

Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science

Linda Butler and Ian McAllister

Australian National University

Evaluations of research quality in universities are now widely used in the advanced economies. The UK's Research Assessment Exercise (RAE) is the most highly developed of these research evaluations. This article uses the results from the 2001 RAE in political science to assess the utility of citations as a measure of outcome, relative to other possible indicators. The data come from the 4,400 submissions to the RAE political science panel. The 28,128 citations analysed relate not only to journal articles, but to all submitted publications – including authored and edited books and book chapters. The results show that citations are the most important predictor of the RAE outcome, followed by whether or not a department had a representative on the RAE panel. The results highlight the need to develop robust quantitative indicators to evaluate research quality which would obviate the need for a peer evaluation based on a large committee. Bibliometrics should form the main component of such a portfolio of quantitative indicators.

Citations have long been used to measure academic status and prestige in universities. Departments that are ranked at or near the top in such exercises can claim a greater share of the university's resources and can recruit the most able scholars and students, compared to those ranked near the bottom. Recent moves to align government funding more closely with research performance have largely been based on subjective indicators such as peer evaluation. What role can objective indicators such as citations play in predicting research performance and thereby determine the allocation of government resources to universities? In this article we explore the outcome of the 2001 Research Assessment Exercise (RAE) in political science in the UK in order to test whether citations could have been used to replace subjective peer evaluation.

The UK is an ideal case study for such an analysis. Among the advanced economies, the UK has the longest record of evaluating research performance and linking funding to the outcomes of the evaluation. The Research Assessment Exercise commenced in 1986 and by 1996 had become a major evaluation exercise, determining the allocation of a significant proportion of government funding for higher education. One of the core principles of the RAE has been transparency and the data on which the evaluations are made have always been publicly available.¹ This provides a unique opportunity to test how effective citations are in predicting the outcome of the RAE in UK political science departments in 2001. Our study has major implications for similar peer evaluations in other countries and disciplines. Could citations wholly or partly replace peer review for evaluating research quality?

Measuring Research Quality: Some Issues

Quantitative indicators of research performance – or 'metrics' – have risen to central prominence in proposals to change the way research funding is distributed in a number of



countries. Recent developments in the UK and Australia provide clear indications of the directions in which public policy is moving. In both countries, there has been general agreement that a range of indicators, subjective and objective, are necessary in any research assessment process, particularly when the outcome will determine the distribution of significant amounts of research funding (DEST, 2006; DfES, 2006a). In the UK, there was a clear rejection of relying solely on metrics to shape external funding in the design of the RAE (DfES, 2006b). The result was support for the development of a portfolio of indicators that would draw on measures covering three broad areas (research income, research infrastructure and research quality) with a process that saw the results from these indicators interpreted by a small group of experts.

The dilemma in most large-scale research evaluation exercises is to develop measures that have cross-disciplinary comparability. The vast majority of respondents to the Department of Education and Science consultation process agreed that it would be a retrograde step to develop separate assessment methods for science, technology, engineering and mathematics (STEM) subjects and non-STEM subjects. Herein lies the difficulty for most social science disciplines. While any metrics adopted in relation to research income and research infrastructure are likely to be applicable across all subjects, providing discipline-specific characteristics are incorporated into any model, the vexed area of output metrics presents real problems for most humanities and social sciences (HASS) subjects. The political science discipline is no exception.

The HASS sector in Australia has been successful in raising awareness of the difficulties its disciplines face in identifying suitable quality metrics (CHASS, 2005). The most frequently proposed quality measures are those based on citations in the journal literature. However, less than a quarter of HASS output appears in the Thomson (formerly Institute for Scientific Information [ISI]) citation databases, the most commonly used data source for these measures.² Much investigation has been undertaken to identify metrics that are equivalent to the standard STEM bibliometric measures, but that are appropriate for HASS disciplines. These studies have followed two paths: producing ranked lists of publication outlets (such as book publishers, conferences, performing or creative arts venues and journals); and investigating novel bibliometric measures (Butler and Visser, 2006; Lester, 1990).

In terms of novel bibliometric measures, the approach has been to use the Thomson databases to identify citations to all the most important publication outlets in a discipline. In the case of political science, this means extracting citations to books, book chapters and journals not indexed in the databases, in addition to citations to those journals that are indexed. Pilot studies have applied this approach to two HASS disciplines – history and political science – and brought together senior academics in each field to assess the results. Both disciplines gave strong support to the adoption of such measures, though with the proviso that any metrics based on this data must be interpreted by peers and not simply fed into a formula for the mechanistic distribution of funds (CHASS, 2006).

