Political Studies Review
Political Studies Association
POLITICAL STUDIES REVIEW: 2009 VOL 7, 73–83

Gradgrinding the Social Sciences: The Politics of Metrics of Political Science

Claire Donovan
The Australian National University

This article employs an interpretive approach, and in the light of contributions to this symposium by Butler and McAllister, and McLean *et al.*, holds that metrics of research 'quality' are socially constructed and hence are as 'subjective' as peer review. Thus it rejects the use of stand-alone metrics as an 'objective' basis to inform funding allocations. Rather, the optimum method of 'quality' assessment is a panel-based exercise with expert judgement informed by a range of discipline-sensitive metrics and peer review of publications. The article maintains that the politics of metrics of political science conceals interests about the foundations of social scientific knowledge, and so the dispute over metrics and peer review is a metaphor for the conflicting epistemological preferences of UK political scientists. It is also argued that metrics-led assessment subjects political science to 'Gradgrinding' on two fronts: that political science departments amount to less than the sum of their parts, and the audit culture strips the discipline of its humanism.

In 2004, through the use of an innovative bibliometric analysis of research output, a *Political Studies Review* article brought a new quantitative dimension to benchmarking the relative standing of British political science departments internationally (Hix, 2004). The timing of the article's publication was auspicious as this marked the midway point of the assessment period for the last-ever Research Assessment Exercise (RAE) and also initial UK government interest in developing 'quality' metrics to assist in distributing block funding to university departments.[1] The results of the final RAE will be released to universities during December 2008 and January 2009, and made public in the following months.[2] This panel-based assessment of research quality will then be replaced by the Research Excellence Framework (REF), the form of which is, as yet, undecided. UK government preferences for how the new REF will be applied to the social sciences have shifted from a simple funding formula centred on the value of research income generated, to generic and discipline-specific metrics perhaps combined with 'light-touch' peer review.[3] Internationally, government policies have also fluctuated between preferences for peer review and metrics-only approaches, Australia being a prime example of this (Butler, 2007; Donovan, 2007a; Donovan, 2007b).[4] Yet despite these vacillations, it is clear that national research assessment exercises are like moths being inexorably drawn towards the flame of publication and citation metrics.

Despite government pressure to incorporate bibliometrics into panel proceedings for the 2008 RAE, the sub-panel for politics and international studies chose not to employ citation metrics to aid their deliberations, and declined to establish a list of relative journal rankings as it 'recognises that some types of research are published in less prominent or more specialist journals' (HEFCE, 2006, p. 31, p. 34). And in a submission to a Department for Education and Skills (DfES) consultation on the reform of higher education assessment and

research funding, the British International Studies Association and the Political Studies Association (BISA/PSA) argued against a greater use of metrics as 'there is no straightforward relationship between metrics and research quality in this subject area, and we strongly believe that peer review should be retained as the best and most credible means of assessing research quality' (BISA/PSA, 2006, p. 2).

This is the context within which the articles contributed to this *Political Studies Review* symposium by Linda Butler and Ian McAllister, and Iain McLean, André Blais, James C. Garand and Micheal Giles, are considered. This article employs an interpretive analysis, and builds on previous work (Donovan, 2007c) to maintain that, in the assessment of political science research, 'quality' metrics are as socially constructed and as 'subjective' as peer review. Three key issues emerge. First, the politics of metrics of political science: support for metrics or peer review conceals interests about the foundations of social scientific knowledge and, as such, this debate is a metaphor for the conflicting epistemological preferences of UK political scientists. Second, we are witnessing a Gradgrinding[5] of political science in two dimensions: (a) that metrics-led assessment can make political science departments amount to less than the sum of their parts; and (b) metrics-based audit bereft of peer review strips the discipline of its humanism through diminishing the importance of interpretation and expertise. Finally, following best practice in the bibliometrics community, the optimum method of 'quality' assessment is a panel-based exercise with expert judgements informed by a range of discipline-sensitive bibliometric data and peer review of publications. It is also symbolic of encouraging pluralism in UK political science.

