

The nature of publishing and assessment in Geography and Environmental Studies: evidence from the Research Assessment Exercise 2008

Keith Richards¹, Mike Batty², Kevin Edwards³, Allan Findlay⁴, Giles Foody⁵, Lynne Frostick⁶, Kelvyn Jones⁷, Roger Lee⁸, David Livingstone⁹, Terry Marsden¹⁰, Judith Petts¹¹, Chris Philo¹², David Simon¹³, Susan Smith¹⁴ and David Thomas¹⁵

¹Department of Geography, University of Cambridge, Cambridge CB2 3EN

Email: ksr10@cam.ac.uk

²CASA, University College London, London WC1E 7HB

³Geography and Environment, University of Aberdeen, Aberdeen AB24 3UF

⁴School of Social & Environmental Sciences, University of Dundee, Dundee DD1 4HN

⁵School of Geography, University of Nottingham, Nottingham NG7 2RD

⁶Department of Geography, University of Hull, Hull HU6 7RX

⁷School of Geographical Sciences, University of Bristol, Bristol BS8 1SS

⁸Department of Geography, Queen Mary, University of London, London E1 4NS

⁹School of Geography, Archaeology and Palaeoecology, Queen's University, Belfast BT7 1NN

¹⁰School of City & Regional Planning, University of Cardiff, Cardiff CF10 3WA

¹¹School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, B15 2TT

¹²Department of Geographical and Earth Sciences, University of Glasgow, Glasgow G12 8QQ

¹³Department of Geography, Royal Holloway, University of London, Egham, Surrey TW20 0EX

¹⁴Department of Geography, Durham University, Durham DH1 3LE

¹⁵School of Geography and the Environment, University of Oxford, Oxford OX1 3QY

Revised manuscript received 20 May 2009

We present a summary of the kinds of outputs submitted to the Geography and Environmental Studies sub-panel (H-32) for the 2008 Research Assessment Exercise (RAE), and examine the relationships between the peer assessment of research quality that the RAE process has typified, and alternative modes of assessment based on bibliometrics. This comparison is effected using (in aggregate form) some of the results from the RAE, together with citation data gathered after completion of the RAE assessment, specifically for the purpose of this paper. We conclude that, if it continues to be necessary and desirable to assess, in some measure and however imprecisely, research quality, then peer assessment cannot be replaced by bibliometrics. Bibliometrics permit measurement of something that may be linked to quality but is essentially a different phenomenon – a measure of 'impact', for example.

Key words: RAE, output types, peer review, bibliometrics

Introduction

This paper is based on an analysis of the research published by UK geographers in the period 2001–2007 and submitted as a sample of their best outputs for the 2008 Research Assessment Exercise (RAE). Its purpose is to examine the nature of this published output in terms of the forms of publication involved, and the particular publishing outlets favoured by UK geographers. This analysis may itself provide some insights into the characteristics of Geography and Environmental Studies, independently of analysis of the substantive content of the work submitted (RAE sub-panel H-32 2008; and in preparation). It may also offer a different approach from some analyses based purely on citations, which can focus less on the output of individual researchers, and more on the aggregate citation character of different fields (van Raan *et al.* 2007). As bibliometric methods develop, new approaches emerge that analyse different dimensions and attributes of publication (e.g. Bollen *et al.* 2009), and permit different inferences to be made about the structure and quality of knowledge; these may complement peer review processes, by providing additional but distinctive information. Since there is debate about whether the future form of research assessment in the UK (to be called the Research Excellence Framework,

or REF) should involve more emphasis on metric data, including bibliometrics such as citation counts, the evidence presented here may both inform that debate and serve to provide assistance to UK geographers who have to make decisions about how best to manage their publication strategy in the future when faced with the REF.

The evidence discussed in this paper relates to the outputs submitted for assessment to the Geography and Environmental Studies sub-panel (H-32) in the 2008 Research Assessment Exercise.¹ Each UK academic was required to submit for assessment up to four research outputs perceived to be the best, published in the period 2001–2007. These were then peer reviewed by a panel of 15 academic geographers, assisted where necessary by additional specialist advisers or by referral to another panel, and were graded on a four-point scale (Table 1). A fifth category ('Unclassified') was used if the panel judged a submitted piece of work either to fall below the standard of nationally recognised work, or not to be within the definition of 'research' for RAE purposes (Table 1). Where an individual researcher was not in a post requiring research activity for the whole 7-year assessment period (for reasons of initial appointment after 2001, part-time employment, career break or illness), the RAE criteria (HEFCE 2005) allowed submission of fewer than four

Table 1 Definitions of the quality levels used in the 2008 Research Assessment Exercise and definition of research (from HEFCE 2005)

Grade	Descriptor
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 work which does not meet the published definition of research for the purposes of this assessment

Definition of research for the RAE

'Research' for the purpose of the RAE is to be understood as original investigation undertaken in order to gain knowledge and understanding. It includes work of direct relevance to the needs of commerce, industry and to the public and voluntary sectors; scholarship*; the invention and generation of ideas, images, performances, artefacts including design, where these lead to new or substantially improved insights; and the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction. It excludes routine testing and routine analysis of materials, components and processes, such as for the maintenance of national standards, as distinct from the development of new analytical techniques. It also excludes the development of teaching materials that do not embody original research.

