

Colman, A. M., Dhillon, D., & Coulthard, B. (1995). A bibliometric evaluation of the research performance of British university politics departments: Publications in leading journals. *Scientometrics*, 32, 49-66.

A Bibliometric Evaluation of the Research Performance of British University Politics Departments: Publications in Leading Journals

**ANDREW M. COLMAN, DEBRA DHILLON
AND BEN COULTHARD***

ABSTRACT

The research performance of 41 British university politics departments was evaluated through an analysis of articles published between 1987 and 1992 in nine European politics journals with the highest citation impact factors. Annual performance scores were obtained by dividing each department's number of publications in these journals in each year (departmental productivity) by the corresponding departmental size. These scores were summed to obtain a research performance score for each department over the period of assessment. They correlate significantly with research performance scores from two previous studies using different methodologies: Crewe's per capita simple publication count for the years 1978 to 1984, and the Universities Funding Council's research selectivity ratings covering the years 1989 to 1992.

Since the mid 1980s, the funding of British universities has been determined to an increasing extent by evaluations of departmental research performance. Rightly or wrongly, there is a momentum towards the use of such evaluations for the purposes of resource allocation. Although research selectivity is now an established aspect of higher education policy in Britain, there is controversy about how departmental research performance ought to be measured. The problem is exacerbated by the dichotomous roles of British universities – as teaching institutions and as centres of innovation and research. Most assessment exercises to date have focused on research performance, although teaching performance has also received some attention.

The concept of research performance requires clarification, especially in the light of the widespread confusion of research performance with research productivity. To rate departments according to research productivity – that is, according to their relative research outputs – would be to disregard input factors such as department size, research grant income, equipment grants and other recurrent income, and technical, secretarial, and administrative staff. Measures of productivity that ignore input factors give an unfair advantage to larger and better funded departments. Research performance, on the other hand, is a measure of output per unit of input and therefore reflects the efficiency of a department's research productivity. Although the most important input factors can be identified without too much difficulty, the assessment of research performance is not as straightforward as it may at first appear, for it should ideally take account of both the quantity and the quality of the research output of departments.

Various attempts have been made, with varying degrees of success, to measure departmental research performance. In 1986 the University Grants Committee (UGC), which

was later transmogrified into the Universities Funding Council (UFC) and then the Higher Education Funding Council (HEFC), conducted the first of a series of research selectivity exercises across the range of British university departments from accountancy to zoology. This exercise involved the comparative evaluation of research performance, partly on the basis of a limited number of publications submitted by each department in each subject area. It became the target of scathing attacks from critics, including Bentham¹ on behalf of geography and Gillett² on behalf of psychology, the essential thrust of the criticisms being applicable to all subject areas. The critics claimed that the UGC ratings were unrelated to either the quality or the quantity of research produced by departments, and they suggested that input factors appeared to have been used in a wholly inappropriate way as performance indicators. Considerable weight was accorded, in particular, to research grant income as a performance indicator. The rationale rested on an assumption that these factors are somehow a reflection of research performance. As both Gillett and Bentham demonstrated, however, this assumption was not grounded in empirical fact. As a consequence, Gillett concluded that the UGC research selectivity exercise was based on ‘patently unsound material’ and that it possessed ‘approximately zero validity’.³ Colman, Garner and Jolly⁴ argued that the use of research grant income as a criterion for distributing resources was a grotesquely inequitable application of the Matthew Principle: “For unto every one that hath shall be given, and he shall have abundance; but from him that hath not shall be taken away even that which he hath” (Matt. 25: 29).

The UGC research selectivity exercise of 1986 and the ones that followed it are examples – although perhaps not very flattering ones – of peer review, which is a generic term covering all forms of assessment of academic departments (or individuals, or manuscripts submitted for publication, or research grant applications) by colleagues assumed to be knowledgeable in the relevant fields of study. The application of peer review to the evaluation of university departments has a long pedigree in the United States. As long ago as 1964 Somit and Tanenhaus⁵ attempted to measure the strength of political science doctoral programs in the United States by drawing a random sample of more than 400 members of the American Political Science Association to serve as peer reviewers. A similar ‘reputational’ approach was used by Roose and Andersen⁶ in their assessment of American graduate training programs in psychology.

These survey methods may be more satisfactory than judgements made by small panels of appointed experts, because they are more likely to be representative of the opinions of the entire profession rather than of a select few. Nevertheless, like all impressionistic peer review assessments of departmental research performance, they are seriously flawed. Gillett⁷ listed no fewer than nine problems associated with the technique, and Johnes⁸ mentioned several more. One of the most important arises from the fact that academic peers tend to be specialists, with the consequence that they are usually comparatively ignorant of sub-fields that lie outside their own relatively narrow spheres of expertise. As a consequence, their competence to pass judgement on research in other sub-fields, or a fortiori on the entire range of a department’s research output, is severely limited. Additionally, certain high-profile specialisms attract greater numbers of researchers and considerably more attention than others, and this can potentially result in pronounced biases against less prominent but no less worthy specialisms. It hardly needs to be said that peer evaluators, however eminent or competent they may be, are only human and prone to the same prejudices and misperceptions as other people. Any reputation-based method of departmental evaluation is therefore liable to distortions arising from prevalent but erroneous or outdated assumptions and stereotypes.