Making Political Science Count

In its response to a Higher Education Founding Council for England (HEFCE) consultation on the REF, The British Academy signalled that it was in principle supportive of using metrics to aid research assessment, yet 'the metrics that currently exist for the humanities and social sciences are not acceptable as measures of quality', and so concluded, 'Much more work needs to be done to identify and clarify such metrics before they could supply a robust and accurate information [*sic*], as well as being sensitive to subject-specific differences' (2008, p. 1). The articles provided for this symposium by Butler and McAllister and McLean *et al*. are important contributions as they seek to develop bibliometric indicators specifically to assess political science research. Butler and McAllister (2009, p. 4) note that experimental work has taken two directions in the social sciences and humanities: producing ranked publication output lists, and testing novel bibliometric measures. We are fortunate, then, that McLean *et al*. tackle the former, and Butler and McAllister the latter.

Butler and McAllister: Evaluating the 2001 Research Assessment Exercise

The BISA/PSA submission to the DfES consultation on the reform of higher education research funding recognised that a major reason for resistance to using bibliometric measures to evaluate the quality of political science research was that standard citation measures neglect the importance of the single-authored book as 'the key research product

in this study area' (BISA/PSA, 2006, p. 2).[6] Butler and McAllister provide valuable data on the types of publication submitted for peer review in the 2001 RAE (journal articles 51.2 per cent, book chapters 23.2 per cent, books 20.0 per cent, edited books 2.1 per cent), confirming that in politics and international studies, journal articles account for only half of submitted output. Their major innovation is to undertake the laborious process of searching the Web of Science through the 'back door'[7] to mine the citations indexed journal papers make to books, chapters and non-indexed articles. However, the success of this enterprise is tempered by two important considerations. First, their citation data are 'asymmetrical' (Clemens *et al.*, 1995, p. 449; Cronin *et al.*, 1997, p. 270) as they do not capture citations made by or between books, for example. Yet their study does provide significant data that would otherwise be missed, illustrating that within this expanded ISI citation universe, books attract a mean of 18.2 citations per publication, compared with 4.1 citations per publication for journal articles. Second, and crucially, it seems citations are allowed to accrue for more than five years beyond the 2001 RAE's assessment period: while some fascinating data are supplied we shall see that, at the heart of the analysis, like is not compared with like.

A key aim of Butler and McAllister's study was to 'explore the outcome of the 2001 Research Assessment Exercise (RAE) in political science in the UK in order to test whether citations could have been used to replace subjective peer evaluation' (Butler and McAllister, 2009, p. 3). To facilitate this, they constructed a database comprising all staff members' nominated research publications, then gauged research 'quality' on the basis of mean citations per department (p. 7), and found that mean citations were the most important predictor of the 2001 RAE outcome for a department (p. 9). Butler and McAllister's primary motivation was to 'consider to what extent the use of peer evaluation to measure research quality could be dispensed with, relying instead solely on objective measures such as citations, income and student numbers' (p. 12). They concluded that a metrics-based approach, with bibliometrics as its central pillar, will yield similar results to a peer-based evaluation, and that parsimony suggested metrics-based evaluations 'are a much more efficient and transparent approach to measuring research quality' (p. 13).

The outcome of the citation analysis is, therefore, central to Butler and McAllister's argument. Yet we encounter a surprising lack of transparency. With the exception of naming three books that account for the high mean citation scores of Birkbeck and the University of Westminster, no other examples of highly cited publications are given. There is no discussion of the limitations or practical difficulties of mining the Web of Science for non-standard citation data. Essential methodological detail about the citation window used is not supplied. In effect, the list of nominated publications is fed into a black box and we are presented with data on total, proportion and mean citations per publication by publication type, and some information on the distribution of citations across departments.

A closer examination of citations made to the three named books raises several concerns about the methodology of the citation analysis. Butler and McAllister tell us that three highly cited books are: Chantal Mouffe (Verso, 2000) *The Democratic Paradox* (118 citations); Barry Buzan, Ole Wæver and Jaap de Wilde (Lynne Rienner Publishers, 1998) *Security: A New Framework for Analysis* (141 citations); and Paul Hirst [and Grahame Thompson] (Polity, first

Table 1: Accumulated Total Citations to Butler and McAllister's Three Highly Cited Books

	Total citations accumulated			
	*December 2000**	*Butler and McAllister*	*December 2006**	*September 2008**
Chantal Mouffe (2000) *The Democratic Paradox*	2	118	129	185
Barry Buzan *et al.* (1998) *Security: A New Framework for Analysis*	38	141	149	212
Paul Hirst and Grahame Thompson (1996) *Globalization in Question*	316	711	724	769

*Source: *Thomson Reuters ISI Web of Science, cited reference search (conducted 17 September 2008).*

edition, 1996) *Globalization in Question: The International Economy and the Possibilities of Governance* (711 citations). A detailed Web of Science 'cited reference search' for citations to these publications received during the 2001 RAE assessment period (1996 to 2000) produces 2, 38 and 316 citations, respectively, far below Butler and McAllister's total (see Table 1).