* Scholarship for the RAE is defined as the creation, development and maintenance of the intellectual infrastructure of subjects and disciplines, in forms such as dictionaries, scholarly editions, catalogues and contributions to major research databases.

outputs. A total of 1210 individuals offered 4590 outputs for assessment, a process completed in October 2008; the outcomes showed that 40 per cent of these researchers produced at least one output rated as world-leading (4*) in the assessment period.

The 2008 RAE was the first in which a (sub)-panel covering Geography and Environmental Studies was employed (RAE sub-panel H-32 2008); previous panels had been for Geography alone. The Higher Education Funding Council defined 67 sub-panels, grouped into 15 main panels; no particular intellectual coherence was claimed to justify construction of a sub-panel for Geography and Environmental Studies. Virtually all UK University Departments of Geography submitted human, environmental and physical geography research to this sub-panel; however, a small number submitted their physical geographers to the Earth Systems and Environmental Sciences sub-panel (E-17), and some returned their work to other sub-panels such as Town and Country Planning (H-31). Some other units undertaking environmental research submitted to sub-panel H-32, but these were very variable in nature, including some with an interdisciplinary focus on environmental issues, and others with a narrower focus on, for example, biology or geology.

Types of publication: quantity and quality

In the total submission of some 4590 items there were several different kinds of output, with each item being defined as one of a series of predetermined output types at the time of submission. In the set of outputs with which this paper is concerned, 'Journal articles' clearly dominated, constituting 86.5 per cent of all outputs. In fact, they were probably over 89 per cent of the total because the majority of items classed as 'Internet publications' were actually 'on-line-early' versions of journal articles. The other main forms of output were those defined as 'Chapters in books' (5.3%) and 'Authored books' (4.7%), with small numbers of 'Other forms of assessable output' (there were only six of these), 'Edited books' (18 in total), 'Conference contributions' (12), 'Software' (2) and 'Digital and visual media' (2). Only 0.4 per cent of the total submitted outputs (17 examples) consisted of 'Research reports for external bodies', which might seem a disappointing representation of explicitly applied work, although many other outputs were applied in nature, and several items classed as 'Authored books' at the time of submission could easily have been classed as 'Research reports for external bodies' instead.² The low representation of some output types means that sample

sizes are too small to draw any general conclusions about their nature and quality.

Assessment of the outputs began with a presumption on the part of panel members that all types of output were in principle capable of being graded at the highest level. However, after completion of the assessment, it was apparent that the proportions of the different types of output in the highest-rated 4* (world-leading) category differed from their proportions in the total submission. For example, 'Journal articles' together with 'Internet publications' accounted for 82 per cent of the number of outputs in the 4* category. This is less than the proportion (89%) of such publications in the complete output set, which appears to contradict a widespread presumption that journal articles *per se* constitute the best form of publication towards which authors should aim. The representation of 'Chapters in books' in the 4* class (1.7%) was also reduced relative to that in the total submission, an even more significant proportional decrease. This might lend credence to another common view, that book chapters may not always be where the highest quality output is published. However, this requires qualification, since there are also book chapters of the highest quality; for example, chapters in a collection that gain added value from their context, perhaps when several scholars working together produce a benchmark analysis of considerable influence and longevity.

These decreases in representation were countered by a very significant relative over-representation of 'Authored books' in the 4* category – the proportion being, at 15.1 per cent, over three times that in the total submission. This implies a relatively high probability that an authored book submitted in the RAE will have been highly rated, although this is obviously not guaranteed – it depends on the quality of the book. However, where books were published early enough in the assessment period to have been the subject of critical review in the public domain, prior validation of quality by *this* process may have led to informed authorial self-selection of the best book-length outputs, exclusion of books with less supportive reviews, and prior enhancement of the quality of this class of submitted output. Nevertheless, the strong performance of book-length outputs is a significant finding in relation to proposed bibliometric means of assessment, as book citations are currently less reliably monitored.

In view of the concerns expressed about the state of monograph authorship in Geography (Harvey 2006; Ward 2009), this is also important evidence of the continued status of this form of scholarly output, and

the esteem in which it is held. Some authors committed themselves to more than one such output in the 7-year period; in fact, 29 authors produced two or more (and there were even cases of four by the same author). Almost 180 authors or authorial teams submitted book-length works. The qualification – ‘or authorial teams’ – implies, of course, that the definition being used here in referring to monographs is that which emphasises a focused and in-depth scholarly treatment of a narrow field, rather than a necessarily solitary authorship (see Ward 2009).