Given the support for these new bibliometric approaches in Australia, we sought to apply them to the UK to determine their effectiveness in any future metrics-based RAE. There

are additional advantages in undertaking a study based on UK data. While the Australian experts believed the novel bibliometric measures presented an accurate picture of the relative strengths and weaknesses of their discipline in Australia, there are no objective rankings against which to test this proposition. In the UK, by contrast, it is possible to use the 2001 RAE ratings to assess the efficacy of this approach. We did not assume that measures based on citations alone would necessarily correlate highly with ISI ratings in political science, despite findings supporting this in other disciplines (Oppenheim, 1997; Oppenheim and Norris, 2003). Instead, we sought to model the RAE outcomes using the data available to panels – external earnings, percentage of staff submitted, student data – but replacing the peer assessment of the four nominated outputs by citation data.

RAE Background and Procedures

The main purpose of the Research Assessment Exercise is to enable the higher education funding bodies to distribute public funds for research selectively on the basis of quality. Prior to the 1980s, university research funding in the UK was allocated by subject-based committees using largely subjective criteria on which to base their evaluations. The lack of transparency in this process, together with arguments about the contribution of university research to the national economy, resulted in the design of a more transparent, objective system of research assessment. The first RAE was conducted in 1986 as a small-scale pilot project, with a more comprehensive and transparent system following in 1989. The 1989 exercise was largely replicated in the following RAE, conducted in 1992. In 1996 a much bigger exercise was mounted, involving over 50,000 academic staff nominated as ‘research active’ by their universities who were evaluated by 69 committees representing discrete subject areas. The 1996 RAE represented the transition from ‘a quality assurance process to a competition for funding’ (UK Parliament, 2002).

The 2001 RAE – the fifth in the series – was substantially based on the 1996 exercise, though it was refined to take account of criticisms, most notably by making better provision for multidisciplinary research and by ensuring the consistency of scoring across the 68 subject area committees.³ The 2001 exercise considered a research output of over 150,000 publications⁴ from almost 50,000 academic staff in 2,598 submissions from 173 higher education institutions. In political science, 69 academic units were assessed, from 66 higher education institutions. In the case of three institutions – King’s College London, Nottingham Trent University and the University of Sussex – two assessable units were entered, as all institutions were free to do.⁵

Each assessable unit was asked to nominate their ‘research-active’ staff members, and to provide up to four publications for each nominated staff member ‘that reflect their best quality work over the period covered by the RAE’ (RAE, 2001a).⁶ The guidelines allowed for the possibility that a member would not have four assessable publications, particularly in the case of new entrants, those involved in major monographs and where there were ‘unusual or particular circumstances’. Of the 69 departments, 14 returned their maximum allocation of four submissions, with most of the remainder submitting at least 90 per cent of their maximum allocation. Only Huddersfield fell significantly below this level, with just

two-thirds of the number of publications they were able to put forward given their six academics.

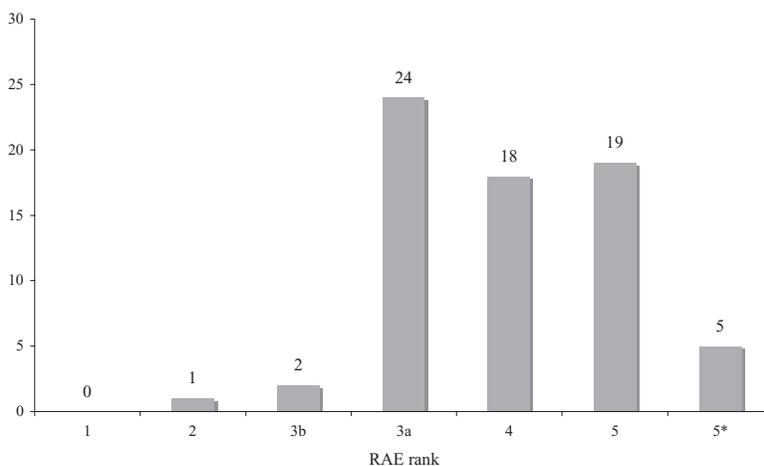
Modelling the 2001 RAE Outcome

The dependent variable is the rank given to each of the 69 academic departments in the RAE. The scoring adopted by the RAE was on a seven-point scale running from 1 to 5, with 3 being divided into '3a' and '3b', and 5 into '5' and '5-star'. A department was awarded the top 5-star ranking if, in the opinion of the panel members, more than half of its research activity was of international excellence, and the remainder demonstrated it had attained levels of national excellence. At the other end of the scale, a department was awarded a ranking of 1 if virtually none of the research activity it submitted was judged to have attained national excellence.

Figure 1 shows the distribution of the RAE scores across the political science discipline; no department received the lowest grade of 1, and just one – Southampton Institute – received a score of 2, with two others (Derby and Middlesex) receiving a grade of 3b.⁷ Five departments – Aberystwyth, Essex, King's College (War Studies), Oxford and Sheffield – received the highest grade of 5-star. The median political science department received a grade of 4, compared to 3a in 1996.⁸ For the purposes of the analyses that follow, these grades have been converted into a 1-to-6 ordinal scale, where 1 represents an RAE grade of 2, and 6 an RAE grade of 5-star.