An extended search including all citations to date[8] reveals too many citations: 185, 212 and 769, respectively. The closest match to Butler and McAllister is for citations to the end of 2006. So according to current ISI data, we may infer that their figures relate to a citation window ending at some point in 2006, or perhaps a later time as ISI data are regularly updated retrospectively. In any case, it appears that Butler and McAllister's citation window extends at least five years beyond the 2001 RAE's assessment period. This is problematic because, as we have seen, Butler and McAllister set out to 'test whether citations could have been used to replace subjective peer evaluation' for politics and international studies in the 2001 RAE (p. 3). To extend the citation window beyond the RAE assessment period is to generate data that would not have been available at that time, and so like is not compared with like.

Yet Table 1 demonstrates the importance of appropriate citation windows for political science research, and that optimum citation data for books do not necessarily complement the length of the RAE cycle. This is particularly so for books published near the end of the assessment period, Mouffe's two citations and Buzan *et al.*'s 38 citations being striking examples of this, although one would assume that peer review would have identified these as high-quality texts. This raises the interesting question of how long an appropriate citation window is for various types of political science publication, and how we might balance this with the length of the RAE cycle and the government's desire to allocate funds on the basis of recently published research.[9]

As we have seen, Butler and McAllister use the Web of Science for their 'back door' citation analysis. This database was not set up to provide data on citations to books at the press of a button, and so does present some obstacles. One of its major limitations is that in the case of jointly authored books, only the first author is listed. And so a further issue arises when

Table 2: Accumulated Citations for First and Second Editions of *Globalization in Question*

	December 2000	*December 2006*	*September 2008*
1999 edition	8	153	176
1996 edition	308	565	587

Source: Thomson Reuters ISI Web of Science, cited reference search (conducted 17 September 2008). Note: Six citations are missing for December 2006 and September 2008 as incorrect publication years are entered into the ISI database.

we consider the most highly cited book in Butler and McAllister's study: their article credits the late Paul Hirst as the author of *Globalization in Question* and does not mention the book's co-author Grahame Thompson (p. 8). In the light of figures supplied on mean citations per submitted work across departments (p. 7 and endnote 11), it appears that the Open University has not been credited with the 711 citations.

A further issue arises when we consider that both Hirst and Thompson submitted the second edition of *Globalization in Question* for peer assessment (Polity, 1999).[10] Table 2 illustrates how the first (1996) and second editions have accumulated citations.

Butler and McAllister have included both editions in their calculations, which is fair given that both editions were published within the 2001 RAE assessment period. Yet we can imagine that as part of a centralised process, a HEFCE data entry clerk might have performed a 'cited reference search' limited to the second edition and the 1996–2000 assessment period, producing eight citations rather than the 711 Butler and McAllister claim. This raises another concern: imagine a first edition of a hypothetical highly cited book is published in 1995 and so before the assessment period, and then a substantially revised and also highly cited second edition is published in 1999. Should the citations that accumulate to the first edition during the assessment period be counted, given that the second edition will have little opportunity to be cited? If the second edition appeared in 1997, would the calculation of citations be guaranteed to exclude citations to the 1995 first edition?

This discussion of issues highlighted by the data provided on the three highly cited books throws up some very thorny issues about the transparency and validity of Butler and McAllister's citation study: and there are another 877 books, 2,251 journal articles, 1,021 chapters and 155 edited books to consider. First, and foremost, their study does not provide an accurate test of whether citations could have been used in place of peer review in the 2001 RAE as the citation window employed extends at least five years beyond any data that could have been supplied at that time. While the citation data are illuminating, they do not do what they purport, and so this central pillar of the study does not support the reworked departmental rankings or the claim that a metrics-based system may safely replace or reduce the scope of peer review. It would be extremely interesting to view the outcomes of a revised analysis that limited citation data to the 2001 RAE assessment period.

