

Special Articles

How should financial support for research be distributed to Universities? The Research Assessment Exercise (RAE) in England and Wales

SIR DAVID GOLDBERG and ANTHONY MANN

SUMMARY. In the United Kingdom, the "Research Assessment Exercise" is used by central government as a way of distributing infra-structure funds to University departments to support research. Departments with the highest ratings get extra support, while departments with low ratings may have their existing support withdrawn. The paper describes an exercise aimed at improving the rating obtained by one such department.

Declaration of Interest: none.

KEY WORDS: research support, improving quality, Research Assessment Exercise, RAE.

Received 27.11.2005 – Final version received 04.12.2005 – Accepted on 04.12.2005.

The Research Assessment Exercise (RAE) was introduced in 1993, and repeated in 1997 & 2001 as a way of determining how financial support for research should be distributed to Universities. Research funding is distributed according to the "dual support system", with some of the money being obtained by competitive tendering from government research councils and major charities, and some obtained directly in order to support research infrastructure. It is the latter that is now distributed on advice from the RAE. Thus, of the £1.8 billion science budget, about 40% is distributed to research councils, and the rest is distributed between universities.

It is worth recalling the situation before the first RAE, where quality of research produced by a university department played little or no part in determining the level of funding, and this resulted in poorly organised research programmes often without aiming for international excellence. In many university departments, some members of staff carried out very little research after achieving their university positions, and others pursued idiosyncratic projects without obtaining much benefit from collaboration with others.

The RAE provides quality ratings for research across all disciplines, and is aimed at ensuring that the institutions doing the best research get the most money. It measures how each university is doing, mostly on the basis of how many staff are involved in research, the number of journal articles they have published, numbers of postgraduate students completing projects, and amount of funding received from research councils and medical charities. The judgements about individual staff members are reduced to a single rating on a scale ranging from 1 to 5*, according to what proportion of the staff are producing work of "national" or "international" standard. A 5* rating means that more than half the staff are producing work of international standard.

All financial support for research was removed from Departments gaining the lowest ratings, and transferred to those achieving 5 and 5*.

By the time of the 1996 assessment, there was a great improvement in the number of departments achieving the top ratings, and this continued in the 2001 assessment - whereas in 1996 573 departments achieved these ratings, and these departments were employing 31% of the research active staff in the country, by 2001 no fewer than 1,081 departments got the top ratings, and these employed 55% of the staff. This dramatic improvement in the research activities of universities should have been rewarded with an increase of £205M .

Address for correspondence: Sir D. Goldberg, Health Services Research Department, Institute of Psychiatry, King's College London, de Crespigny Park, London SE5 8AF (United Kingdom).

Fax: +44-020-7848.0515

E-mail: spjudpb@hotmail.com

(Euro 297M), but in fact only £30M (Euro 43.5M) was allocated. This in turn meant that the level at which a rating was rewarded was greatly decreased, and there were increased numbers of departments receiving no funding at all, or greatly reduced amounts of support (House of Commons Select Committee, 2004).

If a researcher moved from University A to University B just before an assessment time, then the research carried out at University A was credited to University B - inevitably, this led to a flurry of active recruitment of staff just before the RAE was due.

Where psychiatry is concerned, half of the 22 departments rated in 2001 achieved the top ratings, and almost three quarters were rated as 4 or above, meaning that at least some of their research was judged to be at international standards. The downside of this was that fewer scientists were being submitted, and 4 Universities stopped submitting psychiatry to the RAE.

There has been a gradual shift in the way that quality has been judged - whereas in 1992 the total number of publications, together with the impact factors of the journals in which these articles appeared was used, by 2001 the focus was on the four best papers produced in the period since the previous assessment. In the next exercise in 2008 researchers will have to submit copies of the actual papers, and assessors will make quality judgements within the category of "international" research.