Whatever weight one accords to these problems, peer review cannot avoid one particular obstacle: it is usually based at least partly on departmental reputations, and reputational approaches rest on fundamentally subjective judgements. Consequently, even if the reputational ratings are highly representative of professional opinion in the relevant subject areas, this does not mean that those opinions are necessarily valid. Assessments of departmental reputation cannot be empirically validated because they are not grounded in objective performance data. Peer review of academic departments therefore lacks one crucial component of any assessment procedure, namely a 'reality check'.⁹

If subjective peer review is inadequate, as the evidence suggests that it is, then some more objective method is required. Citation analysis has often been used as an objective measure, with researchers in various fields among its exponents.¹⁰ When applied to the evaluation of departmental research performance, citation analysis involves counting the number of times publications by members of each department have been cited by their peers in journal articles. The rationale for its use is relatively simple: the quality of a department's or a researcher's work is reflected in the attention it receives subsequently from researchers. Citation data are relatively accessible from the Science Citation Index, the Social Sciences Citation Index, and the Arts and Humanities Citation Index, or from electronic versions of these databases such as the Bath Information and Documentation Service (BIDS). Citation counting is certainly more objective than peer review. Closer inspection, however, reveals potentially serious inadequacies, not least in the ambiguous motivations behind article citation.¹¹ This concern is shared by Johnes¹² and Chapman,¹³ both of whom have criticized citation analysis. Chapman, in particular, listed no fewer than 25 shortcomings, biases, deficiencies, and limitations which add up to a major indictment of citation counting as a method of measuring research performance.

Some of the more prominent problems with citation counting are the following. All sources of citations are weighted equally although some are clearly of greater quality or worth than others; only first authors receive credit when multi-authored articles are cited, thereby overestimating their research contributions whilst neglecting the contributions of co-authors; citation indexes often fail to distinguish between authors with the same surname; and individuals with commonly misspelt names or whose surnames change (such as married women) may find themselves credited with spuriously low citation counts. Even if these pitfalls could be avoided, serious problems would remain. Newcomers, whether individuals or departments, would be disadvantaged because citations typically take several years to accumulate; authors would continue to receive massively inflated credit for single publications that happen to report much-used tests or procedures, or that contain flawed arguments that become the butt of repeated attacks in the literature; self-citation, although often entirely innocent and justified, would continue to be practised in order to enhance scholastic rankings; and the related problem of mutual citation conspiracies, whereby groups of researchers either formally and explicitly or informally and implicitly cite each other preferentially for mutual benefit, would persist. Inadequacies such as these cannot easily be overcome. Self-citations are often excluded from citation counts, but this can unjustly discriminate against proper and legitimate self-citation, and mutual citation conspiracies are almost undetectable and almost certainly unprovable.

Rushton¹⁴ has defended the validity of citation counts by offering empirical evidence that some of these deficiencies may not be as serious as initially thought, especially when citation analysis is applied to individuals. Although each deficiency may not be very serious on its own, taken together they may none the less combine to cause quite serious distortions,

especially when applied to the assessment of departmental research performance. Rushton contends that gross distortions cannot be very frequent, because the correlation between scientific eminence determined by citation counts and by peer assessment has been shown to be quite high.¹⁵ But the correlation between impressionistic peer assessments and citation counts is not necessarily flattering to citation counts, bearing in mind the mounting criticisms of impressionistic peer assessments. In any event, some correlation between these two assessment procedures is inevitable because they are alike in many ways. Citation analysis, although objective, is grounded in citation behaviours, which continue to be poorly understood. In short, the claim that aggregate citation counts, even when corrected for by input measures, accurately reflect the research performance of academic departments, is not entirely convincing.

What then are the alternatives to peer review and citation analysis? The most straightforward alternative would be to conduct a simple publication count of the output of the members of departments and then to correct these raw productivity figures for department size. Robey¹⁶ developed such a measure of research performance, based on the number of publications by members of university politics departments from 1968 to 1977 in the United States. Similarly, Crewe¹⁷ has conducted an almost exhaustive study of the publication records of British university departments of politics between 1978 and 1984. Apart from the obvious difficulty of locating all of a department's publications, the problem with simple publication counts is that the sheer numbers of publications, even after correcting for departmental size, may be too crude an indicator of departments' research performance, because they fail to take the quality of the publications into account. Accordingly, several researchers have recently taken this approach one step further. They have attempted to introduce a degree of quality control by restricting publication counts to articles in restricted sets of 'high quality' journals. The assumption behind this approach is that although the articles still differ qualitatively, they must all be above a reasonably high threshold of quality to have been accepted for publication in the restricted journal set.

Evaluation of departmental research performance using quality-controlled restricted publication counts was pioneered in the United States. It was applied initially to university departments of psychology¹⁸ but has also been applied to university departments of politics.¹⁹ Its use in Britain has been limited to two studies of university psychology departments.²⁰ The research reported in this article represents a modest attempt to extend the technique to British university departments of politics.

