Committee, *Minutes of Evidence, 23rd January 2002*, (London, HMSO). para 17.
- Sizer, M. (1979), Assessing institutional performance: An overview. *International Journal of Institutional Management in Higher Education*, 3 (1) pp.49-77.
- Smyth, F., & J. Anderson. (1987), *University Performance Indicators*, Science Policy Support Group Seminar, Aide-Memoire No 1, (London, SPSG).
- Times Higher Education Supplement*, Research Assessment, 5th June 1987, xvi. (no author given)
- Universities Funding Council. (1989), *Report on the 1989 RAE*. (London, UFC).
- University Grants Committee (1987), *University Management Statistics and PIs*. (London, CVCP/UGC).
- University Grants Committee, (1988), *Research Selectivity Exercise 1986. Circular letter 45/88*. (London, UGC).
- University Grants Council, (1988), *Research Selectivity Exercise 1989, Circular letter 45/88*, (London, UGC), para. 9.
- Velody, I. (1999), Knowledge for What?- The Intellectual consequences of the RAE, *History of the Human Sciences*, 12 (4) pp.112.
- Wells, A. (1998/99), *Exploring the development of the independent, electronic, scholarly journal. Unpublished MSc Thesis*, (University of Sheffield, Department. of Information Studies)
- Williams, G. (1998), Misleading, unscientific and unjust: the UK's RAE. *British Medical Journal*, 316, pp.1079-1082.
- Willmott, H. (1995), Managing the academics: Commodification and control in the development of university education in the UK. *Human Relations*, 48, pp. 993-1025.
-

Figure 1 The Research Production Model



Source: adapted from (6)

Table 1 Changes in the RAE rating scales over time.

1992 Rating Scale and description		1996 Rating Scale and description		2001 Rating Scale and description	
		5*	Research quality that equates to attainable levels of international excellence in a majority of sub-areas of activity and attainable levels of national excellence in all others.	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.
5	Research quality that equates to attainable levels of international excellence in some sub-areas of activity and to attainable levels of national excellence in virtually all others.	5	Research quality that equates to attainable levels of international excellence in some sub-areas of activity and to attainable levels of national excellence in virtually all others.	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.
4	Research quality that equates to attainable levels of national excellence in virtually all sub-areas of activity, possibly showing some evidence of international excellence, or to international level in some and at least national level in a majority.	4	Research quality that equates to attainable levels of national excellence in virtually all sub-areas of activity, possibly showing some evidence of international excellence, or to international level in some and at least national level in a majority.	4	Quality that equates to attainable levels of national excellence in virtually all of the research activity submitted, showing some evidence of international excellence.
		3a	Research quality that equates to attainable levels of national excellence in virtually all sub-areas of activity, or to international level in some and to national level in others together comprising a majority.	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.
3	Research quality that equates to attainable levels of national excellence in a majority of the sub-areas of activity, or to international level in some.	3b	Research quality that equates to attainable levels of national excellence in the majority of sub-areas of activity.	3b	Quality that equates to attainable levels of national excellence in more than half of the research activity submitted.
2	Research quality that equates to attainable levels of national excellence in up to half the sub-areas of activity.	2	Research quality that equates to attainable levels of national excellence in up to half the sub-areas of activity.	2	Quality that equates to attainable levels of national excellence in up to half of the research activity submitted.
1	Research quality that equates to attainable levels of national excellence in none, or virtually none, of the sub-areas of activity.	1	Research quality that equates to attainable levels of national excellence in none, or virtually none, of the sub-areas of activity.	1	Quality that equates to attainable levels of national excellence in none, or virtually none, of the research activity submitted.

Red text = no funding awarded for these graded institutions

Source: (34) (35)

Table 2 Comparison of RAE ratings 1992, 1996, 2001

Ratings 1992	Submissions 1992	Ratings 1996	Submissions 1996	% staff 1996	Ratings 2001	Submissions 2001	% staff 2001
1	423 (15%)	1	236 (8%)	3.4	1	18 (1%)	0.2
2	613 (22%)	2	464 (16%)	9	2	140 (5%)	2.4
3	837 (30%)	3b	422 (15%)	10.9	3b	278 (11%)	5.5
		3a	528 (18%)	18.4	3a	499 (19%)	12.4
4	560 (20%)	4	671 (23%)	27.6	4	664 (26%)	24.8
5	350 (13%)	5	403 (14%)	20	5	715 (28%)	36.0
		5*	170 (6%)	10.8	5*	284 (11%)	18.7
Total	2783	Total	2894 (100%)		Total	2598 (100%)	

Source: Adapted from (69) (70)

Word count: 6150 + 80 Abstract + 630 in Tables etc.

No Appendices
