Reputations and research quality in British political science: the importance of journal and publisher rankings in the 2008 RAE

Nicholas Allen, Royal Holloway, University of London

Oliver Heath, Royal Holloway, University of London

Abstract

This paper analyses the results of the 2008 Research Assessment Exercise (RAE). It demonstrates that the reputations of political science journals and scholarly publishers can explain the performance of institutions submitted to the RAE's Politics and International Studies sub-panel, and that there were also clear relationships between types of output and research quality. Outputs in top journals and with top presses were strongly associated with 4* quality and research excellence. Moreover, press and journal reputations appeared to have a greater impact than the type of publication. These findings should encourage policy makers to consider more cost-effective and efficient ways of evaluating research.

Introduction

This short paper began life as a minor difference of opinion between colleagues. The exchange was cordial and inconsequential, insofar as departmental harmony was concerned, but it did prompt us to engage with broader questions about the measurement of research quality in political science. The exchange began when one colleague circulated an email with details of yet another attempt to rank political science journals (McLean et al., 2009a). He also suggested that some institutions were using this list, and others like it, as a rule of thumb for gauging research quality ahead of the 2014 Research Excellence Framework (REF). A more senior colleague responded by rehearsing a number of well-known flaws in such rankings and noting that pre-constructed rankings of any kind had been banned from the REF sub-panel responsible for Politics and International Relations. This colleague was absolutely right, of course. Such rankings have indeed been banned, just as they were banned from the 2008 Research Assessment Exercise (RAE), the most recent authoritative audit of research in British Higher Education institutions. And so there the discussion ended. Except that it led us to ask whether such rankings were, in fact, consistent with past judgements about research quality.

To this end, we analysed the results of the 2008 RAE. More specifically, we set out to test whether the reputations of political science journals and scholarly publishers, as measured in surveys conducted before the RAE census date (Garand et al., 2009; Goodson et al.1999; McLean et al., 2009a), could explain the relative performance of institutions that were submitted to the Politics and International Studies sub-panel in 2008. Even more specifically, we set out to test whether the number of books published with 'top' presses and the number of articles published in 'top' journals, as defined by the rankings and as a proportion of all the submitted

outputs, were associated with the percentage of an institution's output that was graded 4*, the top band of research quality, in 2008. We also set out to test whether there were any systematic differences in how types of research were coded, for example, whether monographs were more strongly associated with 4* ratings than articles, and whether articles were more strongly associated than book chapters. The short answer to all these questions was broadly 'yes', as we will see. The analysis set out in the following pages attempts to flesh out this answer and address some of the normative implications.

Before going any further, however, it is worth emphasising two caveats for the benefit of readers. The first caveat, which is implicit in the previous paragraph, is that we are interested primarily in accounting for research judgements at the top end of the spectrum i.e. the proportion of submissions that were judged to be 4*. Attempts to explain previous RAE results have used dependent variables that 'average out' institutions' performances in different quality bands (see, for example, Butler and McAllister). Such averages require assumptions about the relative merit of different bands, for example, whether there is a linear progression in quality between 1* work, 2* work, 3* work and 4* work, or whether quality follows some form of curvilinear progression. We focus on 4* research, partly to avoid making such assumptions, but largely because it is much the most important level of research quality. It signifies excellence, and it has the most bearing on how public money is distributed to Higher Education institutions.[1]

The second caveat relates to the limited nature of our aim. The question that motivates us is largely empirical: it is whether the media of research outputs submitted to the 2008 RAE correlate with the authoritative judgements about research quality. The broader normative question of whether reputational surveys—or some

other form of metric—should replace peer review by a small elite panel of experts is very much secondary in our considerations. To that end, we do not seek to produce a comprehensive model for measuring research quality, assuming that such a model is possible, let alone desirable. That is not our purpose. Nevertheless, our findings obviously speak to the metric-versus-peer-review debate (Russell, 2009; Weale, 2009), and those interested in this debate should find them of great interest.

The remainder of this paper is divided into six parts. The first part provides a brief recap of the 2008 RAE, including its working methods and results. The second part briefly reviews attempts to rank scholarly publishers and journals, with a particular focus on the use of expert surveys. The third part describes our data and the key variables, and the fourth part presents our attempt to model the outcome of the RAE in respect of Politics and International Studies. The fifth part compares an especially parsimonious model's predictions of research quality with the actual results, and the sixth and final part concludes.

The 2008 Research Assessment Exercise

The purpose of the 2008 RAE, much like the purpose of previous RAEs, was to arrive at an authoritative judgement about the quality of research being undertaken in British Higher Education institutions. The 2008 RAE recognised 67 academic disciplines or 'units of assessment', ranging from Accounting and Finance to Theology, Divinity and Religious Studies. Each unit of assessment was evaluated by a separate sub-panel of experts, drawn from universities and other research organisations, and each of the 67 sub-panels was supervised by one of 15 main panels. The main panels brought together cognate disciplines and were intended to ensure a consistent approach in terms of working methods and criteria. Sub-panel 39 was charged with assessing

research in the field of Politics and International Studies and was overseen by main panel J, which was responsible for reviewing and endorsing sub-panel 39's criteria and working methods and confirming the 'quality profile' awarded to each submission. Sub-panel 39 was responsible for assessing in detail the various submissions made in the field of Politics and International Studies and making recommendations as to how each submission should be graded (RAE, 2006).

