



OPINION PIECE

Measuring research quality: the United Kingdom Government's Research Assessment Exercise

Ray J. Paul¹

¹*Department of Information Systems and Computing, Brunel University, Uxbridge UB8 3 PH, U.K.*

Correspondence: Ray J. Paul, Department of Information Systems and Computing, Brunel University, Uxbridge UB8 3 PH, U.K.
Tel: +44 1895 266027;
E-mail: Ray.Paul@brunel.ac.uk

Abstract

Can research quality be measured meaningfully? Whether it can or not does not interfere with the desire to do so if the motivation is strong enough. This paper discusses the United Kingdom (U.K.) Government's approach to measuring research quality in U.K. universities, known in the U.K. as the Research Assessment Exercise (RAE). The RAE is held at irregular intervals of time, the latest being held in 2008 (RAE 2008). The motivation is accountability of government spending, but in fact many different objectives are ascribed to the exercise. RAE 2008 is fully described so that it can be used to raise debating issues around the general issue of research quality, a subject that appears to be of universal interest in Information Systems. One major conclusion appears to be that journal rankings are not a good indicator of the quality of any paper published in that journal, nor necessarily of the combined quality of all the papers.

European Journal of Information Systems (2008) 0, 000–000.
doi:10.1057/ejis.2008.31

Measuring research quality

The U. K. Government spends a lot of money on research, particularly in Higher Education. Some of the latter money is channelled through Government agencies, who tend to award grants to successful research proposals, and for which the money allocated is spent according to the research proposal plan (researchers' salaries, equipment, etc.). So here the measure of research quality is competitive bidding for grants, judged through a peer review process and against predetermined funding budgets for subject or research topic areas.

A second level of Government funding for research is through the direct award to a university of monies that are related to the volume and quality that each university has historically attained. In this financial year 2007/08 the money given to the English universities alone was £1.415 billion for the year, a considerable amount of funding (U.S. \$2.8 billion). Analysis of the funding allocations shows the usual Pareto type distribution of funding, with the four universities with the highest income (Cambridge, Oxford, University College London and Imperial College) receiving nearly 30% of the total (Cambridge received £107 million), the top 10 receiving over 50% of the total, the top 20 receiving over 70% and the bottom 80 out of 130 or so institutions receiving less than 6%. This money is allocated without conditions on how it should be spent and is the only 'free' money that universities get to spend how they like. So not only is it a lot of money,

but its value is greater still because of the flexibility and choice it gives the beneficiary.

So how is this money allocated to the universities? The official answer is on the basis of historical research quality. What this means, in practice, is discussed in the rest of this paper starting with the official declaration of intent in the next section, and ending with some debating issues and some surprising conclusions.

Although I am a member of the peer review group carrying out this assessment for the U.K. Government this year, I want to make it quite clear what my reservations about measuring research quality are. I will start with one of my favourite quotes,

Recall Chou En-lai's response, when asked to assess the results of the French Revolution: it is too soon to tell. (<http://www.law.nyu.edu/eecr/vol11num3/special/krygier.html>)

Apart from the issue of how long it takes for a research result to have an impact, there are issues to do with defining quality at all. What would a world leading edge piece of research consist of? Is it Einsteinium in its impact? If so, there is not a lot of that about. Does it change research directions, or is it heavily referred to by all and sundry (but a good review paper will achieve this), or what? We shall see what the RAE is making of such definitions below, and then later what impact this has on the assessment of quality. Lord Kelvin is often quoted in this context, but usually incorrectly interpreted in my view. One version of the quote is:

In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of *Science*, whatever the matter may be. [PLA, vol. 1, 'Electrical Units of Measurement', 1883-05-03] (<http://zapatopi.net/kelvin/quotes/>)

This quote is often I believe misunderstood. The attachment of numbers to what is being measured is easily undertaken without any questioning of whether this can be done or not. Kelvin does not say that using numbers means you know something, but 'when you can measure' then ... Note that Kelvin's quote also covers those situations 'when you cannot measure it'. In other words, there are cases when you cannot measure. Even if you attach numbers that have relative ordering as meaning (for example, league tables), this is not equivalent to having absolute values. And if you cannot even do relative numbers, you have no meaning. And then the quote goes on to imply that you then have little knowledge.