BENEFITS OF THE RAE

There is no doubt that the research output of the country as a whole has benefited immensely from the RAE, with Departments carrying out the best research receiving far more central support, and producing more focused research aiming at making maximum impact. In terms of numbers of scientific papers, the United Kingdom had a fairly similar output as France and Germany in the 1980ies, but has steadily increased its output since the RAE started, and is now far ahead of other European countries, and second only to the United States. The council responsible for distributing research support carried out a survey and claimed to find strong support for the system that funds only the best research: "there was overwhelming support from all respondents that the Council should continue to fund research selectively on the basis of quality. And 98% expressed support for retaining a process of research assessment based on peer review." (Higher Education Funding Council, 2003; Roberts, 2003).

The Institute of Psychiatry is an excellent example of a Department that has done extremely well out of the RAE, as we have received substantial extra monies that have allowed further expansion. New monies have been used to attract top class researchers both from the United States and Europe, and thus an already strong organisation has become much stronger. Unfortunately there are many places that would report a different experience.

DAMAGE DONE BY THE RAE

From the outset, the trade union for academic staff has vigorously opposed the whole process: "the RAE has had a disastrous impact on the higher education system, leading to the closure of departments with strong research profiles and healthy student recruitment. The RAE has been responsible for job losses, discriminatory practices, widespread demoralisation of staff, the narrowing of research opportunities through the over-concentration of funding and the undermining of the relationship between teaching and research, with a consequent reduction in the quality of higher education available to students (Association of University Teachers, 2005).

Williams (1998) argued that the assessment criteria used are "restrictive, flawed, and unscientific" and produce a distorted picture of research activity that can threaten the survival of active and productive research units. The assessment exercise is described as unaccountable, inefficient, time consuming, and expensive. Tomlinson (2000) writes that the research assessment exercise can result in major shifts in funding to or from individual medical schools and has led to a loss of status for teaching compared with research. It could result in the emergence of a small number of centres of excellence where internationally competitive research is undertaken while the remaining schools follow a spiral of decline to become "teaching only" medical schools. Since research informs teaching and clinical practice such a development would be undesirable both for those medical schools and the NHS.

Banatvala *et al.* (2005) refer to "the corporate insanity afflicting UK medical schools, where the combined shortfall of university funding has forced deans of medical schools to behave like managers of Premier Division soccer clubs, recruiting potential research stars to improve or maintain RAE ratings, at the expense of teaching and clinical practice. The numbers of clinical academic staff have been cut to recruit basic scientists whose research is likely to favour RAE ratings. The knock-on effect of reductions in academic staff is for staff in the NHS to carry an additional teaching burden,

which is often difficult to fulfill. It is disturbing that some schools, to save money, have failed to appoint professors when chair-holders retire or have redistributed funding to appoint non-medical scientists”.

THE TASK IN HAND

It is clear that the RAE causes major anxiety in all Universities: if ratings drop below the work being judged to be below “National” standards all funding will cease, while if one drops down from say 5* to 5, the loss of funding may necessitate staff being declared redundant, and programmes cut back. Rather than face such a prospect, Universities may decide to declare only their top quality departments, in order to retain a high rating.

In October 2004 we were asked by a University to do whatever we could to improve the rating likely to be attained by one of its large departments. The parent University might otherwise decide to close the department down before the assessment date approached.

In 2008 Papers published by members of the department between 2001 and the end of 2007 would be assessed. At a preliminary internal college review this department had not fared well, less than 40% of its senior staff being rated as having output of international importance. These ratings would have to improve if the department was to survive. The assessment of each individual is in fact based on the 4 papers each researcher submitted, and the papers had to be data based and published in high quality journals.

While the department has been very productive in terms of numbers of papers published, many of the staff had been content to publish in journals reflecting their area of expertise or in minor un-refereed publications. Unfortunately such journals tend to carry a low impact factor, although read widely by members of what may be a small sub-specialty, and are thus a sensible way of disseminating useful research findings. If impact factor had been the basis for rating of output, then an exercise done without any sensitivity to this fact would result in low ratings. Second, within the department in question, communication about the importance of RAE was not very well appreciated; thus some of the staff didn't feel that the RAE really involved them.