Overall, sub-panel 39 assessed 59 submissions from 58 institutions.[2] As a rule, submissions were usually organised on a departmental basis, and the bulk of each submission comprised the research outputs of individual members of staff. These outputs included monographs, journal articles, edited books, chapters in edited books, as well as web-sites, conference papers, consultancy reports and other outputs. Individuals were expected to submit four outputs, subject to special circumstances. In addition to individuals' outputs, participating institutions were also assessed on the basis of their research environment, defined by the number of their research students and research studentships, their research income, their research structure, their research strategy and their staffing policy. Finally, institutions were also assessed on the basis of their esteem, defined by the national and international recognition that related to the individuals submitted. Each submission's quality profile was determined largely by individual outputs, although a quarter of the profile would reflect the research environment and esteem.

At the heart of sub-panel 39's work was the detailed assessment of individual research outputs (see Table 1 below). Each output was read in detail by at least two members of the sub-panel and assigned one of five quality levels:

4*—quality that is world-leading in terms of originality, significance and rigour.

3*—quality that is internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence.

2*—quality that is recognised internationally in terms of originality, significance and rigour.

1*— quality that is recognised nationally in terms of originality, significance and rigour.

Unclassified—quality that falls below the standard of nationally recognised work or which does not meet the published definition of research for the purposes of this assessment. (RAE, 2006, pp. 19-20)

Each quality level was then carried forward to an overall institutional profile, so that a certain proportion of outputs were graded as 4*, another proportion as 3*, yet another proportion as 2*, and so on. Together, the grading of individual outputs would constitute 75 percent of a department's rating. Research environment (constituting 20 percent) and esteem (5 percent) were graded according to the same criteria and added to these scores. Finally, the scores were rounded into 5-point bands.

Two further features of the sub-panel's working procedures are worth highlighting, since they relate directly to the questions we address. First, sub-panel 39 made clear that all forms of research would be treated equally: it would 'not rank nor regard any particular *form* of output as of greater or lesser quality than another *per se*' (RAE, 2006, p. 31). In other words, monographs would not necessarily be regarded as being of a higher quality than journal articles or book chapters. This policy in 2008

was a departure from practice in 2001, when monographs were explicitly judged to be more important than peer-reviewed articles and other outputs (Butler and McAllister, 2009, p. 7). Second, sub-panel 39 explicitly rejected any use of pre-constructed rankings for evaluating monographs or journal articles: 'The sub-panel will not establish a list of the relative standing of publishers…. 'The sub-panel will not establish a list of the relative standing of journals' (RAE, 2006, p. 31). In both cases, it recognised that some types of research were published by less prominent or more specialist publishers and in less prominent or more specialist journals.

The RAE results were published in December 2008. Figure 1 reports the distribution of quality levels across submissions to sub-panel 39. The dominant rating was 2*. On average (at the institutional level), nearly 37 percent of departments' submissions were graded in this way. The least common rating (excluding unclassified) was 4*, which covered, on average, just 11 percent of departments' submissions. The Universities of Essex and Sheffield headed the league table: in both cases, 45 percent of their submissions were judged to be 4*. For a majority of British institutions (39), no more than a tenth of the overall submission was judged to be 4*. No department had half of its submissions judged to be 4*. Eleven departments failed to achieve a 4* rating for any proportion of their submission.

Figure 1 about here.

Reputations, rankings and perceptions

The 2008 RAE was the latest in a long line of authoritative audits of research at British Higher Education institutions. But political scientists, like members of every other discipline, also make judgements about the quality of individuals' research on a

day-to-day basis, for example, when looking to appoint or promote members of staff or when evaluating grant applications. Rightly or wrongly, we tend to look at candidates' past publications when undertaking such work and draw inferences about their potential on the basis of where they have previously published their research. First-hand knowledge of candidates' work can greatly alter such opinions, but our expectations, and thus our evaluations, are likely to be shaped by the knowledge of what and where an individual publishes.

Reputations are a 'noisy' signal of quality (McLean et al., 2009a, p. 19). Top presses and journals may publish bad research, and lesser presses and journals may publish good research. But there are reasons for supposing that a press's or journal's reputation has some relationship with the work it publishes. From an author's point of view, we know that certain presses are more selective in the manuscripts they publish and subject all manuscripts to peer review. From the same point of view, we also know that some journals' reviewers seem to employ more rigorous criteria when evaluating submissions than others. As the Chair of the 2001 RAE notes, 'with highly ranked journals there is likely to be competition for publication and to the extent to which competition is a quality filter, the journal name will provide some evidence of quality' (Weale, 2009, p. 46). From a reader's point of view, we know that some journals and publishers tend to print more significant and rigorously-conducted research than others.