Some seem to believe that if a journal is ranked highly, has good citation or impact measures of performance, is difficult and time consuming to get into, etc. that this shows something about the quality of the papers published. Maybe it does, but does it show research quality? At the end of the paper I will explain why this is unlikely to be the case.

Measuring research quality reminds me of what the populist philosopher Bryan Magee says in his autobiography (Magee, 1997) about his first encounter with a simple every day activity. At the age of 5 he realised that every night he fell asleep – but when exactly? He stayed awake to find out; he asked his big sister when that failed; he asked his parents; and he realised that some questions do not appear to have answers. To determine when you fall asleep, you have to define what you mean by 'falling asleep' and any such definition is then a tautology. Similarly for research quality – any measurement requires a definition that will then lead to measurement that meets the definition – but this is not adequate for a universal measure of research quality which needs to be independent of the definition if it to be meaningful. For example if you wish to measure the quality of wine, and you make the observation that in general you have found French wine to be the best, it should be no surprise to you that among the top ranked wines there are a disproportionate number of French wines.

Nevertheless, the RAE attempts to measure research quality in the manner now to be described.

The research assessment exercise – RAE 2008: intent

This section sets out the intention of the RAE using the publicly available document RAE 01/2004. Most of the words in this section are taken from this document with occasional rewording to précis the document into the relevant story line of this paper. The RAE authorities' state:

'We undertake periodic assessments of the quality of research carried out in higher education institutions (HEIs) in all disciplines and across the U.K.' The purpose is to inform our allocations of grant for research, and to support our shared policy of promoting continuous improvement in the quality of the U.K. research base and its economic and social impact.

The RAE will continue to be a discipline-based expert review process in which judgements on the quality of research are made by researchers and experts active in that discipline. Decisions on ratings will be made by some 15–20 main panels, based upon detailed assessment work by around 70 sub-panels. The outcomes will be published as quality profiles for around 70 units of assessment (see Table 1 below). The main panels and sub-panels will work closely together.

We have decided to have four starred levels of excellence. In developing descriptors, we shall keep in mind the definitions for the top two starred levels proposed by the review group (roughly, the upper and

Table 1 Sample quality profile^a

Unit of assessment A FTE staff submitted for assessment		Percentage of research activity in the submission judged to meet the standard for				
		Four star	Three star	Two star	One star	Unclassified
University X	50	15	25	40	15	5
University Y	20	0	5	40	45	10

^aThe figures are for fictional universities. They do not indicate expected proportions.

lower half of the body of work that would have been defined as reaching 'international' levels of excellence in the 2001 RAE). The lower two starred levels will fit in below this and will probably cover between them the work that would have been classified as reaching 'national' levels of excellence in 2001.

RAE 2008: the approach

Having declared the intent for RAE 2008, later documents set out how this is to be achieved. In this section the approach taken uses the words in the publicly available document RAE 03/2005. Most of the words in this section are taken from this document with occasional rewording to précis the document into the relevant story line of this paper. The RAE authorities' state:

1. The RAE is conducted jointly by the Higher Education Council for England (HEFCE), the Scottish Higher Education Funding Council (SHEFC), the Higher Education Funding Council for Wales (HEFCW) and the Department for Employment and Learning, Northern Ireland (DEL). The RAE is managed by the RAE team, based at HEFCE, on behalf of the four U.K. higher education funding bodies. In this document, 'we' refers to the RAE team.
2. The primary purpose of RAE 2008, the fourth such national exercise since 1992, is to produce quality profiles (see Table 1 above) for each submission of research activity made by institutions. The four higher education funding bodies intend to use the quality profiles to determine their grant for research to the institutions which they fund with effect from 2009–10. Any HEI (Higher Education Institution) in the U.K. that is eligible to receive research funding from one of these bodies is eligible to participate.
3. The definition of research for RAE 2008 is below. It is virtually the same as that used in RAE 2001: we have made minor amendments to the phrasing of the definition used in 2001 that clarify but do not change its meaning.