The categories we initially used for rating individuals by their output were as follows:

- I - international
- XI - potentially international by the time of the RAE
- N - national
- SN - below national

The internal review had classified only 10 members of the 26 staff who could possibly be declared as International, with a further 6 as on the margin between “I” and National. The University wanted the proportion rated “I” to increase substantially to at least 50% by the time of RAE.

Time scale

For a paper to be published by end 2007, we estimate that it would need to be submitted 18 months earlier, allowing six months for review and modification before acceptance and one year to publication after acceptance.

The approach

If staff members were to improve their own rating, they had a further 18 months in which to prepare and submit new papers to high quality journals. There would not be time to conduct new field work and analyse a new piece of research in that time. Thus, as outsiders, we believed our best help would be to review their planned work programme for the next eighteen months with the RAE in mind. Our rules were clear - publication in high impact journals were central to a good rating, and “international” ratings were more likely if at least one of these journals was a high quality American journal.

At the RAE itself the actual papers will have to be submitted, and their quality assessed directly - but as many papers were not yet written, we used the impact factor of the journal as a proxy measure.

We began by asking all members of the department who were to be declared as part of the department's submission to provide a list of existing publications, papers in preparation and of data sets to which they had access upon which more analyses were possible.

We then met each staff member individually using this information as the basis of discussion. Our plan was to have a further meeting after six months to see whether suggestions made at the first meeting were being followed up and a last meeting after eighteen months to decide which papers will be the four to be returned by that individual. We would then do our own rating of their quality independent of actual or likely impact factor. For those who were still in difficulties at the six months interview, we allowed the possibility of another review before the final one in 2006.

Those already regarded as of international standard by the University committee would be seen as part of the process, but the focus of the work was to be on those who needed to improve.

Strategies used in interview

It was first necessary to explain the rules of RAE and the emphasis on high quality output and that papers declared must be data based, thus eliminating most review and chapters from consideration. Although, we explained that our personal view was that the rules could be rigid and unfair, they were still the rules that were guiding our work.

We began by agreeing which if any of the already published papers would be suitable to return, an impact factor of 4 or more being the criterion for that. Fortunately the British Journal of Psychiatry, the most commonly used journal for the department's research, more than satisfies this criterion.

We then went on to see whether new papers in the pipeline could be well placed. For this, we asked for an outline of papers in preparation and assessed their likely success in reaching a high impact journal. For some we could suggest straight away the individual should raise his or her horizon for a particular paper, trying for journals that they hadn't considered. In particular, we asked whether the paper would be or could be made of interest to an American readership. For others, we asked for a paper in the queue for writing to be prioritised over the one currently receiving attention, if the former seemed to have a better chance of a high impact journal. For those with work that didn't seem to have a chance of a high impact journal but should find a home in a journal with impact factors between 2 and 4, we suggested that the paper should be submitted to a non-British journal with the same level of impact, thus indicating international interest in their work. We actively discouraged work on papers that would go to a journal with an impact factor of less than 2. We also actively discouraged such activities as chapter writing until the four papers had been submitted.

The third part of our interview was to go over the data sets to which the individual still had access, to discuss what had already been published and what could yet be mined. This proved fruitful in producing suggestions for new work. Of course, the analysis and write up would have to be done well within 18 months. Part of the discussion was thus to check the reality of such a proposal, and to decide what support might be necessary, perhaps from a statistician.

At the end of the interview with us, taking about an hour, every individual went away with an agreed prospectus for the next 18 months and an appointment to review progress in six months.

Responses

Individuals seen reacted to the intervention in a variety of ways. Many felt that it was awkward and a little humiliating to appear before peers in this way. They would have known that despite their seniority, they had not been rated as of international standard. Some dealt with this by feeling misjudged, others by belittling the procedure, and others were over-apologetic. To these it was necessary to point out the risks to the department's survival if ratings didn't improve and that we were there to advise, not judge. Some were pleased to have this opportunity to have advice on improving their CV.