The technique is not without its critics, but it does avoid many of the problems associated with other methods of evaluating departmental research performance. Unlike peer review, it does not rely on any unrealistic assumption of omniscience on the part of panel members, and it is entirely objective. Unlike citation counting, all authors receive appropriate credit, misattributions and other errors are virtually eliminated, newcomers are not unfairly disadvantaged, and self-citations, mutual citation conspiracies and the like do not distort the results. Its strength lies in the fact that each article is located at source, counted only once, and then checked carefully for departmental affiliations. For these reasons, the use of restricted publication counts for evaluating departmental research performance has definite advantages over other methods, and it can be implemented relatively accurately.

The study described below was based on a count of publications by members of British university departments of politics in nine leading European journals of politics during the years 1987–1992 inclusive. A six-year period was thought to be long enough to iron out short-term fluctuations and to provide a reasonable sample of departmental research output,

but not so long as to include data of merely historical interest.

METHOD

Institutions and Departments

The units of analysis of the present study were British university departments of politics. In most universities there is a department in which politics is taught and researched, and in the majority of cases it is easy to identify. It is usually called a department of politics, political theory, political studies, government, political science, international relations, or European studies. But politics is a diverse discipline with a correspondingly diverse organizational structure, and there were several instances in which it was unclear whether departments should be included or excluded, especially where their names or structures had changed during the period covered by the survey. In most cases, however, where there was just one department with one of the names listed above, or with a combination of the above names (for example, Aberdeen University's department of Politics and International Relations), there was no ambiguity about inclusion or exclusion.²¹

The 41 universities finally chosen for inclusion are shown in Table 1. Politics departments in new universities had regrettably to be excluded, because staff lists were publically available only from 1992. As many university departments of politics as possible were included in the study, but departments were excluded if their existence did not span the entire period of the study (e.g., University of Manchester Institute of Science and Technology), if they or their staff lists were not readily identifiable (e.g., King's College London and University College of Cardiff), or if they were too wide-ranging and diffuse to enable fair comparisons with more conventional politics departments (Bath, Cambridge, and East Anglia Universities). Departments in the latter category, whilst undeniably teaching and researching politics, do so under large, broadly named faculties, in which specific politics staff cannot easily be identified. Finally, there were a few politics departments that also covered other allied but separate disciplines (Dundee University's Department of Political Science and Social Policy; Liverpool University's Department of Politics and Communication Studies; Birkbeck College London's Department of Politics and Sociology; Salford University's Department of Politics and Contemporary History, and Ulster University's Department of Philosophy and Politics). In these cases, it was generally only the politics staff who contributed to the journals that came under scrutiny in this study, and the research performance scores of these departments should be interpreted in the light of that fact.

Restricted Journal Set

In order to implement a quality check of articles, it was necessary to define a restricted set of indubitably high-quality journals on which to base the publication count. Previous studies that have used this methodology have sometimes defined their restricted journal sets as those published by the leading academic society or professional association of the discipline, such as the American Psychological Association²² or the British Psychological Society.²³ In other cases restricted journal sets have been defined as a select number of supposedly 'reputable' journals.²⁴ Neither of these approaches is entirely satisfactory; they are either arbitrary or subjective and do not guarantee that the restricted journal sets consist of the journals of the highest possible quality. To optimize the choice of journals for the research described below, the restricted set was defined as the European journals of politics with the highest mean citations per published article, otherwise known as citation impact. A journal's citation

impact factor is generally regarded as a rough index of its prestige and scholarly influence in the discipline.²⁵ The citation impact factors of the journals in this study were obtained from the Journal Citation Reports, published by the Institute for Scientific Information.²⁶

Only European journals were included, because publications by British researchers in American and other non-European journals are too sparse to form the basis of meaningful or reliable publication counts. The 10 European journals of politics with the highest citation impact factors were, in descending order, the New Left Review, Problems of Communism, Public Choice, the British Journal of Political Science, Policy and Politics, Studies in Comparative Communism, the European Journal of Political Research, Political Studies, Politics, and Electoral Studies. The New Left Review was excluded because it does not provide institutional affiliations or addresses of its authors and is not, in any event, a scholarly journal in the same sense as the others. The restricted publication count was based on the remaining nine journals. Two of them, namely Problems of Communism and Studies in Comparative Communism, are relatively specialized, and their inclusion may be thought to favour departments specializing in studies of communism, but this problem turned out to be inconsequential because there were very few contributions to either journal by members of British university politics departments.

Rating Procedure

Each issue of each of the nine selected journals was searched meticulously for the years 1987 to 1992 inclusive for articles written by members of university departments of politics in the United Kingdom.

Multiple authorships were credited according to the following formula, which was borrowed from Howard, Cole, and Maxwell:²⁷

$$\text{credit}_i = (1.5^{n-i}) / (\sum 1.5^{i-1}),$$

where i is the specified author's ordinal position in the article's byline, n is the total number of authors, and the summation is over i from $i = 1$ to n . The formula is designed to assign appropriate credit to authors for their relative contributions according to the rank order of their authorship positions. According to the formula, an article by a single author scores one unit for that author's department, an article by two authors scores 0.6 and 0.4 units for the first and second authors' departments respectively, a triple-authored paper scores 0.47, 0.32, and 0.21 respectively, and so on.