McLean et al.*: Comparative Journal Ratings*

Butler and McAllister note that the first experimental route in developing bibliometric indicators for the humanities and social sciences is constructing ranked publication output lists (p. 4), and this is the direction McLean *et al.* take with their political science journal ratings. Their motivation is dissatisfaction with the shortcomings of citation metrics and the need to develop alternative measures specific to the social sciences, which they situate as intermediate between the natural sciences and humanities.

Theirs is an impressive contribution as it reports the results of three separate surveys of the Canadian, UK and US political science profession and so allows for comparative analysis of these ranked lists of 92 journals. Respondents were asked to rate journals they were familiar with out of ten for the quality of the articles these publish. Most notably, to borrow a phrase, there is a 'fusion of horizons' in the top 10 of each list: in terms of impact,[11] Canada and the UK share 8 in common, Canada and the US 8 and the UK and US 7. Although, on a less harmonious note, while *Canadian Public Policy* and *Canadian Public Administration* are rated 32 and 36, respectively, by Canadian political scientists, both the US and UK place these 91st and 92nd, pointing to obvious regional preferences.[12]

While an important development, and while the ratings do include ISI and non-ISI journals, book publication is excluded. The approach was previously adapted for this purpose by the Australian Political Studies Association (AusPSA) which undertook a consultation to produce ranked bands of journals and academic book publishers.[13] Perhaps the next wave of the survey could seek similar data? It is a pity that data on preferences by sub-field and methodological approach most often employed[14] remain pending, as this would be a particularly illuminating contribution to debates about the degree to which journal rankings reinforce 'positivistic' data gathering.

McLean *et al.* assert that their 'expert judgements of journal rankings are robust and may be confidently used by those tasked with grading people, publications or university departments' (McLean *et al.*, 2009, p. 35). The use of journal rankings to support promotion or appointment processes is the least controversial notion,[15] but how might ranked journal lists be adapted to national research assessments? The idea of journal rankings generated by consultation with political scientists is appealing as the approach is bottom up and democratic. Yet the academic community tends to rebel against the idea that all publications in highly ranked journals are *de facto* of higher quality than the others, and *vice versa*.[16] When applied to national (or international) research evaluation, ranked lists have proven controversial. For example, in the UK, after strong academic opposition the Arts and Humanities Research Council was forced to abandon an attempt to construct lists of the top ten 'most significant and important' journals in various subject areas for quality evaluation purposes; and the construction of ranked journal lists to inform panel decisions for Excellence in Research for Australia (ERA) has proved similarly controversial as has the development of the European Reference Index for the Humanities (Donovan, 2008). Yet Norway successfully employs a publications-based funding formula divided into tiers of publisher prestige applied to ISI and non-ISI journal papers, book chapters and books (Sivertsen, 2006). It would be informative to know how McLean *et al.* would wish to

operationalise their rankings at the national level: in tandem with peer review of publications? As part of a suite of publication metrics provided to expert peers? Or as a stand-alone bibliometric measure?

The Politics of Metrics of Political Science

This article maintains that the politics of metrics of political science conceals interests about the foundations of social scientific knowledge, and so the dispute over metrics and peer review is a metaphor for the conflicting epistemological preferences of UK political scientists. When championing discipline-sensitive metrics, we should not adopt bibliometric tools that intentionally or inadvertently reward the 'right' kind of political science and punish the 'wrong' sort. It is also essential to maintain a central role for peer review within the policy process: problems with implementation may mean that discipline-specific measures are compressed or abandoned, and only peer review of publications and expert peer consideration of metrics will safeguard the interests of all areas of political science.

As noted above, there has been an international policy drift towards using bibliometrics in national research evaluation exercises, and this is often grudgingly accompanied by governments allowing 'light-touch' peer review for the social sciences. The British Academy noted in its submission to the REF consultation: 'There is a growing recognition of the difficulties of reconciling a light-touch approach with the additional demands that a greater use of, and reliance on, metrics will bring' (British Academy, 2008, p. 8), a notion that will now be explored further. There is little discussion of what, precisely, constitutes 'light-touch' peer review.[17] It can mean two things. One option is to present expert panels with the results of a traditional peer-review exercise and a separate metrics-based assessment, where the task of the panel is to consider both sets of data in order to form a final quality judgement (Moed, 2007). Another is for metrics to be the dominant form of information which panels of peers then consider (Butler, 2007), and actual review of 'outputs' will only take place in rare cases where there are no appropriate quality metrics. So in one, traditional peer review is retained and metrics provide additional data for panels to consider; in the other, assessment is metrics driven, and reviewing publications and other 'outputs' is a last resort.