We then went on to see whether new papers in the pipeline could be well placed. For this, we asked for an outline of papers in preparation and assessed their likely success in reaching a high impact journal. For some we could suggest straight away the individual should raise his or her horizon for a particular paper, trying for journals that they hadn't considered. In particular, we asked whether the paper would be or could be made of interest to an American readership. For others, we asked for a paper in the queue for writing to be prioritised over the one currently receiving attention, if the former seemed to have a better chance of a high impact journal. For those with work that didn't seem to have a chance of a high impact journal but should find a home in a journal with impact factors between 2 and 4, we suggested that the paper should be submitted to a non-British journal with the same level of impact, thus indicating international interest in their work. We actively discouraged work on papers that would go to a journal with an impact factor of less than 2. We also actively discouraged such activities as chapter writing until the four papers had been submitted.

The third part of our interview was to go over the data sets to which the individual still had access, to discuss what had already been published and what could yet be mined. This proved fruitful in producing suggestions for new work. Of course, the analysis and write up would have to be done well within 18 months. Part of the discussion was thus to check the reality of such a proposal, and to decide what support might be necessary, perhaps from a statistician.

At the end of the interview with us, taking about an hour, every individual went away with an agreed prospectus for the next 18 months and an appointment to review progress in six months.

RESULTS

Table I shows our initial assessments, the changes we rated as occurring by 6 months and what our prediction of the classification is likely to be in eighteen months time:

Table I.

	Not rated	S/N	XN	N	?XI	XI	I	Total rated
Our first assessment	(2)	3	1	8	2	1	10	25
Our second assessment	n/a	1	0	7	3	3	13	27
What might be achieved	n/a	0	1	1	3	9	14	28

These are:

- 4* - world-leading in terms of originality, significance and rigour,
- 3* - internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence, and 2* - recognised internationally in terms of originality, significance and rigour.

We thought it would be invidious to attempt to rate our colleagues using these highly subjective criteria, and to our great relief the head of the relevant department asked us not to do so.

CONCLUSION

The experience has been a new one, not only for ourselves but for our colleagues. There is an undoubted subjective element in the procedure, especially since our assessments depended upon or judgment of papers not yet written, and certainly not yet accepted by Editors. We

It can be seen that we have assessed most of the staff either of having already achieved international status, or at least of having that potential, if all goes well. At the RAE itself ratings will be made on a somewhat different system, as there will be 3 grades of "international" rating.

tried hard not to appear too judgmental, and we hope that we were consistently encouraging. To an extent we were successful in this, as some of our senior colleagues, highly sensitive at their first interview, became more relaxed and grateful for our comments on their draft papers. We can only wait to find how things turn out!

REFERENCES

- Association of University Teachers (2005). Retrieved March 13, 2006, from <http://www.aut.org.uk/index.cfm?articleid=130>
- Banatvala J., Bell P. & Symonds M. (2005). The Research Assessment Exercise is bad for UK medicine. *Lancet* 365(9464), 1025-1026.
- Higher Education Funding Council (2003). Retrieved March 13, 2006, from <http://www.hefce.ac.uk/research/assessment/>
- House of Commons Science & Technology Committee (2004). *Research Assessment Exercise: a Re-assessment. 11th Report on the Research Assessment Exercise*. HC 586. HMSO: London.
- Roberts G. (2003). Consultation by the UK funding bodies on the review. Retrieved March 13, 2006, from http://www.hefce.ac.uk/pubs/hefce/2003/03_22.htm
- Tomlinson S. (2000). The research assessment exercise and medical research *British Medical Journal* 320, 636-639.
- Williams G. (1998). Misleading, unscientific, and unjust: the United Kingdom's research assessment exercise. *British Medical Journal* 316, 1079-1082.