Definition of research for the RAE

(Changes in phrasing from the definition used for the 2001 RAE are in *italics*.)

'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, industry, *and* to the

public and voluntary sectors; scholarship;¹ the invention and generation of ideas, images, performances, 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 *routine* analysis of materials, components and processes *such as* for the maintenance of national standards, as distinct from the development of new analytical techniques. It also excludes the development of teaching materials that do not embody original research.

4. Each submission will contain the core data detailed in paragraphs a–h following.

- (a) Overall staff summary: summary information on research active staff selected (FTE and headcount) and related academic support staff (FTE) in the unit of assessment. NB FTE means Full-Time-Equivalent.
- (b) Research active individuals: detailed information on individuals selected by the institution for inclusion as research active.
- (c) Research output: up to four items of research output produced during the publication period (1 January 2001 to 31 December 2007) by each individual named as research active and in post on the census date (31 October 2007).
- (d) Research students: numbers of full-time and part-time postgraduate research students and degrees awarded.
- (e) Research studentships: numbers of postgraduate research studentships and source of funding.
- (f) External research income: amounts and sources of external funding.
- (g) Textual description: including information about the research environment and indicators of esteem.
- (h) Individual staff circumstances.

RAE 2008: an example of panel criteria and working methods

I have chosen Panel I whose sub-panels are

Economics and Econometrics
Accounting and Finance

¹Scholarship for the RAE is defined as the creation, development and maintenance of the intellectual infrastructure of subjects and disciplines, in forms such as dictionaries, scholarly editions, catalogues and contributions to major research databases.

Business and Management Studies
Library and Information Management

I am a member of the last two sub-panels, providing the link between them which is largely in the discipline of information systems. Quite rightly I am under confidentiality constraints so in this section I am only going to highlight information publicly available to all and sundry. The particular document this section is based on is RAE 01/2006 (I) which describes the panel criteria and working methods of Panel I and its four sub-panels listed above. Most of the words in this section are taken from this document with occasional rewording to précis the document into the relevant story line of this paper. The document discusses how consistency of quality levels will be determined in the paragraphs numbered as per below:

11. The main panel and its sub-panels regard the cited outputs as the most important indicator of research quality, and this is reflected in the weightings allocated to this element of the quality profile. Research outputs will therefore carry a weighting of 70% towards the overall quality profiles.

12. Indicators of the research environment provide important evidence of the infrastructure supporting high-quality research, and departments' contribution to the development of their fields. These indicators will collectively carry a weighting of 20% towards the quality profile. Within this figure, the sub-panels will recommend a quality profile for research environment based upon their assessment of the evidence relating to: the research infrastructure; research training and support for staff in developing their research; and impact and user engagement.

13. Indicators of esteem and impact will carry a weighting of 10%. These may apply to individuals, research groups, or the department as a whole.

14. All sub-panels have carefully considered the standard RAE definition of the quality levels – from 4* to Unclassified – and have worked collaboratively to develop criteria for assessing the originality, significance and rigour of submitted research. Cross-membership between sub-panels will assist the chairs in ensuring that these criteria are applied consistently (my role). The main panel will also examine a sample of submissions at key points during the assessment process to ensure that subpanels continue to apply consistent standards of quality.