This article maintains that we should follow best practice in the bibliometrics community, where the optimum method of 'quality' assessment is a panel-based exercise with expert judgements informed by a range of discipline-sensitive bibliometric data and peer review of publications. This is also symbolic of encouraging pluralism in UK political science.

Gradgrinding the Social Sciences

We are witnessing a Gradgrinding of political science on two fronts. The first is that metrics-led assessment can make political science departments amount to less than the sum of their parts. In *Hard Times* a definition of a horse is given as follows:

> Quadruped. Graminivorous. Forty teeth, namely twenty-four grinders, four eye-teeth, and twelve incisive. Sheds coat in the spring; in marshy countries sheds

> hoofs, too. Hoofs hard, but requiring to be shod with iron. Age known by marks in mouth Thus (and much more) ... Now girl number twenty', said Mr Gradgrind. 'You know what a horse is' (Dickens, 1987, p. 18).

A horse is clearly more than this. By analogy, describing a department's research quality through recourse to metrics alone leaves a fragmented picture built from quality indicators demonstrated not to be direct measures of research quality (Donovan, 2007b, p. 586, p. 592). While these 'facts' may provide some useful data, a clear picture may only be provided by additional and qualitative description.

Second, metrics-based audits which minimise or exclude the role of peer review strip the discipline of its humanism through diminishing the importance of interpretation and expertise. To argue that metrics are more efficient, transparent and objective is a utility-driven posture that overlooks the fact that the 'objective evidence' of journal rankings and citations is the sum of many 'subjective' judgements: these contingent choices could have been different and so are socially constructed. We should be aware that metrics-led approaches may disadvantage some sub-fields or methodologies (Donovan, 2007c, pp. 672–4), and so be an indirect means of adding strings to block funding. This Gradgrinding replaces values, interpretation and 'subjective' judgements with scientific 'facts' and the shibboleth of quantitative data. Metrics have the potential to become technologies of governance (Donovan, 2007c, pp. 670–7), and we should remain aware that while metrics may simplify assessment, they may also simplify the scope and aspirations of political science through privileging 'positivistic' knowledge to the exclusion of interpretive knowledge.

Conclusion

When Howard White *et al.* consider alternative approaches to research assessment in the humanities and social sciences, they conclude that we 'someday have to decide whether it is worse to have [our] hearts broken qualitatively or quantitatively' (2008, p. 2). This article has argued that it is better for UK political science to be a sum of its parts, and so have its heart broken both ways. The purpose of this article has been to highlight that apparently 'objective' bibliometric data are the sum of 'subjective' judgements and thus socially constructed. Bibliometrics should not therefore be accorded superior status to quality judgements arrived at by expert panels informed by generic and discipline-sensitive metrics and the peer review of publications. This article therefore supports the development of a range of metrics to aid the panel process, and recognises the importance of the novel citation study and comparative journal rankings presented in this *Political Studies Review* symposium as steps towards this.

It is ironic that *Political Studies Review* is a journal that is not currently indexed by ISI. It seems, by way of conclusion, fitting to see how Simon Hix's article, 'A Global Ranking of Political Science Departments', published in this journal in 2004 has fared in terms of the citations it has attracted from three popular databases: Google Scholar, Scopus and Web of Science. Table 3 shows that the article's performance varies according to which 'objective' measure is used.

Table 3: Citations to Hix's 'A Global Ranking of Political Science Departments'

Database	*Raw data*	*Cleaned data*
Google Scholar	59	40
ISI Web of Science (cited reference search)	18	18
ISI Web of Science (standard search)	0	0
Elsevier Scopus	0	0

Note: Search conducted 17 September 2008.