There are two basic approaches for estimating the quality of a journal or press (Christensen and Sigelman, 1985). The first of these is the impact or 'bibliometric approach', which has been applied primarily to journals. It evaluates the quality of a given journal on the basis of how many citations its articles receive within a given database, for example, the ISI Web of Science or even Google Scholar (if we treat the

World Wide Web as a database). Journals may be ranked by the total number of citations of the articles they carry, the average number of citations per article, the impact of articles (the number of citations for a journal's 'average article' in the first two years after publication) or even by the number of articles exceeding a certain citation threshold and published in a defined period (see Plümper, 2007). Such metrics have, in turn, been used to rate political science departments (Hix, 2004). They have also been in attempts to explain earlier RAE outcomes (see, for example, Butler and McAllister, 2009).

The bibliometric approach for estimating quality is indirectly reputational, in that such citation metrics may be measuring scholars' behavioural response—in terms of submitting and citing work—to publishers' and journals' reputations. In this sense, Weale (2009, p. 41) is right to reject the suggestion that such metrics are 'objective'; after all, they ultimately reflect a large number judgements by individual scholars about whether to cite particular works or not. An alternative approach is to employ an expert survey and to measure perceived reputations directly. This method was first used to measure American political scientists' perceptions of journals (Giles and Wright, 1975; Giles et al., 1989) before being extended to Britain (Crewe and Norris, 1991). It has also been used to measure perceptions of publishers (Garand and Giles, 2011; Goodson et al., 1999).

The most recent expert survey of British political scientists' perceptions of journals, and the survey that informs our analysis of the 2008 RAE, was conducted by McLean et al. (2009a) as part of a comparative study of members of American, British and Canadian political science departments. The British component of the survey was based on the population of native political scientists as defined by the 2007 edition of the Political Studies Association's *Directory*. It was fielded in 2007.

Respondents were asked to rate the quality of each of 92 journals on a 0-10 scale (where 10 was outstanding), so long as they were familiar with it, and they were also invited to add and rate other journals that were not included in the 92. Using a previously established method premised on the basis of journal quality and visibility (Garand, 1990; Crewe and Norris, 1991), mean journal evaluations were combined with measures of familiarity to create a measure of perceived 'impact'. Clearly, any such evaluations will reflect subjective opinions: different people, with different specialisms, will rate different publishers and journals very differently. Partly for this reason, the impact measure has been criticised for being essentially arbitrary (Johnston, 2009, p. 53). Others, however, have praised it for its 'validity' (Weale, 2009, p. 46). One obvious advantage of the measure is that it is based on the evaluations of members of the discipline as a whole. The rankings broadly reflect what we—as producers and consumers of research—value.

According to the survey, the *British Journal of Political Science* emerged as the top journal, followed by the *American Political Science Review* and *Political Studies* (see Mclean et al., 2009a, pp. 27-29 for full details).[3] The importance of familiarity, in addition to perceived journal quality, meant that there were some surprises: *Government and Opposition*, for example, ranked above *Journal of Politics*. The results also tended to favour broad-based journals over highly-esteemed but more specialist journals. In this respect, McLean et al. (2009b, p. 89) are probably correct when they aver that 'this is a value position held by many political scientists'. Finally, the results indicated broad similarities between American and British perceptions of

journals: 7 of the top-10 journals in Britain were also in the top-10 of journals in the United States.[1]

This last finding is important when trying to identify the top scholarly publishers in Britain. In the absence of a similar British survey, we rely on Goodson et al.'s (1999) ranking of scholarly publishers among American political scientists to determine our list of top-10 publishers. This list includes Cambridge and Oxford University Presses, as well as eight leading American university presses.[4] The cross-national similarities in respect of journal rankings give us confidence in using these results in respect of books.

Data and variables

In order to examine whether journal and press reputations can be used as valid indicators of research quality we first coded every individual research output submitted to the 2008 RAE sub-panel 39.[5] We coded on the following basis: first, whether an output was a monograph,[6] an edited book, a journal article, a chapter in an edited volume, or some other output; second, whether a monograph, edited book or chapter appeared with a top university press (the top ten as defined in Goodson et al., 1999); and third, whether an article was published in a top journal (the top ten and top twenty as defined by McLean et al., 2009).

Table 1 reports a general breakdown of submitted outputs. Articles constituted the bulk of them (60 per cent in total), followed by books (21 per cent in total). Within

---

[1] To a certain extent then it does not make a great deal of difference whether we use the top 10 as voted for by British academics or whether we use the top 10 as voted internationally (by British, US and Canadian academics) since by and large they contain the same journals. However, the second 10 journals are much more distinctive, and in fact there is no overlap between the journals included on the British and the international list. Our analysis shows that the British list has a significant and positive association with 4 star ratings, whereas the international list does not have a significant association (and the direction of the relationship is in fact negative). This could be interpreted as evidence that the RAE panel was parochial rather than international. Or it could be interpreted as being more in line with British evaluations of quality than North American evaluations.

these broad classifications, 6 percent of outputs were articles in a top-ten journal, 13 per cent in a top-20 journal, and 4 per cent were books published with a top press.