15. The quality levels are defined as follows:

- 4* – work assessed as reaching the 4* level will clearly demonstrate levels of originality, significance and rigour which are comparable to the best work in the field or sub-field whether conducted in the U.K. or elsewhere. Such work has been, or will be, recognised as making a significant or substantial contribution to knowledge, theory, policy, or practice in its field or sub-field. It has become, or is likely to become, a primary point of reference in its field or sub-field.
- 3* – work assessed as reaching the 3* level will demonstrate international standards of originality, significance and rigour. It has advanced, or is likely to advance, knowledge, theory, policy or practice in its field or subfield. It has become, or is likely to become, a major point of reference in its field or sub-field.
- 2* – work assessed as reaching the 2* level will demonstrate quality that is internationally recognised in terms of originality, significance and rigour. It has made, or will make, a contribution to knowledge, theory, policy or practice in its field or sub-field.
- 1* – work assessed as reaching the 1* level will demonstrate quality that is nationally recognised in terms of originality, significance and rigour. It has made, or will make, a limited contribution to knowledge, theory, policy or practice in its field or sub-field.
- Unclassified – work assessed as Unclassified will fall below the standard of nationally recognised work, or fail to meet the definition of research as set out above.

Some comments about RAE 2008: determining a research profile

There are three issues arising from the current assessment which are worth raising even though the assessment is ongoing at the time of writing. These are discussed in this and the next two sections. Again confidentiality restricts me to discussing general principles and practicalities, and any inference any one makes from what I write here about how the sub-panels are operating is entirely the product of that person's imagination.

In general, the search for a portfolio of outcomes for research publications could be addressed in a number of ways, but to make the point I intend, I shall look at just two of them.

The first is to identify a research activity within the submission to the sub-panel. This could be suggested by the submission itself, where the opportunity to label research groups and identify which paper is in which group is provided for. Alternatively, one could look at the submission and construct such groups from the unstructured set of publications. Then each group could be looked at 'in the round' and a rating allocated. This would allow a world leading edge group to be rated at 4* even though not all parts of the publications in that group are 4*. But then, why take a leading, or the world leading edge group, and allocate less than a 4* by giving bits of it 4*, 3*, etc.? The advantage of this approach is that from a broad view of all the publications, an indicative ranking could be determined. This could be reinforced or destabilised by selective reading of publications felt to be representative of the group research quality. Then, having established a tentative research rating, the less well known publications could be sampled as a test of publication quality consistency.

The second way of looking at research quality is by assigning a rating to each publication and then summing up all the ratings. Publications would again be sampled from both the familiar and the unfamiliar to ensure that

the ratings are largely robust. This is a very tedious process, and ignores any idea that the sum of the parts may not be as good as the whole. So in general overall assessments would be lower compared to the first method. The only advantage of this approach is that one can sit down and just get on with the rating without the need to look at a bigger picture. Maybe that can be added in after (how?).

Quite a dilemma for the assessment.

Some comments about RAE 2008: research environment and esteem

Having looked at research outputs, what should one make of research environment and esteem? One could look at these solely on the basis of the submission's text on these issues and make a judgement. But my subpanels have a maximum weighting of 70% for research publications because this is considered to be the prime indicator. And if the portfolio for research publication is (w%, x%, y%, z%), then I would have thought that this shows a research environment that at least is good enough to have sustained the publications, and one would hope the esteem of the group would be recognised externally for sustaining this portfolio.

So I would normally expect the research environment and esteem portfolios to at least match the publications output portfolio, and if the exit velocity of the research activity warrants it, to be slightly better than the research publications portfolio. Exceptionally, a research group in decline may accumulate research publication ratings that are better than its research environment and esteem for the end of the assessment period. Thus the overall portfolio would be worse than the publications portfolio indicating the declining exit velocity.

But apart from exit velocity, how can the outcomes for research environment and esteem be much different to the research output profile? Consider Table 2, a list of possible diverse results.

Do any of the rows in this table make any sense? Since at the end of the day research dissemination is the only evidence of any research having actually taken place, it follows for me that the other factors are derivatives of it. I look forward to hearing why I am wrong about this.

Some comments about RAE 2008: research quality and journal rankings

Journal league tables are now such a strong influence on researchers and their employers, that performance targets for researchers can be easily constructed around the number of 'A' grade journal papers and supporting non-'A' grade journal papers that are published or accepted for publication in any given time period. The implication might seem that quality papers are published in 'A' grade journals and vice versa.