An attempt was made to include only genuine articles in the publication count, but some journals were found to contain ambiguous items. In these instances common-sense judgements about inclusion or exclusion were made, and in practice the ambiguities were usually resolved with relative ease. For example, review articles that were written in the style and form of regular articles were included, but review articles in Studies in Comparative Communism, which resembled book reviews rather than genuine articles, were excluded from the count. Similarly, research notes that possessed many of the features of genuine articles, such as abstracts and lists of references, were included, but no credit was assigned for book reviews or book notes, stand-alone abstracts, letters, editorial notes, or other ephemera.

In order to calculate a department's research performance as distinct from its research output or productivity, a measure of departmental size was needed. This was determined for each of the years 1987 to 1992 inclusive from information published in the annual Commonwealth Universities Yearbook (CUY) from 1988 to 1993 (the CUY lists the staff of departments in the previous academic year). Fluctuations in departmental size were therefore

properly controlled for. Emeritus professors, visiting and honorary members of staff, demonstrators, research assistants and research fellows were not counted, but research professors were. Temporary staff were counted in the same way as tenured staff. In calculating departmental size, each full-time staff member was counted as one unit, and each part-time staff member or person holding appointments in two departments was counted as half a unit.

For each of the years 1987 to 1992, the author credit score for each department was divided by the corresponding department size score, giving an annual research performance score. These annual performance scores were then summed to give an overall performance score for the entire six-year period. Departmental sizes and hence performance scores for institutions containing several departments of politics (e.g., Birmingham University) were combined to yield a single research performance score for each institution.

RESULTS

Table 1 about here

Departmental research performance scores for the years 1987 to 1992 inclusive are shown in Table 1 in descending rank order. Also shown in Table 1 are per capita simple publication scores as calculated by Crewe²⁸ for the years 1978 to 1984 inclusive, with corresponding ranks, and UFC research evaluation ratings for the period 1989–1992. The research performance scores could not be calculated for three departments (Bath, Cambridge, and East Anglia Universities) for reasons previously discussed. Nevertheless, their total numbers of publications in the quality-controlled restricted journal set over the six-year time period are shown in Table 1, together with their per capita simple publication scores as calculated by Crewe, and the UFC rating for East Anglia University is also shown. In addition, for the sake of completeness, two institutions that were not included in our study, but were assessed by both Crewe and the UFC, are also listed – King’s College London and the Institute of Commonwealth Studies.

With regard to the ‘typical’ department of politics, Table 1 reveals that the mean research performance scores for the 41 departments was 0.28 ($SD = 0.22$), the mean per capita simple publication score was 8.00 ($SD = 3.88$), and the mean UFC rating was 3.54 ($SD = 0.90$). The leading department of politics according to research performance was Strathclyde University, with a performance score of 1.10, more than 3.7 standard deviations above the mean and approaching twice the score of the next-placed department, Essex University, which had a performance score of 0.65.

Pearson’s product-moment correlation coefficients were calculated between the performance scores and the per capita simple publication scores (excluding Ulster and Stirling Universities, for which per capita simple per capita publication scores were not reported by Crewe²⁹), between the performance scores and the UFC ratings (excluding Salford and Loughborough Universities, which were assessed separately by a different UFC subject panel in 1992: European Studies, rather than Politics and International Studies), and between the per capita simple publication scores and the UFC ratings, for which scores for Salford, Loughborough, and Ulster Universities were omitted but scores for King’s College London and the Institute of Commonwealth Studies were included. The highest correlation is between the research performance scores and the per capita simple publication scores: $r(37) = .52$ ($p < .01$). Research performance scores correlated significantly with UFC ratings, $r(37) =$

.49 ($p < .01$), and per capita simple publication scores correlated significantly with UFC ratings, $r(38) = .47$ ($p < .01$). These three correlations are fairly close to one another, and the differences between them are all statistically non-significant.

DISCUSSION

The Universities Funding Council awarded seven departments of politics maximum ratings of 5, supposedly indicating that these departments, and they alone, produced 'research quality that equates to attainable levels of international excellence in some sub-areas of activity and to attainable levels of national excellence in virtually all others'.³⁰ They were Strathclyde, Essex, Hull, Glasgow, and Manchester, Universities, the London School of Economics and Political Science, and King's College London.

The research performance scores reported above show a considerable range of values from department to department, with Strathclyde University emerging as the clear and undisputed leader – its performance score was about 70 per cent higher than that of Essex University, the second-placed department. The top UFC ratings of Strathclyde, Essex, and Hull Universities are clearly supported by our research performance findings: they were among the top four departments in terms of research performance (the fourth, Salford University, was rated 4 by the European Studies subject panel of the UFC). Conversely, eight of the top ten departments in terms of research performance received UFC ratings of 4 or 5; the two exceptions were Aberdeen and Bristol Universities, which were rated only 3 by the UFC. This concordance is reflected in the significant correlation obtained between research performance and UFC ratings ($r = .49$).

The UFC ratings are only partly corroborated by the evidence of research performance, however, and several discrepancies merit comment. Some departments that were not regarded by the UFC as among the very best (notably Salford, Aberdeen, and Newcastle upon Tyne Universities), were ahead of other better-known and more prestigious rivals (e.g., Oxford University) in research performance. Furthermore, other top-rated departments according to the UFC were Glasgow University, Manchester University, and the London School of Economics, but according to research performance scores, however, they ranked 10th, 16th and 17th respectively out of 41 – high ranks, but in the latter two cases by no means exceptional ones.