Each of the 69 subject area panels laid down clear rules in advance for evaluating their respective fields. In the case of political science, the panel identified three evaluation criteria:

Figure 1: RAE Grade in 69 Political Science Departments



Notes: Figures are the number of political science departments allocated to each of seven possible grades used in the 2001 RAE.

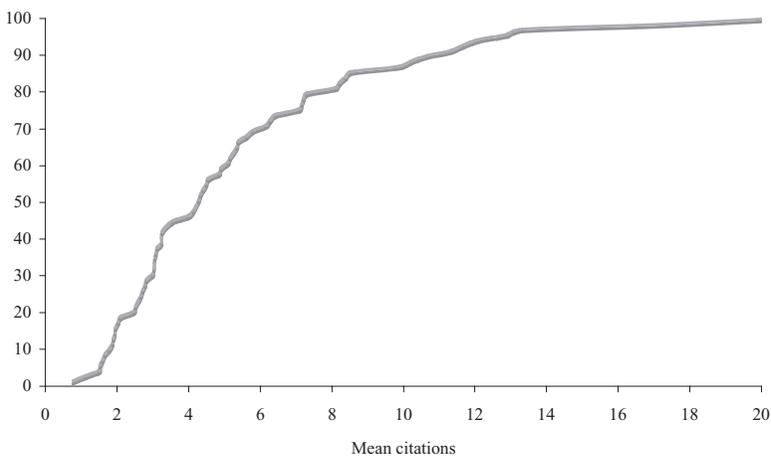
- (1) the quality of publications or other outputs;
- (2) evidence of the vitality of the research culture; and
- (3) evidence of the quality of research activity as reflected by a number of other factors such as research postgraduates, esteem indicators and contributions to academic public goods (RAE, 2001a).

In the case of publications, the panel committed to collectively reading at least one-quarter of the submitted output of a department, with a greater proportion of publications being examined if the outcome was judged to be borderline. The procedure was to evaluate the publications using 'the standards of research that are employed in a peer-review process' (RAE, 2001a). Authored books were judged most important, followed by journal articles and chapters in edited books, and lastly edited books.⁹ The quality of the publications is measured here by the mean citations received for the publications entered by each staff member.¹⁰

At this point, it is worth noting an important and novel aspect of our citation analysis. Our counts were obtained by interrogating ISI's Web of Science for citations to all 4,400 publications submitted for assessment to the RAE. They were not restricted to citations received by articles in journals indexed by ISI, and hence are far more comprehensive than standard citation measures. They are also more defensible for a discipline such as political science, where only a quarter of submitted outputs were articles in ISI journals.

As we would expect, the distribution of citations across the 69 departments is not linear, and for that reason we use a log transformation in the multivariate analyses (Figure 2). While the actual number of mean citations per submitted work ranges from a low of 0.78 (for Middlesex) to a high of 20.21 (for Birkbeck), the overall mean is 5.34. Indeed, while there

Figure 2: The Distribution of Citations



*Notes: Figures are the mean citations per submitted work per department. N = 69 academic political science departments.*¹³

are 41 departments that scored below 5 citations per submitted work, there were just 2 departments – Westminster and Birkbeck – that scored more than 15 and only 9 in total that scored more than 10.¹¹ In the case of Westminster and Birkbeck, their exceptional performance on citations was associated with a small number of highly cited works. In Westminster, one of the submitted books received 141 citations and another 118 citations,¹² while in Birkbeck the high score was substantially attributable to the 711 citations received by one book, Paul Hirst's *Globalization in Question* (Polity, first edition 1996).

Several other variables can be identified to represent 'the vitality of the research culture' and the 'quality of the research activity'. The model includes the annual research income from research councils covered by the Office of Science and Technology and the British Academy (OST/BA). The mean OST/BA income across the 69 departments was £8,630, but varied enormously, with twelve departments receiving nothing, and one – Sussex-SPRU – receiving £119,930 (see Table 1). For ease of interpretation, this is measured here in thousands of pounds. Research 'esteem', always a difficult concept to quantify, is measured here by membership of national academies. The 3 national academies in the UK are the Royal Society of Edinburgh, the Royal Irish Academy and the British Academy, and 13 of the 69 departments had at least one academy member among their staff, with Oxford having the largest number, 9, followed by the LSE with 3.¹⁴

The size of a particular department is often regarded as an indicator of research activity in its own right, with larger departments capitalising on their size to generate an active research culture and research groupings that attract further interest and resources (Jackman and Siverson, 1996). The model includes the number of academic staff submitted in the RAE – in other words, the number considered 'research active'.¹⁵ The mean number of such research-active staff was 17, ranging from a low of 3 at Wolverhampton to 80 at the LSE. The number of graduate students is also a good measure of research activity, since high-profile researchers will attract students who want to work with them. The mean number of graduate students, measured as full-time equivalents per research-active staff, was 6.02, with York having the largest ratio at 55.15.¹⁶