But the RAE is assessing research quality and with that in mind the published outputs are being read and assessed. And the journal papers published in the top ranking journals have a quality that is not consistent

Table 2 Hypothetical combinations of profiles

<i>Research output</i>	<i>Research environment</i>	<i>Esteem</i>
Good	Poor	
Good		Poor
Poor	Good	
Poor		Good
Average	Good	
Average	Poor	
Average		Good
Average		Poor

with this expectation. In fact, for one journal that publishes information systems papers, the papers are largely given very low assessments even though where the journal appears in a league table, it will be at or near the top of the table. The reason for this is that the journal, with a very large circulation, has concentrated on making the research results clear to its wider and varied readership, and therefore wants little in the paper about the research content. So when assessing such a paper for research quality – there is hardly any!

That could of course be an anomaly of that particular journal, but in fact the lack of correspondence between the journal ranking and the perceived quality of any of its papers is quite striking. I offer the following tentative reasons for this, in order to generate for debate:

1. The RAE sub-panel members who are measuring the research quality of a paper are the only members of the U.K. research community doing this. So what is everyone else doing?
2. Journal referees are not measuring research quality, they are giving advice about whether the paper is publishable in the journal in its current state, or whether publishable after some agreed suggested amendments. The higher the ranking of the journal, the more demand to get published in it. This can allow for a long and meticulous review process before publication. And what gets published? The combined views of authors and reviewers in a paper which has had any content that is not rigorously supportable removed from it. So it is not uncommon for the final published version of the paper to be rather bland, self-evident and endlessly citable, since it is hardly likely to contradict the paper in which it is cited. A publication threshold says nothing about quality, especially how far above the threshold a paper might or might not be.
3. What about the readers of the journals? I guess most researchers, being very busy people, look at papers to seek anything relevant to their interests. A paper of little interest is neither good nor bad quality, it is just not of interest. If the paper is of interest, it is enthusiastically taken on board with an enthusiasm proportional to the interest. Note, this is not the same as research quality.

Given the above observations, it would appear to me that journal league tables are to do with promotion procedures and rites of passage in academia, not quality research.

And if this observation was not enough cause for concern, there is the impact on the RAE results to consider. Table 1 shows the four scoring values. Looking at the descriptors for these scoring values (fourth section), we can observe that 4* is going to be relatively scarce and that without careful retrospective overall evaluation of the rating process, the modal and median scores could well be a 2. Averages of the portfolio results will then rarely be above 3, mostly 2 point a bit, and often 1 point something. In a world where a four point score is used for student evaluation, and where a Grade Point Average (GPA) of 3 or more is required for graduate work (GPAs of less than 3 is usually only acceptable for undergraduate performance), this is not a smart outcome for all the effort going into the RAE.

References

- MAGEE B (1997) *Confessions of a Philosopher*. Weidenfeld and Nicolson, London.
- RAE 01/2004 – RAE2008: Initial decisions by the U.K. funding bodies, February. RAE documents: <http://www.rae.ac.uk/pubs/>.

But the RAE is about resource allocation not GPAs! The resource allocation model is derived after the results of the RAE are published. So the resource allocation can go wherever it is so desired just by constructing the model appropriately.

I suppose we should all be grateful for the opportunity to be shown the meaning of research quality so much more clearly.

I end with a thought based on a quote by Mark Twain: Wagner's music is better than it sounds.

- *Mark Twain's Autobiography* (re-quoting humorist Bill Nye)

<http://www.twainquotes.com/Opera.html>

So I guess one way to assess research output would be to assume

'The research quality of a research publication is better than it reads'

RAE 03/2005 – RAE2008: Guidance on submissions, June. RAE documents: <http://www.rae.ac.uk/pubs/>.

RAE 01/2006 (I) – RAE2008: panel criteria and working methods panel I. RAE documents: <http://www.rae.ac.uk/pubs/>.