Hix's article attracts most citations from Google Scholar, although the 'raw data' do not control for citation from peer-reviewed literature; a cleaned list gives 40 citations. A standard Web of Science search produces no hits, and no citations are found on Scopus either. Following Butler and McAllister's lead, a rather circuitous 'cited reference search' of the Web of Science produces eighteen citations. Although this test was conducted for entertainment rather than scientific purposes, it is interesting to note that Google Scholar detects more than twice the number of citations from peer-reviewed publications than the Web of Science. It is, however, clear that there is a long way to go in developing citation databases which can be applied with confidence to the various forms of publication political scientists produce.

(*Accepted*: 19 September 2008)

About the Author

Claire Donovan, Research Evaluation and Policy Project, Research School of Social Sciences, The Australian National University, ACT 0200, Australia; email: *claire.donovan@anu.edu.au*

Notes

Thanks are due to Phil Larkin for his comments on this article.

1 This followed a recommendation of the Roberts Review of Research Assessment (2003) that the funding councils begin to develop discipline-specific metrics of research quality.

2 See http://www.rae.ac.uk/pubs/2008/cl/01/ [Accessed 17 September 2008].

3 See http://www.hefce.ac.uk/Research/ref/ [Accessed 17 September 2008].

4 These papers form part of a special edition of *Science and Public Policy* published in October 2007, and trace the transition in Australia from a simple funding formula based on research grant income, student data and publication output to a Research Quality Framework (RQF) incorporating peer review of publications, discipline-specific metrics and the assessment of the extra-academic impact of research. However, in December 2007 the RQF was scrapped by the new Labor government and in February 2008 replaced by Excellence in Research for Australia (ERA), a metrics-based quality assessment that will use discipline-specific bibliometric data to inform 'light-touch' peer review (see http://www.arc.gov.au/era/default.htm) [Accessed 17 September 2008].

5 I use the term 'Gradgrinding' to refer to a reliance on data and 'facts' over expertise and interpretation. Thomas Gradgrind is the headmaster in Charles Dickens' *Hard Times*, for whom the world consists of facts and facts alone, and children are 'little pitchers ... who were to be filled so full of facts' (Dickens, 1987, p. 16). He is supposedly modelled on James Mill and the overly rigorous education he gave to his son, John Stuart Mill.

6 'Standard' citation analyses rely on data supplied by Thomson Reuters Institute of Scientific Information (ISI) Web of Science, and capture the citations made between indexed journals. This database was originally constructed to map publication patterns in the natural sciences, and so excludes citations to and from 'non-standard' publications such as books, book chapters, non-indexed journals and conference proceedings.

7 The Web of Science contains data on all the citations made by indexed papers, including those to non-source items. It is possible to conduct a 'cited reference search' to locate citations made to a particular book, for example. However, as the database was not constructed with this use in mind, data entry is not uniform and often requires substantial cleaning.

8 This search was conducted on 17 September 2008.

9 We can imagine that in a metrics-only assessment system, rather than rushing to meet the RAE assessment deadline, academics would aim to just miss it, thus increasing their chance to accumulate citations during the next round of assessment.

10 See http://www.hero.ac.uk/rae/rae_dynamic.cfm?myURL=http://195.194.167.103/submissions/Inst.asp?UoA=39 [Accessed 17 September 2008].

11 Impact = (mean rating) + (familiarity * mean rating).

12 The *Political Studies Review* readership may be interested to note that in terms of 'impact', *Political Studies* is ranked 3rd in the UK list, 19th in the Canadian list and 49th in the US list.

13 See http://www.auspsa.org.au/images/stories/RQF/final%20journal%20rankings%20nov%2007.xls for the journal ranking and http://www.auspsa.org.au/images/stories/RQF/publishers%20ranking%20nov%2007.xls for the publisher ranking [Accessed 17 September 2008].

14 The survey gives the following options for methodological approach most often employed: quantitative, qualitative, normative and formal. However, an improvement would be to investigate epistemological inclination; for example categories such as 'positivist'/ empiricist' and 'interpretive' would prove more informative as either might use qualitative and quantitative methodologies.

15 But see Donovan (2007c, pp. 674–6) as a critique.

16 For a similar perspective from an RAE panel member see Paul (2008).

17 An REF pilot study document says that 'light-touch' peer review will be used in areas where bibliometrics are difficult to apply, but does not clarify what 'light-touch' peer review is. See http://www.hefce.ac.uk/research/ref/pilot/REF.pdf [Accessed 17 September 2008].