Table 1: Variables, Definitions, Means

<i>Variable</i>	<i>Scoring</i>	<i>Mean</i>
RAE ranking	1 = 2, 2 = 3a, 3 = 3b, 4 = 4, 5 = 5, 6 = 5-star	3.97
Citations (log)	Mean citations per submitted work, logged	1.45
Research Culture		
OST/BA annual income	£1,000s	8.63
National academy membership	1 = member(s) in department, 0 = none	0.19
Department size		
Academic staff	Number	17.00
Graduate students	Mean students per staff member	6.02
Department member on RAE panel	1 = yes, 0 = no	0.17
(N)		(69)

Finally, the model controls for whether or not the academic department had a member on the RAE panel. The panel consisted of 12 members, with half being drawn from the top-ranked departments in the previous RAE; of the 4 departments ranked as 5-star in 1996, 3 were represented on the 2001 panel, as were 3 of the 6 departments ranked 5 in 1996. How effective is the model in predicting the outcome of the RAE for the 69 political science departments? The results of the OLS regression analysis¹⁷ in Table 2 suggest that the model works well, explaining almost two-thirds of the variance in the RAE outcome. The mean number of citations a work attracts significantly improves the RAE outcome for a department and is the most important predictor in the model. Citations are, therefore, an important indicator of research quality as judged by peer evaluation through the RAE, as indeed we would expect.

The second important predictor of outcome for a department is having a member on the RAE panel; net of other things, the partial coefficient of 0.871 suggests that this is worth almost one point on the RAE scale. For example, a department that was awarded a 4 ranking in the RAE but did not have a staff member on the panel could have expected to receive close to a 5 ranking if one of its members had been on the panel – net of a wide variety of other factors. This is a substantial effect and suggests the importance of understanding the RAE evaluation system and of devising a strategy to maximise a department's strengths. It is also an indication that, for general acceptance within the sector, panel

Table 2: Predicting RAE Outcome (OLS Estimates)

	<i>RAE outcome</i>		
	<i>Partial</i>	<i>(SE)</i>	<i>Standard.</i>
Citations (log)	0.541**	(0.148)	0.340**
Department member on RAE panel	0.871**	(0.241)	0.307**
Department size			
Academic staff	0.019	(0.008)	0.233*
Graduate students	0.030*	(0.012)	0.203*
Research culture			
OST/BA annual income	0.007	(0.005)	0.116
National academy membership	0.147	(0.271)	0.053
Constant	2.435	(0.495)	
Adj R-squared	0.622		
(N)	(69)		

**statistically significant at $p < 0.01$; * $p < 0.05$.

Notes: Ordinary least squares regression equations showing partial and standardised coefficients and standard errors predicting RAE outcome. See Table 1 for details of variables and scoring. N = 69 political science departments.

membership needs to be drawn from the elite echelon of researchers who are normally associated with highly rated departments.

The remaining variables in the model are of lesser importance. The size of the academic department is important, with larger departments attracting a higher ranking, net of other things. Similarly, a larger graduate ratio increases the likelihood of a more favourable RAE outcome. For example, York, with a ratio of 55 students per staff member, could have expected to receive an RAE outcome more than one point higher than Edinburgh, with five students per staff member, due solely to York's much larger student size.¹⁸ By contrast, the two measures of research culture, external income and national academy memberships, are not significant predictors, net of other things. This finding in relation to external income is particularly significant given original proposals for the new 'metrics RAE' to be based solely on this data (DfES, 2006c). At least for political science, this approach would not work.

The model works well, therefore, in predicting RAE outcome, with citations – as a leading objective indicator of research quality – being the most important predictor of the outcome, as we would expect. More unexpectedly, having a department member on the RAE panel is also a very significant influence on the outcome. This finding reflects the complexity of the evaluation process and the importance of a department gaining good intelligence in order to adopt an effective assessment strategy. It also highlights the inevitability of drawing panel memberships primarily from departments ranked highly in previous exercises.

Disaggregating Citations

Citations were, then, the most important factor in predicting the outcome of the 2001 RAE in political science. But in our methodology, citations cover a range of different publications, all of which are associated with different patterns and rates of citation. Which types of publication were more likely to attract citations and therefore to feed into a better RAE outcome? The formal advice of the panel was that they would rate authored books most highly, followed by peer-reviewed journal articles. The panel indicated that chapters in edited books would be evaluated 'by the same standards of research excellence as articles in journals' (RAE, 2001a), although presumably the absence of peer review would in most cases mean that chapters were less influential than articles. Edited books were to be ranked less highly than authored books, although it was recognised 'that, on occasion, an edited book may involve a high degree of research effort' (RAE, 2001a) and in that circumstance an edited book would be given greater weight.