Table 1 about here

We then aggregated the coded outputs by institution (59 altogether). Although we cannot tell how individual outputs were graded, it is possible to establish if there is a general relationship between how departments were graded overall, in terms of the proportion of output judged to be 4*, 3* and so on, and the types of output that they submitted. If journal reputation is not a valid indicator of an article's quality, then we would not expect there to be an association between the proportion of top journals in a department's submission and its proportion of 4* work in the RAE results. Similarly, if publisher reputation is not a valid indicator of quality, then we would not expect to see a relationship between the percentage of top university-press books submitted and the institutions' 4* ratings.

Table 2 reports the bivariate association between our variables and the 2008 RAE results. Statistically significant correlations ($p < 0.05$) are indicated in bold. At the top of the table are several variables that broadly relate to a department's 'research environment'. As Butler and McAllister (2009, p. 8) note, the size of a department is often regarded as an indicator of research activity in its own right, with larger departments taking advantage of their size to generate an active research culture that attracts further interest and resources. We examine the total number of outputs submitted by each department, which arguably provides a better indication of research activity than the number of people submitted, since many people did not submit the full quota of four outputs and were not as 'research active' as others. The mean size of

departmental submissions was 80 outputs, ranging from 4 at the University of Greenwich to 344 at the University of Oxford. Perhaps not surprisingly, big departments, as defined by the total number of submissions, were more likely to be awarded a higher 4* ranking than smaller departments. There was a similar albeit weaker relationship in respect of 3* work, and a negative relationship between the number of submissions and 2* and 1* work. Top politics departments in 2008 were likely to be big.

Table 2 about here

Another indicator of research environment is research income. We look at external research income over the RAE cycle from 2001 to 2007. The mean amount of research income over this period was just over £2 million, ranging from zero (as reported by four departments) to over £30 million (as reported by Kings College London). The median amount was £845,000. Because income was heavily skewed we also examine the log of research income. From the table we can see that departments which reported large levels of research income tended to get a higher 4* rating than departments which received only a little (or none at all). Interestingly, however, the correlation for log income and 3* rating was slightly stronger than the correlation between log income and 4* rating. This suggests that research income may not have affected the balance between departments 4* and 3* performance. Nevertheless, those departments that raised large amounts of research income were far less likely to have a large proportion of their output rated 1*.

Turning to research outputs, if we look at monographs in general, there were no significant correlations between the number of books as a proportion of a

department's submission and how its research was rated other than in respect of 1*

quality. Departments that submitted many books tended to receive a low 1* rating.

However, when we disaggregate books according to the publisher, two markedly

different patterns emerge. First, departments that submitted a large proportion of

books published with a top press did very much better, especially in respect of their

4* rating. Second, departments that submitted a large proportion of books published

by other presses only benefited in respect of their 2* rating.

     In respect of edited books, there were no significant correlations between how

many a department submitted and how its research was assessed. However, if we

disaggregate edited books according to whether or not they were published by a top

press, then a clearer pattern emerges. Top-press edited books had a significant and

positive correlation with 4* work, though the size of the correlation was much weaker

than for top-press monographs. By contrast, there were no significant correlations for

edited books published with other presses.

     The widespread view that book chapters constitute a weak submission appears

to be borne out by the data. Departments that submitted a large proportion of chapters

tended to get a higher 1* rating. But even here, we see differences according to where

chapters were published. There were no significant correlations between top-press

chapters and research quality; however, when chapters were published with other

presses, the correlation with 1* quality was very strong. Indeed, out of all the output

types, the submission of such chapters had the strongest correlation with 1* quality.

     We next turn to the importance of journals. If we look at all articles without

taking into account where they were published, there were no clear patterns in terms

of how research was judged. This finding almost certainly reflects the enormous

variability in journal quality. Because so many articles were submitted, we

disaggregated them into three bands according to McLean et al.'s (2009a) perceived impact rankings: whether they were published in a 'top 10' journal, a 'second 10' journal (number 11-20 in the rankings) or a journal outside the top 20. Although this division is somewhat arbitrary, it will suffice for testing the general proposition that journal reputations matter.

The ordering of articles in this way appears to have a degree of validity. The correlation between the proportion of top-10 journals and the percentage of work rated 4* was slightly stronger than the correlation between second-10 journals and 4* work. Similarly, the correlation between second-10 journals and work rated 3* was somewhat stronger than the correlation between top-10 journals and 3* work. Moreover, from the signs attached to the various coefficients in Table 2, we can see that there was an especially pronounced split between top-20 journals, which had positive correlations with 4* and 3* ratings and negative correlations with 2* and 1* ratings, and other journals, where the directions of the associations were reversed. Put simply, departments that submitted large numbers of top-20 journal articles as a proportion of their outputs tended to get higher 4* ratings than departments that submitted only a few. By contrast, departments that submitted many articles from journals outside the top 20 tended to get a higher 1* rating.