Once again, the Universities Funding Council may have mistaken size for performance, because both Manchester University and the London School of Economics had large departments of politics (seventh and third largest respectively). Other anomalies provide further evidence for this apparent 'big is beautiful' bias in the UFC ratings. Swansea, Bradford, and Aberystwyth Universities, which all had much larger politics departments than average, were given ratings of 4, indicating 'research quality that equates to attainable levels of national excellence in virtually all sub-areas of activity, possibly showing some evidence of international excellence, or to international level in some and at least national level in a majority'.³¹ In stark contrast to this, our findings rank them 35th, 36th, and joint bottom respectively, with very low research performance scores – Aberystwyth's department published virtually no articles in high citation impact European journals between 1987 and 1992. (It is only fair to point out that all three of these departments would probably have fared better had some of the journals in the restricted set been devoted to international relations, because they all have strengths in that area.) If there was a confusion of size with performance on the part of the UFC, it may not have been deliberate or even conscious; it may have arisen through the so-called availability heuristic, a well established cognitive

judgement strategy that causes people to overestimate the frequencies or probabilities of memorable or notable events when they can easily recall instances of them.³² Peer reviewers are, of course, likely to be able to recall more achievements associated with large departments than small ones, other things being equal, and if this effect is not controlled for it is very likely to lead to inflated subjective impressions of large departments.

The findings of this study provide evidence to support several tentative conclusions concerning the Universities Funding Council's research selectivity exercise.³³ In contrast to the previous research selectivity exercises, which were notoriously lacking in validity,³⁴ the 1992 UFC exercise does appear to demonstrate modest validity. Cause for concern remains, however, in the existence of serious discrepancies between UFC ratings and research performance scores. The nature of these discrepancies suggest that the UFC ratings lacked adequate control for input measures such as department size.

If the Universities Funding Council 1992 ratings, which were based on peer review, left something to be desired, how did Crewe's³⁵ simple publication count accord with the findings of our quality-controlled restricted publication count? One might expect quite a high correlation between the two, because they were based on partly similar and partly identical performance indicators. As expected, our results did correlate significantly with Crewe's: the correlation was $r = .52$. Direct comparison is complicated by the fact that Crewe's data were based on publications during the years 1978 to 1984, whereas ours were based on publications during 1987 to 1992. Discrepancies between the results of the two studies may therefore be due not only to methodological differences but also to real improvements and declines in departmental research performance over time. Had the time span of the two studies been identical, the correlation would undoubtedly have been higher, although it is impossible to say how much higher they would have been.

Despite this difficulty, the obtained correlation is quite high, slightly higher, in fact, than that between research performance and UFC ratings, although this difference is not significant. But Strathclyde University's overwhelming lead according to our quality-controlled restricted publication count was not reflected in Crewe's simple publication count, according to which it ranked in ninth place. Similarly, several departments such as Salford, Newcastle upon Tyne, Keele, Manchester, Leicester, and Birmingham Universities appear to have been grotesquely underrated by Crewe's simple publication count. Unless these departments underwent massive improvements in the space of a decade, which is certainly possible, either Crewe's or our index of research performance is wanting. Perhaps the most likely explanation for the discrepancy is the failure of simple publication counts to control for the quality of publications. Departments such as these, although they evidently did not have spectacularly high overall publication rates, regularly produced articles in the leading journals. Our index of research performance, because it took account not only of quantity but also indirectly of quality of research output, probably gives a truer picture of departmental research performance than can be achieved by a simple publication count.

The index of research performance used in this article – the average number of publications per capita between 1987 and 1992 inclusive in nine high citation impact European journals – has palpable advantages over other assessment measures such as peer review, citation analysis, and simple publication counts. It is, none the less, vulnerable to various methodological criticisms that are worth mentioning. A common criticism is that restricting the publication count to a specified set of leading journals does not control for the residual qualitative and quantitative differences between articles. Quantitative differences could be controlled for by counting published pages rather than published articles, but

journals differ in their formats and typefaces, and only a count of published words would overcome such problems. Qualitative differences are even more problematical. The goal of fine-grained quality control is probably unattainable, but restricting the count to the most cited journals, all of which have very high rejection rates, provides some protection against the inclusion of poor quality articles. Failures of the peer review procedure customarily used to decide the fate of manuscripts submitted for publication are more likely to result in false negatives (the rejection of high quality manuscripts) than false positives (the acceptance of genuinely weak submissions).