The pattern of submissions to the RAE shows the predominance of articles, with just over half of all submissions being journal articles (Table 3). Chapters in edited books accounted for just under one-quarter of total submissions, and authored books one-fifth. In line with their lower standing in the eyes of the panel, edited books accounted for just 3.5 per cent of all submissions. Authored books attracted the greatest number of citations – 16,030 – which represented a mean of 18.2 citations for every submitted book. Journal articles were less well cited, with an average of 4.1 citations per submission. The edited books that were

Table 3: Type of Submitted Publication and Citations

	<i>Submitted works</i>		<i>Citations</i>		<i>Mean cpp</i>
	<i>Total N</i>	<i>Percent</i>	<i>Total N</i>	<i>Percent</i>	
Books	880	20.0	16,030	57.1	18.2
Journal articles	2,251	51.2	9,152	32.5	4.1
Book chapters	1,021	23.2	1,808	6.4	1.8
Edited books	155	3.5	1,072	3.8	6.9
Others	93	2.1	66	0.2	0.7
Total	4,400	100.0	28,128	100.0	6.4

Note: 'Mean cpp' is mean citations per publication.

Table 4: Citations by Type of Publication and RAE Outcome (OLS Estimates)

	<i>RAE outcome</i>		
	<i>Partial</i>	<i>(SE)</i>	<i>Standard.</i>
Citations (log of)			
Authoried books	0.151*	(0.070)	0.223*
Journal articles	0.585**	(0.146)	0.427**
Book chapters	0.115*	(0.046)	0.250*
Edited books	0.002	(0.020)	0.008
Other	0.027	(0.031)	0.089
Constant	3.357	(0.279)	
Adj R-squared	0.383		
(N)	(69)		

**statistically significant at $p < 0.01$; * $p < 0.05$.

Notes: Ordinary least squares regression equations showing partial and standardised coefficients and standard errors predicting RAE outcome. $N = 69$ political science departments.

submitted for assessment attracted an average of 6.9 citations each, perhaps reflecting the fact that the small number that were submitted were of high quality. Chapters in edited books attracted an average of just 1.8 citations for each one that was submitted.

Which type of publication, reflected in mean citations, had the greatest influence on predicting the outcome of the RAE? To make this estimate, Table 4 uses the log of the mean citations of each of the five types of publication to predict RAE outcome, without the controls used previously.¹⁹ The results show that articles were by far the most important of the five types of publication in predicting the RAE outcome, and were almost twice as

important as book citations – perhaps surprisingly given the weight placed on authored books by the panel. Book chapters were similar in importance to books, while edited books and other types of submissions had no significant effect on RAE outcome.

These results suggest that journal articles were of greatest importance – compared to other types of publication – in predicting a department's research quality rating. Given the emphasis on authored books, the much lesser importance of citations to books in forecasting the research outcome is surprising. 'Authored books' as a category does, of course, encompass many types of books, ranging from research monographs with prestigious university publishers to first-year texts with mainstream commercial publishers. Without investigating the books submitted to the RAE in detail, we are unable to analyse this further. What is clear, however, is that edited books, whatever the citations they received, counted for little, though chapters included in such edited volumes did have a significant effect on the outcome.

Metrics and Research Quality

The results presented so far have shown that objective indicators of research quality, led by citations, predict a large proportion of the variance in the 2001 RAE outcome in political science, which was arrived at through peer evaluation. The results also show that membership of the RAE panel had a significant influence on the outcome, net of other things. In this section we consider to what extent the use of peer evaluation to measure research quality could be dispensed with, relying instead solely on objective measures such as citations, income and student numbers. If the 2001 RAE in political science had adopted this metrics-only approach, how would the results have differed from what was actually decided upon, using peer evaluation?

To make these estimates, we use the regression model presented in Table 2, but omitting the variable measuring membership of the RAE panel. This reduced model accounts for 55 per cent of the variance in RAE outcome, just slightly less than that for the model that includes RAE panel membership. To calculate the predicted RAE outcome using this model, the regression equation is re-estimated for each of the 69 departments, substituting the means for each department into the equation and calculating the predicted RAE outcome.²⁰ The predicted outcome is then compared with the actual outcome to see which departments have shifted up or down in grade as a result of moving to a solely metrics-based evaluation. The results of this exercise are shown in Table 5.

Overall, only one of the 69 departments would have received an outcome on the metrics model which would have differed by more than one unit on the peer evaluation grading used by the RAE. This is Southampton Institute which moves from a 2 to a 3a grade. Of the remaining departments, Table 5 shows that eleven departments would have improved their position by one grade, while eighteen departments would have moved down one grade. In general, the twelve departments on the left-hand side of Table 5 that would have improved their position were lower-ranking departments, the only exception being the LSE, which would have moved up from a 5 to a 5-star grade.