References

British Academy (2008) Response to Questions Posed by the HEFCE Consultation on the Research Excellence Framework. Available from: http://www.britac.ac.uk/reports/ref/index.cfm [Accessed 17 September 2008].

British International Studies Association/Political Studies Association (BISA/PSA) (2006) Reform of Higher Education Research Assessment and Funding: Response to DfES Consultation Exercise. Available from: http://www.psa.ac.uk/Publications/2006%20RAE%20consultation.pdf [Accessed 17 September 2008].

Butler, L. (2007) 'Assessing University Research: A Plea for A Balanced Approach', *Science and Public Policy*, 34 (8), 565–74.

Butler, L. and McAllister, I. (2009) 'Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science', *Political Studies Review*, 7 (1), 3–17.

Clemens, E. S., Powell, W. W., McIlwaine, K. and Okamoto, D. (1995) 'Careers in Print: Books, Journals and Scholarly Reputations', *American Journal of Sociology*, 101 (2), 433–94.

Cronin, B., Snyder, H. and Atkins, H. (1997) 'Comparative Citation Rankings of Authors in Monographic and Journal Literature: A Study of Sociology', *Journal of Documentation*, 53 (3), 263–73.

Dickens, C. (1987) *Hard Times*, ed. T. Eagleton. London: Methuen.

Donovan, C. (2007a) 'Introduction: Future Pathways for Science Policy and Research Assessment: Metrics vs. Peer Review, Quality vs. Impact', *Science and Public Policy*, 34 (8), 538–42.

Donovan, C. (2007b) 'The Qualitative Future of Research Evaluation', *Science and Public Policy*, 34 (8), 585–97.

Donovan, C. (2007c) 'The Hidden Perils of Citation Counting for Australasian Political Science', *Australian Journal of Political Science*, 42 (4), 665–78.

Donovan, C. (2008) 'Taming the Pushmi-pullyu: State of the Art in Evaluating Research Quality in Humanities Research Quality-A View from Down Under' ('Das zweiköpfige Lama zähmen: Die australische Suche nach den besten Evaluierungsmethoden für die Geisteswissenschaften'), in E. Lack and C. Markschies (eds), *What the Hell is Quality? Qualitätsstandards in den Geisteswissenschaften*. Frankfurt: Campus, pp. 74–101.

Higher Education Funding Council for England (HEFCE) (2006) *RAE 2008 Panel Criteria and Working Methods: Panel J*. Bristol: HEFCE. Available from: http://www.rae.ac.uk/pubs/2006/01/docs/j39.pdf [Accessed 17 September 2008].

Hix, S. (2004) 'A Global Ranking of Political Science Departments', *Political Studies Review*, 2 (3), 293–313.

McLean, I., Blais, A., Garand, C. and Giles, M. (2009) 'Comparative Journal Ratings: A Survey Report', *Political Studies Review*, 7 (1), 18–38.

Moed, H. F. (2007) 'The Future of Research Evaluation Rests with an Intelligent Combination of Advanced Metrics and Transparent Peer Review', *Science and Public Policy*, 34 (8), 575–83.

Paul, R. J. (2008) 'Measuring Research Quality: The United Kingdom Government's Research Assessment Exercise', *European Journal of Information Systems*, 17 (4), 324–9.

Roberts, G. (2003) RA Review: Review of Research Assessment – Report by Sir Gareth Roberts to the UK Funding Bodies. Available from: http://www.ra-review.ac.uk/reports/roberts.asp [Accessed 17 September 2008].

Sivertsen, G. (2006) 'A Bibliometric Model for Performance Based Budgeting of Research Institutions', in *Proceedings of the 9th International Conference on Science & Technology Indicators held in Leuven, Belgium, 7–9 September 2006*. Leuven, Belgium: Katholieke Universiteit Leuven/Steunpunt O&O Statistieken, pp. 133–5.

White, H. D., Boell, S. K., Yu, H., Davis, M., Wilson, C. S. and Fletcher, C. (2008) 'Libcitations: A Measure for Comparative Assessment of Book Publications in the Humanities and Social Sciences'. Unpublished paper, College of Information Science and Technology, Drexel University/Bibliometrics and Informetrics Research Group, University of New South Wales.