Overall, the correlations reported in Table 2 tell a clear story. Journal and publisher reputations appear to act as good indicators of research quality, as judged by the 2008 RAE. Departments that submitted large proportions of top-press books and top-journal articles tended to do very much better than departments that did not. These findings should not come as a surprise. Indeed, something would be seriously wrong if this was not the case. We are, as members of the discipline, the individuals who review articles and book manuscripts, and if our evaluations about what constitutes

good research were out of step with those of the RAE sub-panel, we would have to question whether its members were truly representative of the discipline as a whole.

However, two caveats are worth mentioning. First, we must be wary of committing the ecological fallacy. Because we are dealing with aggregate- rather than individual-level data, we cannot say with confidence that books published with top presses tended to be rated as 4* work, whereas books published with other presses tended to be rated 2*. We do not know how individual submissions were evaluated. Yet it is entirely plausible to suppose that departments that submitted a large proportion of top-press monographs tended to achieve a higher 4* rating because such outputs tended to be rewarded favourably. Second, when a department submits smaller proportions of one output, it is by necessity submitting more of another. To test the robustness of these correlations, we need to conduct multivariate analysis to establish whether, for example, how many top-10 articles a department submits still matters when we know how many top-press books it submits. It is to this task that we turn next.

Modelling 4* performance in the 2008 RAE

Simple statistical techniques can help us unpack the relative importance of different types of output in explaining institutional performance in the 2008 RAE. We focus our attention on the factors associated with 4* ratings, the key benchmark of research excellence for most members of the profession in Britain. The dependent variable is the percentage of an institutional submission judged to be 4*. Our full model (A), which is reported in the second column of Table 3, includes almost all the significant predictors from Table 2: the number of submissions; log income; top-press monographs; top-press edited books; top-10 journals; and second-10 journals. We also

include a dummy variable to control for whether or not a department had a member on the RAE panel, since analyses of previous RAE results have shown that departments tend to perform significantly better if they have a member of staff involved (Butler and McAllister, 2009). The results of the OLS regression analysis suggest that the model works well, explaining nearly 75 percent of the variance in 4* quality.

Table 3 about here

Dealing first with our measures of research environment, we can see that the coefficient for research income was not significant. This finding is in line with Butler and McAllister's (2009) analysis of the 2001 RAE, and suggests that the importance of research money in these types of evaluations is often over-stated. The coefficient for the total number of submissions, however, was highly significant. Even controlling for everything else, large departments that submitted many outputs did better, on average, than those institutions that submitted fewer. In terms of the strategic decisions that departments face over whether or not to submit a particular member of staff, this finding suggests there may be some advantaging in erring on the side of submitting more outputs rather than fewer.

Turning to outputs, top-press monographs emerged as having the most substantial impact on 4* evaluations. The coefficient was positive and highly significant: for every one-point increase in the proportion of a department's submission that comprised top-press monographs, a department's 4* rating improved by 1.6 points. Of the other variables in the model, the proportion of articles published in top-ten journals also had a significant and positive association with 4* work. However, the submission of edited books and second-10 journal articles did not have

a significant association with 4* outputs when we take into account the other variables in the model.

Finally, the coefficient for membership on the RAE sub-panel was significant and positive. Departments that had a member of staff on the sub-panel saw their 4* rating jump by 6.4 points, all other things being equal. This is a substantial effect, and one that does not have an easy interpretation. It may be that good departments were more likely to have members on the RAE panel and that the 'membership' term was picking up some unmeasured aspect of research environment or esteem, net of research income and department size. Alternatively, the effect could be a consequence of panel membership and the value of first-hand knowledge of its working methods and criteria. Panel members were potentially able to help their own departments establish more effective pre-submission procedures (see Weale, 2009, pp. 44-45). If this latter interpretation is valid, sub-panel 39 arguably failed to communicate its methods and criteria effectively to other departments.

Predictions

The results so far suggest that knowledge about where outputs were published, in particular the rankings of relevant journals and publishers, could predict a large proportion of the variance in the 2008 RAE outcome. To test this specific proposition, we drop all the non-significant variables from model A, as well as whether a department was represented on sub-panel 39, and examine how well departments fared according to their number of submissions and the proportion of their submissions that comprised top-press monographs and top-10 journal articles. The results are reported in the third column of Table 3. This simple model is able to account for 68 percent of the variance in departments' 4* ratings. It does not explain

everything—we would not expect it to given that all outputs were peer reviewed and so subject to measurement error—but it does explain a substantial amount.[7]

To look at where the main discrepancies occurred between our model and the actual RAE results, we estimated the predicted 4* ratings for each department by entering the relevant values for each department into the regression equation from model B. The results of this exercise are shown in Table 4. As noted, the RAE rounded institutional scores into 5-point bands. Overall, our model predicted the proportion of 4* quality awarded to a department within plus or minus 5 points of the actual proportion in 39 out of 59 cases. It thus shows a good degree of consistency with the RAE sub-panel's evaluations. Of the remaining 20 cases, our model predicted that nine institutions should have performed less well than they actually did, based solely on the number of their submissions, top-press monographs and top-10 articles, and eleven institutions should have performed somewhat better.