Table 5: Predicting RAE Outcome (OLS Estimates)

	<i>Over-prediction</i>			<i>Under-prediction</i>	
	<i>Actual</i>	<i>Predicted</i>		<i>Actual</i>	<i>Predicted</i>
Southampton Institute	2	3a	King's College (War)	5*	5
LSE	5	5*	Queens Univ. Belfast	5	4
Sussex (Politics, IR)	4	5	Bradford	5	4
Westminster	4	5	St Andrews	5	4
Univ. College London	3a	4	Reading	5	4
Cambridge	4	5	De Montfort	5	4
Kent	3a	4	Exeter	5	4
Open University	3a	4	Sheffield	5*	5
Derby	3b	3a	Aberystwyth	5*	5
Swansea	3a	4	Bristol	5	4
Southampton	4	5	Essex	5*	5
Wolverhampton	3a	4	Coventry	4	3a
			Birmingham	5	4
			Dundee	4	3a
			Keele	5	4
			Hull	5	4
			Durham	4	3a
			Queen Mary	4	3a

Notes: The 'actual' figure is the 2001 RAE outcome for each department. The 'predicted' figure is the prediction based on the OLS regression model in Table 2, minus membership of the RAE panel; see text for details of estimation. Only departments that changed in grade due to the metrics model are shown, and the predicted figure has been rounded to the nearest grade. The departments are ranked by the size of the difference between the actual and the predicted outcome.

The eighteen departments that would have dropped one grade on the right-hand side of Table 5 are, conversely, disproportionately drawn from the higher-ranking departments. Four of the five departments ranked as 5-star – Aberystwyth, Essex, King's College (War Studies) and Sheffield – would have dropped to a 5 grade, leaving only Oxford (along with the LSE, which would have moved from 5 to 5-star) at that level. At the other end of the scale, four departments – Coventry, Dundee, Durham and Queen Mary – would have dropped from a 4 to a 3a grade.

These results, while affecting the grades of 30 of the 69 departments, suggest that a metrics approach to measuring research quality would produce results which are very close to the results that come from peer evaluation of research. Moreover, the metrics results are more defensible, since they are achieved without recourse to a panel of assessors drawn from departments that are themselves being assessed, and the accompanying risk of accusations of real or imagined conflicts of interest. The robustness of the metrics-based results suggests to us that such evaluations are a much more efficient and transparent approach to measuring research quality.

Discussion

Evaluations of research output in higher education are rapidly becoming the norm across a range of developed countries. The UK's Research Assessment Exercise is perhaps the best developed of these evaluations, starting life in 1986 as a modest pilot project but by the mid-1990s becoming a large-scale national evaluation which has shaped much of the higher education sector. The results of this periodic assessment have been impressive: following the 1996 RAE, 43 per cent of the research that was evaluated was rated as of 'national' or 'international' standard; the same figure after the 2001 RAE was 63 per cent. In 2001, nineteen universities had an average departmental score of 5 or above, up from three in 1996.

Despite this impressive increase in research quality between 1996 and 2001, the funding organisations have been unable to increase significantly the resources required to reward this improvement. This has inevitably led to questions about the time and effort required in preparing for and carrying out such a massive exercise. And in the absence of significantly increased government funding for higher education, the ability of future RAEs to leverage further significant returns to the sector through additional improvement in the quality of UK research is unlikely (Geuna and Martin, 2003).

One obvious approach to streamlining the RAE and reducing the burden to universities is to rely more heavily on quantitative indicators of research. Our aim here has been to assess the role of bibliometrics in such an evaluation – perhaps the most readily available, comprehensive and widely accepted set of quantitative measures available in academic research. Bibliometric analysis is not without its problems,²¹ but we were careful to choose a methodology sympathetic to the publication practices of political science. By examining the 4,400 submitted works to the 2001 RAE political science panel, and the 28,128 citations that these works attracted, we conclude that citations were easily the most important component in predicting the outcome of the RAE, as indeed we would expect. But unexpectedly, whether the department had a member on the RAE panel was the second most important predictor of the RAE outcome.