A second criticism was levelled by Johnes.³⁶ When a performance measure is based on publications in a restricted set of journals, the choice of the journals, and indeed the number of journals chosen, may have a profound influence on subsequent departmental rankings. The problem is essentially that consideration of only a small number of journals tends to lead to the inclusion of journals specializing in some fields but not in others. Thus, departments specializing in, say, international relations (e.g., Swansea, Bradford, and Aberystwyth Universities), local government studies (e.g., Birmingham University), or Southern African studies (e.g., Leicester University), may justifiably feel that their research performance scores do not adequately represent their true research strengths because none of the journals in the restricted set happen to specialize in these sub-fields. This is an intractable problem. The publication count must necessarily be based on a limited number of journals in order to control for qualitative differences between publications. The parallel requirement of fair representation for all specialist areas cannot possibly be met in its entirety: some departments will inevitably be advantaged and others disadvantaged by any set of journals that might be chosen. It is worth bearing in mind that any performance indicator, no matter how carefully it is devised, is bound to favour some departments and to disfavour others.

Perhaps the most obvious criticism of the quality-controlled restricted publication count approach to assessing research performance is that it bases its measurement solely on articles published in journals, to the exclusion of other research outlets such as books and chapters in edited monographs. This is necessary for two reasons. The first is the sheer quantity of published material: locating all of the books and chapters written by each staff member of each department over a six-year time period would be a monstrously difficult and error-prone task. Although some researchers such as Crewe³⁷ have attempted it, it is impossible to verify that the resulting data were fully comprehensive. Second, even if one were able to construct an exhaustive research record for every staff member of every department, how would one weight the various publications? If the qualitative and quantitative differences between articles seem to present difficult problems, they are dwarfed by the enormous differences between books, chapters in edited monographs, and articles in all kinds of periodicals. Various rules for weighting publications have been suggested,³⁸ but they are inevitably arbitrary and conjectural.

Nederhof³⁹ has criticized quality-controlled restricted publication counts for precisely their failure to overcome this problem, suggesting that since they focus on only a small percentage of published research, they are inappropriate measures of research performance. As Howard, Cole, and Maxwell have pointed out, however, the issue is not one of percentages but of generalizability: 'estimates based on even a minuscule proportion of data might in fact be eminently generalizable'.⁴⁰ How well estimates of research performance based on quality-controlled restricted publication counts correlate with estimates based on more comprehensive indices is ultimately an empirical question, to which the findings reported in this article make a small contribution.

NOTES

*Andrew M. Colman is Reader in Psychology, Department of Psychology, University of Leicester, Leicester LE1 7RH, England. Debra Dhillon and Ben Coulthard are graduates from the same department. We are grateful to Ivor Crewe for helpful advice in the planning of this research, and to R. L. Borthwick for valuable critical comments on an earlier draft of this article.

¹ G. Bentham, 'An Evaluation of the UGC's Ratings of the Research of British University Geography Departments', *Area*, 19 (1987), 147-154.

² R. T. Gillett, 'Serious Anomalies in the UGC Comparative Evaluation of the Research Performance of Psychology Departments', *Bulletin of the British Psychological Society*, 40 (1987), 42-49.

³ Gillett, 'Serious Anomalies in the UGC Comparative Evaluation of the Research Performance of Psychology Departments', at p. 42.

⁴ A. M. Colman, A. B. Garner and S. Jolly, 'Research Performance of United Kingdom Psychology Departments', *Studies in Higher Education*, 17 (1992), 97-104.

⁵ A. Somit and J. Tanenhaus, *American Political Science: Profile of a Discipline* (New York: Atherton Press, 1964).

⁶ K. D. Roose and C. J. Andersen, *A Rating of Graduate Programs* (Washington, DC: American Council on Education, 1970).

⁷ R. T. Gillett, 'Research Performance Indicators Based on Peer Review: A Critical Analysis', *Higher Education Quarterly*, 43 (1989), 20-38.

⁸ G. Johnes, 'Research Performance Indicators in the University Sector', *Higher Education Quarterly*, 42 (1988), 54-71.

⁹ Gillett, 'Research Performance Indicators Based on Peer Review: A Critical Analysis', at p. 27.

¹⁰ See, for example, N. S. Endler, J. P. Rushton and H. L. Roediger, III, 'Productivity and Scholarly Impact (Citations) of British, Canadian, and US Departments of Psychology', *American Psychologist*, 33 (1975), 1064-1082; E. Garfield, 'Is Citation Analysis a Legitimate Analysis Tool?', *Scientometrics*, 4 (1979), 359-375; E. Garfield, 'Mapping the World of Science: Is Citation Analysis a Legitimate Evaluation Tool?', in D. N. Jackson and J. P. Rushton, eds, *Scientific Excellence: Origins and Assessment* (Beverly Hills, CA: Sage, 1987); H. Klingemann, 'Ranking the Graduate Departments in the 1980s: Toward Objective Indicators', *PS: Political Science and Politics*, 19 (1986), 651-661; J. L. Matson, W. D. Gouvier and R. Manikam, 'Publication Counts and Scholastic Productivity: Comment on Howard, Cole, and Maxwell', *American Psychologist*, 44 (1989), 737-739; J. P. Rushton, 'Evaluating Research Eminence in Psychology: The Construct Validity of Citation Counts', *Bulletin of the British Psychological Society*, 37 (1984), 33-36; J. P. Rushton, 'A Ten-year Scientometric Revisit of British Psychology Departments', *The Psychologist*, 2 (1989), 64-68.

¹¹ M. Weinstock, 'Citation indexes', *Encyclopedia of Library and Information Science*, 5 (1971), 16-40.