Table 4 about here

In general, the departments that did better than our model predicted were the departments that were represented on the RAE sub-panel. Indeed, the departments with the five largest residuals all had members on the panel (the Universities of Sheffield, Aberystwyth, Bradford, Essex and Durham). By contrast, the departments which did worse than expected were generally smaller departments, the notable exceptions being the Universities of Oxford (which did not have a member on the RAE sub-panel in 2008) and Birmingham (which did).

Discussion

The forthcoming REF, just like the last RAE, has eschewed hierarchies of output in its working procedures. The REF's main panel C, which covers sub-panel 21 Politics and International Studies, states in its published criteria: 'The sub-panels will assess all forms of output on an equal basis, with no preconception of quality attached to the form or medium of an output. No sub-panel will use journal impact factors or any hierarchy of journals in their assessment of outputs' (REF, 2012, p. 64). But just because these rankings will not be used as proxy indicators does not mean that they are not valid indicators of research quality (McLean et al., 2009a, p. 35) For the most part, they are.

Submissions in top journals and with top presses are strongly associated with research excellence. The reputation of the publication appears to make a greater difference than the type of publication. This effect may well be because outputs that appear in top journals and with top presses tend to go through a more competitive and thorough review process. In a sense then, this finding is reassuring. Both RAE judgements and reputational rankings are based on peer review; it is just that the latter is based on the opinions of many more people than the former. For example, the 2008 RAE sub-panel's assessment of politics was based on peer review by just 12 people. Yet, the reputations of journals are implicitly based on the judgements of the many thousands of scholars who submit and review papers, and on the editors who make decisions every day about the quality of the research submitted. Much the same could be said of the reputations of different scholarly publishers. McLean et al. (2009b, p. 91) make a strong case for not replicating the hard work of reviewers: 'Using a panel of scholars [to review work is] grossly inefficient compared to reliance on the perceived quality/prestige of the journal [and, we might add, publisher] in which the work appears'. We agree. We hope our findings may persuade colleagues and policy

makers that there are more cost-effective and efficient ways of evaluating research in a world of limited resources, though how these might be implemented in practice is of course a matter for further discussion.

Notes

---

[1] For example, in January 2009, the Higher Education Funding Council for England (HEFCE), which is responsible for distributing public money, decided to weight its funding allocation for 1*, 2*, 3* and 4* research as follows: 0:1:3:7. In February 2010, the relative weighting for 4* work was increased thus: 0:1:3:9. And in February 2012, the weighting was changed again, with 2* work losing its funding, thus: 0:0:1:3.

[2] The University of Sussex made two submissions to sub-panel 39, one centred on the work of its Department of International Relations, the other on the work of its Science and Technology Policy Research centre.

[3] The top twenty were: (1) *British Journal of Political Science*; (2) *American Political Science Review*; (3) *Political Studies*; (4) *International Organization*; (5) *American Journal of Political Science*; (6) *Comparative Politics*; (7) *World Politics*; (8) *European Journal of Political Research*; (9) *Comparative Political Studies*; (10) *European Journal of International Relations*; (11) *West European Politics*; (12) *Review of International Studies*; (13) *Government and Opposition*; (14) *Journal of Politics*; (15) *Journal of Common Market*; (16) *International Affairs*; (17) *International Studies Quarterly*; and (18) *Party Politics*; (19) *Journal of European Public Policy*; and (20) *Philosophy and Public Affairs*.

[4] The top ten were: (1) Cambridge University Press; (2) Cornell University Press; (3) Harvard University Press; (4) MIT Press; (5) Oxford University Press, including Clarendon Press; (6) Princeton University Press; (7) University of California Press; (8) University of Chicago Press; (9) University of Michigan Press; and (10) Yale University Press.

[5] This list is available on the RAE website at: http://www.rae.ac.uk/Results/. Last accessed on 3 July 2012.

---

[6] We did not distinguish between textbooks and research monographs.

[7] By way of contrast, Butler and McAllister's (2009, p. 11) citation-based model of the 2001 RAE results explained only 38 percent of the variance. However, the two models are not readily comparable, since they involve very different dependent variables.

References

Butler, L., and I. McAllister (2009) 'Metrics or peer review? Evaluating the 2001 UK research assessment exercise in political science', *Political Studies Review*, 7:1, 3-17.

Christensen, J.A., and L. Sigelman (1985) 'Accrediting knowledge: journal stature and citation impact in social science', *Social Science Quarterly*, 66:4, 964-975.

Crewe, I., and P. Norris (1991) 'British and American journal evaluation: divergence or convergence?' *PS: Political Science and Politics*, 24:3, 524-531.

Garand, J.C. (1990) 'An alternative interpretation of recent political science journal evaluations', *PS: Political Science and Politics*, 23:3, 448-451.

Garand, J.C., and M.W. Giles (2011) 'Ranking scholarly publishers in political science: an alternative approach', *PS: Political Science & Politics*, 44:2, 375-383.

Giles, M.W. and G.C. Wright, Jr (1975) 'Political scientists' evaluations of sixty-three journals, *PS: Political Science and Politics*, 8:3, 254-256.