While the RAE took strenuous and worthy efforts to avoid any actual or perceived conflict of interest, inevitably RAE panel membership brings with it an intimate knowledge about the process and procedures of evaluation which will convey indirect but tangible benefits to the person's own department. Inevitably, then, the process highlights the disadvantages of mounting subjective peer evaluation exercises which involve large committees composed of members who are tasked with evaluating each other's departments. The committee members help to shape the rules for evaluation but, more importantly, they understand them and how they can be used to present a department in the most favourable light. Another disadvantage is that panel members are disproportionately drawn from the most highly regarded departments, further advantaging the top departments in what is a highly competitive exercise with very significant resource implications for universities. Our findings suggest, unequivocally, that a metrics-based model, using objective, transparent indicators drawn from a range of readily available measures, will yield results which are very close to those of a peer-based evaluation model. Such a stronger reliance on quantitative

indicators, of which bibliometrics is a central measure, will, most importantly, obviate the need for a large peer review committee and the consequent indirect biases that it introduces into the system. And not least, such an approach would help to reduce one of the most oft-quoted criticisms of the RAE, namely the cost in university resources and in academic staff time. Once a set of quantitative indicators is agreed upon, only a small panel of experts would be required to scrutinise the data for anomalies and to weigh up the merits of any claimed disadvantage from the choice of metrics. The data collection and estimates could then be delegated to a committee of administrators.

Such a metrics-based approach would help to ensure that future RAEs continue to achieve the impressive improvements in research quality that have been achieved in the past. Most importantly, it would do so at significantly reduced cost.

(Accepted: 4 December 2007)

About the Authors

Linda Butler, Research School of Social Sciences, Australian National University, Canberra, ACT 0200, Australia; email: linda.butler@anu.edu.au

Ian McAllister, Research School of Social Sciences, Australian National University, Canberra, ACT 0200, Australia; email: ian.mcallister@anu.edu.au

Notes

Thanks to Kasper Hansen, Ben Martin, Paul Mitchell, Jonathan Tonge, Stephen White and two anonymous reviewers from this journal for comments on earlier drafts of the article. The usual disclaimer applies.

- 1 All of the data used here are available at <http://www.hero.ac.uk/rae/>
- 2 Elsevier is a recent entrant into bibliometric analysis with the establishment of its Scopus database, which is now gaining some following in the academic community.
- 3 The subject areas were reduced by one, to 68 in 2001.
- 4 Research outputs are usually, but not always, publications. Other types of output – including films, exhibitions or contributions to television programmes – are increasingly common (UK Parliament, 2002).
- 5 The units were: King's College London (War Studies and Defence Studies); Nottingham Trent University (International Relations, and Politics and International Studies); and the University of Sussex (SPRU – Science and Technology Policy Research, and Politics and International Studies).
- 6 The reduction in the maximum number of submitted publications per entrant to four was a key change from 1996, and was designed to reduce the workload on the RAE committees.
- 7 Departments judged to be at the bottom end of the scale and likely to receive scores below 3a do not attract any funding, and universities generally do not submit these weaker departments for assessment – the compliance cost of submissions making it not worthwhile.
- 8 RAE (2001b). The report noted that of the 61 departments that submitted in both 1996 and 2001, 41 increased their grade while just 3 departments dropped a grade.
- 9 The panel noted that 'it is recognised that, on occasion, an edited book may involve a high degree of research effort'. Other works were also permitted, such as textbooks, CDs and websites, and would be assessed provided they 'can be shown to contain significant research effort'.
- 10 Other ways of measuring citations are total citations and the Hirsh index (Hirsh, 2006). The measure used here – mean citations per submitted work, across the 69 departments – is the most straightforward. The smallest department (in terms of submissions) was Wolverhampton which had 3 persons making 12 submissions, followed by Liverpool John Moores with 4 persons making 14 submissions. The largest were the LSE (80 persons, 304 submissions) and Oxford (76 persons, 290 submissions). Early analyses controlled for the year in which publication of the submitted work took place, since the period of time that has elapsed might influence the opportunities for the work to be cited. Only works published between 1996 and 2000 could be submitted to the RAE. This control made no substantive difference to the results and has been omitted in the interests of presenting as parsimonious a model as possible.

- 11 Namely: Aberystwyth (10.21); Sussex Politics (10.68); LSE (11.33); Newcastle (11.69); Glasgow (12.09); Southampton (12.89); Strathclyde (13.32); Westminster (17.44); Birkbeck (20.21).
- 12 Namely: Barry Buzan, Ole Wæver and Jaap de Wilde, *Security: A New Framework for Analysis* (Boulder CO/London: Lynne Reiner, 1998); Chantal Mouffe, *The Democratic Paradox* (London: Verso, 2000).
- 13 Citations were extracted from Thomson's Web of Science during September and October 2006.
- 14 The estimate was made in 2001, when the RAE submissions were entered. There are, of course, other more complex ways of measuring research 'esteem' (visiting fellowships, honorary degrees, prestigious overseas appointments and so on), but that is beyond the scope of this exercise.
- 15 A valuable measure would of course be the total number of academic staff in a department, but that was unfortunately not available.
- 16 We have not distinguished between postgraduate taught students – who may be pursuing degrees with very little research component – and postgraduate research students, who should be engaged in more research-intensive study – as that was not available in the published data.
- 17 The estimates were also calculated using logistic regression techniques (to take into account the skewed distribution of the dependent variable). Since the results were substantially the same as the OLS model, we have relied on the latter here, as the results are more easily interpretable.
- 18 The calculation is 55 minus 5 = 50. $50 * 0.030 = 1.5$.
- 19 The introduction of controls, as in Table 2, reduces the weight of the publications in predicting RAE outcome but not the relative ordering. Since the introduction of more independent variables reduces the degrees of freedom in what is already a small N of cases, we use the simpler analysis presented here.
- 20 The equation was:
 Predicted RAE outcome = 2.445 + (log mean citations * 0.605) + (number of academic staff * 0.020) + (graduate students * 0.027) + (income * 0.006) + (national academy membership * 0.500).
- 21 For a recent detailed overview of the role of bibliometrics in a national research assessment, see Butler (2007).