¹² Johnes, 'Research Performance Indicators in the University Sector'.

¹³ A. J. Chapman, 'Assessing Research: Citation-count Shortcomings', *The Psychologist*, 8 (1989), 336-344.

¹⁴ J. P. Rushton, 'Evaluating Research Eminence in Psychology: The Construct Validity

of Citation Counts'; J. P. Rushton, 'A Ten-year Scientometric Revisit of British Psychology Departments'.

¹⁵ See, for example, K. E. Clark, *America's Psychologists: A Survey of a Growing Profession* (Washington, DC: American Psychological Association, 1957); C. R. Myers, 'Journal Citations and Scientific Eminence in Contemporary Psychology', *American Psychologist*, 25 (1970), 1041-1048.

¹⁶ J. S. Robey, 'Political Science Departments: Reputation Vs. Productivity', *PS: Political Science and Politics*, 12 (1979), 202-209.

¹⁷ I. Crewe, 'Reputation, Research and Reality: The Publication Records of UK Departments of Politics, 1978-1984', *Scientometrics*, 14 (1988), 235-250.

¹⁸ W. M. Cox and V. Catt, 'Productivity Ratings of Graduate Programs in Psychology Based Upon Publication in the Journals of the American Psychological Association', *American Psychologist*, 32 (1977), 793-813; G. S. Howard, D. A. Cole and S. E. Maxwell, 'Research Productivity in Psychology Based on Publication in the Journals of the American Psychological Association', *American Psychologist*, 42 (1987), 975-986.

¹⁹ J. M. McCormick and E. L. Bernick, 'Graduate Training and Productivity: A Look at Who Publishes', *Journal of Politics*, 44 (1982), 212-227; D. R. Morgan and M. Fitzgerald, 'Recognition and Productivity Among American Political Science Departments', *Western Political Quarterly*, 30 (1977), 342-350; S. Welch and J. R. Hibbing, 'What Do the New Ratings of Political Science Departments Measure?', *PS: Political Science and Politics*, 16 (1983), 532-540.

²⁰ Colman, Garner and Jolly, 'Research Performance of United Kingdom Psychology Departments'; A. M. Colman, S. P. Grant and E. Henderson 'Performance of British University Psychology Departments as Measured by Number of Publications in BPS Journals', *Current Psychology: Research and Reviews*, 11 (1992), 360-368.

²¹ There were eleven special cases, each of them an old university without a single unambiguously identifiable politics department carrying one of the names mentioned in the text but with more than one department contributing to research in politics. The names of the 'special case' universities that were included in the study are as follows. Birmingham University: Political Science and International Studies; Regional, Urban and Local Government Studies; Russian and East European Studies. Bradford University: European Studies; Peace Studies. Bristol University: School for Advanced Urban Studies; Politics. Edinburgh University: Centre of European Government Studies; Europa Institute; Politics. Keele University: International Relations; Politics. Lancaster University: Politics; Politics and International Relations. Oxford University: Politics; Social Studies (politics staff only). Strathclyde University: Politics; Government; Centre for the Study of Public Policy. Sussex University: International Relations; Politics. Ulster University: Philosophy and Politics (see text); Social Administration and Policy. University of Wales, Swansea: Political Theory and Government; Centre for Russian and East European Studies.

²² W. M. Cox and V. Catt, 'Productivity Ratings of Graduate Programs in Psychology Based Upon Publication in the Journals of the American Psychological Association'; G. S. Howard, D. A. Cole and S. E. Maxwell, 'Research Productivity in Psychology Based on Publication in the Journals of the American Psychological Association'.

²³ Colman, Grant and Henderson 'Performance of British University Psychology Departments as Measured by Number of Publications in BPS Journals'.

²⁴ J. M. McCormick and E. L. Bernick, 'Graduate Training and Productivity: A Look at Who Publishes'; D. R. Morgan and M. Fitzgerald, 'Recognition and Productivity Among

American Political Science Departments'; S. Welch and J. R. Hibbing, 'What Do the New Ratings of Political Science Departments Measure?'

²⁵ E. Garfield, 'Mapping the World of Science: Is Citation Analysis a Legitimate Evaluation Tool?'

²⁶ Institute for Scientific Information, *Journal Citation Reports: A Bibliometric Analysis of Social Science Journals in the ISI Database. Printed Guide to the Microfiche Edition* (Philadelphia, PA: Institute for Scientific Information, 1992).

²⁷ G. S. Howard, D. A. Cole and S. E. Maxwell, 'Research Productivity in Psychology Based on Publication in the Journals of the American Psychological Association'.

²⁸ I. Crewe, 'Reputation, Research and Reality: The Publication Records of UK Departments of Politics', 1978–1984'.

²⁹ I. Crewe, 'Reputation, Research and Reality: The Publication Records of UK Departments of Politics', 1978–1984'.

³⁰ Universities Funding Council, *Research Assessment Exercise 1992: The Outcome: Circular Letter 26/92* (Bristol: Universities Funding Council, 1992), at p. 15.

³¹ Universities Funding Council, *Research Assessment Exercise 1992: The Outcome: Circular Letter 26/92*, at p. 15.

³² See A. Tversky and D. Kahneman, 'Availability: A Heuristic for Judging Frequency and Probability', *Cognitive Psychology*, 5 (1973), 207-232.