Giles, M.W., F. Mizell and D. Patterson (1989) 'Political scientists' journal evaluations revisited', *PS: Political Science and Politics*, 22:3, 613-617.

Goodson, L.P., B. Dillman and A. Hira (1999) 'Ranking the presses: political scientists' evaluations of publisher quality', *PS: Political Science and Politics*, 32:2, 257-262.

Hix, S. (2004) 'A global ranking of political science departments', *Political Studies Review*, 2:3, 293-313.

Johnston, R. (2009) 'Where there are data … Quantifying the unquantifiable', *Political Studies Review*, 7:1, 50-62.

McLean, I., A. Blais, J.C. Garand and M.W. Giles (2009a) 'Comparative journal ratings: a survey report', *Political Studies Review*, 7:1, 18-38.

McLean, I., A. Blais, J.C. Garand and M.W. Giles (2009b) 'Authors' response to reviews', *Political Studies Review*, 7:1, 88-92.

Plümper, T. (2007) 'Academic heavy-weights: the "relevance" of political science journals', *European Political Science*, 6:1, 41-50.

RAE (2006) *RAE 2008 Panel criteria and working methods: Panel J*. Available at: http://www.rae.ac.uk/pubs/2006/01/docs/jall.pdf. Last accessed on 3 July 2012.

REF (2012) *Panel criteria and working methods*. Available at: http://www.ref.ac.uk/media/ref/content/pub/panelcriteriaandworkingmethods/01_12.pdf. Last accessed on 3 July 2012.

Russell, A. (2009) 'Retaining the peers: how peer review triumphs over league tables and faulty accounting in the assessment of political science research', *Political Studies Review*, 7:1, 63-72.

Weale, A. (2009) 'Metrics versus peer review?', *Political Studies Review*, 7:1, 39-49.

Figure 1: Distributions of 4*, 3*, 2* and 1* quality levels, 2008 RAE
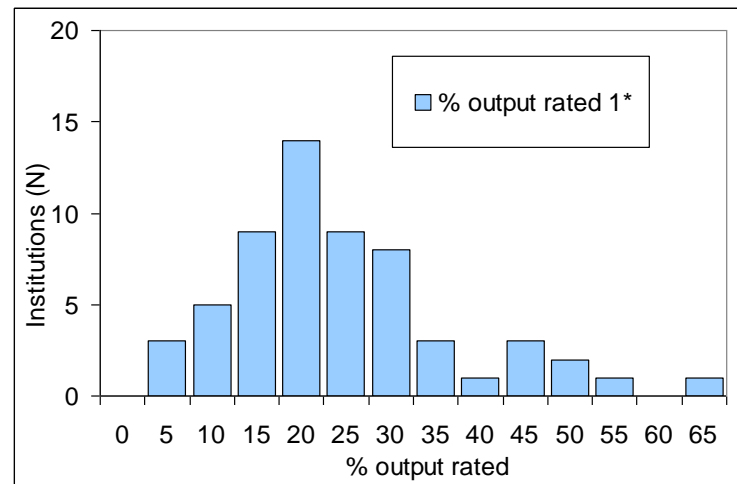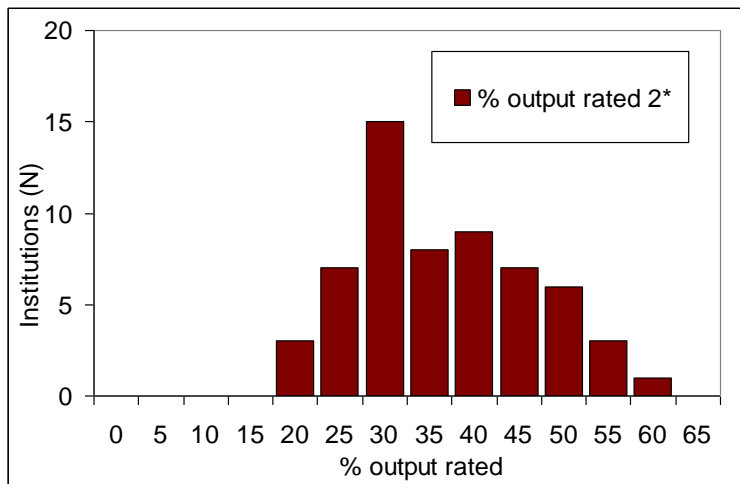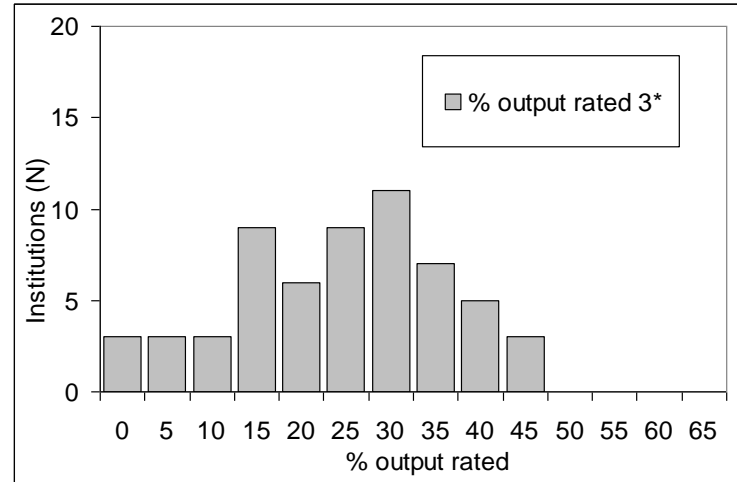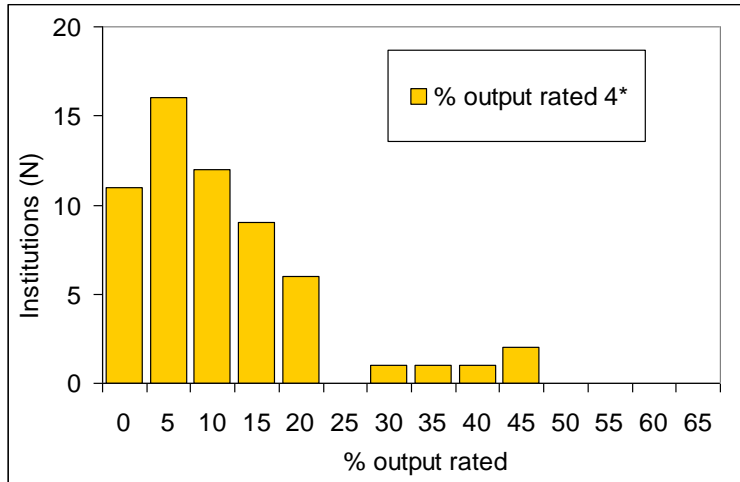
Table 1   Breakdown of coded submitted outputs