References

- Butler, L. (2007) 'Assessing University Research: A Plea for A Balanced Approach', *Science and Public Policy*, 34 (8), 565–74.
- Butler, L. and Visser, M. S. (2006) 'Extending Citation Analysis to Non-source Items', *Scientometrics*, 66 (2), 327–43.
- Council for the Humanities, Arts and Social Sciences (CHASS) (2005) Measures of Quality and Impact of Publicly-Funded Research in the Humanities, Arts and Social Sciences. Available from: <http://www.chass.org.au/op2.pdf> [Accessed 26 April 2007].
- Council for the Humanities, Arts and Social Sciences (CHASS) (2006) CHASS Bibliometrics Project, Political Science and History Panels, Report on Recommendations and Major Issues. Available from: http://www.chass.org.au/papers/bibliometrics/CHASS_Report.pdf [Accessed 26 April 2007].
- Department of Education, Science and Training (DEST) (2006) Research Quality Framework. Assessing the Quality and Impact of Research in Australia. The Recommended RQF. Available from: http://www.dest.gov.au/NR/rdonlyres/7E5FDEBD-3663-4144-8FBE-AE5E6EE47D29/14867/Recommended_RQF_Dec2006.pdf [Accessed 26 April 2007].
- Department for Education and Skills (DfES) (2006a) Reform of Research Assessment and Funding (Letter from DfES to HEFCE). Available from: <http://www.hefce.ac.uk/News/HEFCE/2006/LetterAJtoDY.pdf> [Accessed 26 April 2007].
- Department for Education and Skills (DfES) (2006b) DfES Consultation on the Reform of Higher Education Research Assessment and Funding: Summary of Responses. Available from: <http://www.dfes.gov.uk/consultations/conResults.cfm?consultationId=1404> [Accessed 26 April 2007].
- Department for Education and Skills (DfES) (2006c) DfES Reform of Higher Education Research Assessment and Funding. Available from: <http://www.dfes.gov.uk/consultations/conResults.cfm?consultationId=1404> [Accessed 16 May 2007].
- Geuna, A. and Martin, B. R. (2003) 'University Research Evaluation and Funding: An International Comparison', *Minerva*, 41 (4), 277–304.
- Hirsh, J. E. (2006) 'An Index to Quantify an Individual's Scientific Research Output', *Proceedings of the National Academy of Science*, 102 (46), 16569–72.
- Jackman, R. W. and Siverson, R. M. (1996) 'Rating the Rating: An Analysis of the National Research Council's Appraisal of Political Science PhD Programs', *PS: Political Science and Politics*, 29 (2), 155–60.
- Lester, J. (1990) 'Evaluating the Evaluators: Accrediting Knowledge and the Ranking of Political Science Journals', *PS: Political Science and Politics*, 23 (3), 445–7.

- Oppenheim, C. (1997) 'The Correlation between Citation Counts and the 1992 Research Assessment Exercise Ratings for British Research in Genetics, Anatomy and Archaeology', *Journal of Documentation*, 53 (5), 477–87.
- Oppenheim, C. and Norris, M. (2003) 'Citation Counts and the Research Assessment Exercise V: Archaeology and the 2001 RAE', *Journal of Documentation*, 56 (6), 709–30.
- Research Assessment Exercise (RAE) (2001a) Panels' Criteria and Working Methods. 3.31 Politics and International Studies, UoA 39. Available from: http://admin.hero.ac.uk/rae/Pubs/5_99/ByUoA/crit39.htm [Accessed 14 April 2007].
- Research Assessment Exercise (RAE) (2001b) Report on Disciplines: Panel 39: Politics and International Relations. Available from: <http://www.hero.ac.uk/rae/overview/> [Accessed 2 April 2007].
- UK Parliament (2002) Memorandum Submitted by the Higher Education Funding Council for England (HEFCE). Available from: <http://www.parliament.the-stationery-office.co.uk/pa/cm200102/cmselect/cmsctech/507/2012302.htm#n7> [Accessed 10 April 2007].