³³ Universities Funding Council, *Research Assessment Exercise 1992: The Outcome: Circular Letter 26/92*.

³⁴ See, for example, Gillett, 'Serious Anomalies in the UGC Comparative Evaluation of the Research Performance of Psychology Departments'; Gillett, 'Research Performance Indicators Based on Peer Review: A Critical Analysis'.

³⁵ Crewe, 'Reputation, Research and Reality: The Publication Records of UK Departments of Politics, 1978–1984'.

³⁶ Johnes, 'Research Performance Indicators in the University Sector'.

³⁷ Crewe, 'Reputation, Research and Reality: The Publication Records of UK Departments of Politics, 1978–1984'.

³⁸ See, for example, D. Crane, 'Scientists at Major and Minor Universities: A Study of Productivity and Recognition', *American Sociological Review*, 30 (1965), 699-714; B. Meltzer, 'The Productivity of Social Scientists', *American Journal of Sociology*, 55 (1949), 25-29.

³⁹ A. J. Nederhof, 'Books and Chapters are Not to be Neglected in Measuring Research Productivity', *American Psychologist*, 44 (1989), 734-735.

⁴⁰ G. S. Howard, D.A. Cole and S. E. Maxwell, 'On Productive and Unproductive: Comments About Productivity', *American Psychologist*, 44 (1989), 739-741, at p. 739.

TABLE 1 Research Performance Scores (1987–1992), Per Capita Simple Publication Counts (1978–1984), and UFC Ratings (1989–1992) of Politics Departments

University	Selected Articles ^a	Performance	Rank	All Pub's ^b	Rank	UFC Rating ^c
Strathelyde	15.52	1.098	1	10.74	9	5
Essex	14.00	0.651	2	16.83	2	5
Salford	4.92	0.621	3	6.95	24	(4) ^d
Hull	7.87	0.617	4	16.55	3	5
Aberdeen	5.40	0.488	5	7.62	18	3
Newcastle upon Tyne	6.18	0.460	6	6.24	30	4
Oxford	21.32	0.454	7	11.94	6	4
Exeter	5.60	0.417	8	8.80	14	4
Bristol	12.51	0.404	9	12.53	5	3
Glasgow	4.87	0.402	10	7.26	20	5
Edinburgh	6.40	0.400	11	9.59	11	3
Brunel	4.47	0.349	12	7.61	19	2
Open	2.00	0.343	13	7.21	21	3
York	5.60	0.338	14	6.68	26.5	4
Keele	5.40	0.326	15	5.65	34.5	4
Manchester	8.40	0.318	16	5.65	34.5	5
London: LSEPS	13.60	0.310	17	8.68	15	5
Lancaster	5.60	0.303	18	5.82	33	3
Leicester	2.60	0.278	19.5	5.39	37	3
Liverpool	4.21	0.278	19.5	9.05	13	3
Warwick	6.00	0.259	21	7.92	16	4
Belfast	3.00	0.241	22	6.04	32	3
Reading	3.00	0.228	23	4.50	41	2
Birmingham	12.60	0.222	24	2.51	45	3
Sheffield	2.00	0.200	25	4.84	39	4
Leeds	2.60	0.183	26	6.20	31	3
Loughborough	2.00	0.126	27	4.01	42	(4) ^d
Sussex	2.00	0.125	29.5	3.39	44	3
London: Birkbeck	1.00	0.125	29.5	11.83	7	4
London: QMWC	1.00	0.125	29.5	7.20	22	3
Nottingham	1.00	0.125	29.5	7.66	17	3
London: SOAS	1.00	0.100	32	3.56	43	3
Southampton	1.00	0.071	33	11.32	8	4
Kent	1.00	0.065	34	5.22	38	3
Wales: Swansea	1.00	0.061	35	6.37	28	4

Bradford	1.60	0.059	36	4.68	40	4
Durham	0.40	0.044	37	6.34	29	3
Ulster	1.40	0.043	38	–	–	3
Dundee	0.28	0.024	39	6.81	25	2
Stirling	0.00	0.000	40.5	–	–	3
Wales: Aberystwyth	0.00	0.000	40.5	6.68	26.5	4
Bath	4.94	–	–	5.48	36	–
Cambridge	4.40	–	–	22.05	1	–
East Anglia	3.00	–	–	9.17	12	3
London: King's	–	–	–	14.43	4	5
London: ICS	–	–	–	7.05	23	2
Mean	4.83	0.28	–	8.00	–	3.54
Standard Deviation	4.73	0.22	–	3.88	–	0.90

^aArticles in nine high citation impact European journals, 1987–1992, uncorrected for departmental size – corrected per capita scores are shown in the following column as performance scores.

^bPer capita simple publication counts, 1978–1984, according to data reported by Crewe (1988, Table 3).

^cRatings on a scale of 1(low) to 5 (high) issued by the Universities Funding Council (1992) following its research selectivity exercise covering the period 1989–1992.

^dThese UFC ratings were given by a different subject panel (European Studies), which may have used different criteria and standards from the Politics and International Studies subject panel, and the ratings were therefore not included in the statistical analysis reported in the text.