	Submissions	**% of all outputs**
Monographs	1,028	21.8
Top press	178	3.8
Other press	850	18.1
Edited books	95	2.0
Top press	20	0.4
Other press	75	1.6
Chapters in edited book	658	14
Top press	164	3.5
Other press	494	10.5
Journals	2,832	60.2
Top 10	258	5.5
Second 10	333	7.1
Top 20	591	12.6
Non-top 20	2,241	47.6
Other outputs	95	2
Total	4,708	100

Table 2: Correlation coefficients

	% rated 4*	% rated 3*	% rated 2*	% rated 1*
Submissions	**0.597**	**0.345**	**-0.280**	**-0.473**
Research income	**0.263**	0.178	-0.212	-0.196
Log income	**0.399**	**0.486**	**-0.122**	**-0.451**
Monographs	0.199	0.088	0.221	**-0.264**
Top press	**0.777**	**0.421**	**-0.367**	**-0.580**
Other press	-0.144	-0.097	**0.382**	-0.009
Edited books	0.239	0.066	-0.231	-0.051
Top press	0.**286**	0.181	-0.142	-0.219
Other press	0.160	-0.004	-0.218	0.042
Chapters in edited book	-0.118	-0.182	-0.182	0.**322**
Top press	0.235	0.194	-0.201	-0.165
Other press	-0.239	**-0.296**	-0.123	**0.445**
Journals	-0.086	0.043	0.058	-0.061
Top 10	**0.492**	**0.330**	-0.152	**-0.431**
Second 10	**0.462**	**0.570**	-0.172	**-0.543**
Top 20	**0.591**	**0.566**	-0.202	**-0.608**
Non-top 20	**-0.482**	**-0.345**	0.192	**0.356**
Other outputs	-0.049	0.071	-0.088	0.063

Notes: Statistically significant correlations (p < 0.05) are indicated in bold.

Table 3: Modelling the % of output rated 4* in the 2008 RAE (OLS)

	Model A	Model B
Submissions	**0.041 (0.017)**	**0.049 (0.017)**
Log income	0.180 (0.235)	
Top-press monographs	**1.633 (0.373)**	**2.007 (0.368)**
Top-press edited book	1.309 (0.886)	
Top-10 journals	**0.447 (0.195)**	**0.471 (0.204)**
Second-10 journals	-0.101 (0.208)	
Department member on panel	**6.375 (2.399)**	
Constant	-2.349 (2.797)	-0.508 (1.549)
R2	0.741	0.680
Adj. R2	0.706	0.662

Notes: Cell entries are coefficients (standard errors). Statistically significant

correlations (p < 0.05) are indicated in bold.

Table 4: Using 'model B' to predict the 2008 RAE

	Under-prediction			Over-prediction	
Institution	**Actual**	**Predicted**	**Institution**	**Actual**	**Predicted**
Sheffield	45	25	Strathclyde	0	16
Aberystwyth	40	29	Royal Holloway	5	16
Bradford	15	5	Aberdeen	5	15
Essex	45	36	Oxford	35	43
Durham	15	6	Liverpool	0	7
Dundee	10	1	Kent	0	7
Newcastle	15	7	Edinburgh	10	17
Sussex A	20	12	Robert Gordon	0	6
Ulster	10	4	Queen Mary	5	11
			Southampton	5	11
			Birmingham	5	11