Patterns of journal use

There were two somewhat unexpected characteristics of UK Geography and Environmental Studies revealed by the RAE submissions, which for the first time were detailed in digital and searchable databases, and therefore provided much valuable information.

The balance of human and physical geography

The first of these was that the balance between outputs broadly classifiable as either human or physical geography was very close to 50 : 50. In fact, there were approximately 2220 outputs classified as physical geography and environmental science, while those in human geography and social scientific environmental studies numbered about 2370. As noted above, some additional physical geography outputs were submitted to the Earth Systems and Environmental Sciences panel, so the balance was perhaps even closer to 50 : 50.³ This implies that Geography and

Environmental Studies make a more significant contribution to environmental science in the UK than has been recognised (RAE sub-panel H-32 2008), with important implications for the funding levels necessary to sustain its associated laboratory- and field-based research. Unfortunately, a historic presumption that the proportion of physical geography is significantly less than 50 per cent, which probably reflects the balance of content in conferences such as those of the RGS-IBG and the AAG (which have largely become *de facto* human geography meetings), has not been to the benefit of the discipline, which in the UK continually has to remind the funding bodies of its part-science status.

Diversity of journal outlets

The second, rather remarkable, characteristic was the considerable diversity of the journal outlets employed by UK geographers to report their work. In total, there were just short of 750 individual journal titles in which the 3950 journal articles were published, and of these, just over 400 journals were used only once. A large number of outlets is in one sense unsurprising, in that there are significantly non-overlapping journal sets used by human and physical geographers, and the various sub-fields within each. Nevertheless, both physical and human geographers are promiscuous in their choice of journals, with these non-overlapping sets used by each group each including about 300 journals. Figure 1 is an overall summary of the rank order of journals according to the number of papers published in them (in this selected sample of outputs),

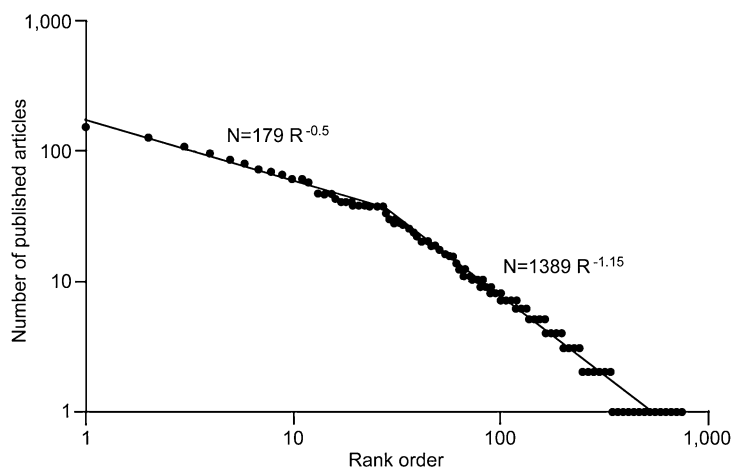


Figure 1 The relationship between the number of published articles (submitted for the 2008 RAE) in a journal and its rank

and this shows that as the rank decreases from 1 to 30, there is a relatively slow rate of decline in the numbers of papers published in successively lower-ranking journals, whereas beyond a rank of about 30th, the rate of decline increases. We suspect that this reflects a number of factors, ranging from perception of a higher standing for the journals in the top group, through the effects of censoring of outputs as a result of submissions being limited to the best four per person, to the effect on curve-fitting of large numbers of lower ranking journals having equal numbers of outputs.

Table 2 lists the 50 journals that published the most articles submitted to the Geography and Environmental Studies RAE return. The list shows a mix of journals used by both human and physical geographers, with general geography journals and more specialist journals both appearing, and even general science journals such as *Nature* and *Science* making the top 50. A cautionary note, however, is that while some multi-part journals such as *Environment and Planning* are treated as independent publications, this is not true of others, such as the *Journal of Geophysical Research*. Table 2 shows that the dominant 50 journals account for about 55 per cent of all outputs, implying that a further 700 are required to account for the remainder, and as noted above, a particularly surprising outcome is that there are some 400 journals that were only used once by authors returning their outputs to the panel.

Classifying journals crudely as either 'physical' or 'human', and plotting the numbers of the separate sets of articles against rank as in Figure 1 replicates the evidence of a break at about the 30th position. This is slightly more pronounced in physical geography, where the two exponents are -0.47 and -1.38 ; for human geography journals they are -0.74 and -1.17 . This implies that there is more continuity in the decline in the frequency of journal use in human geography than in physical. The set of physical geography journals numbers about 293, with 45 per cent of these having single submissions, while the human geography set of about 457 includes 60 per cent with single submissions. This may seem surprising, as physical geography has an anecdotal reputation for publishing in a more widely dispersed set of journals than human geography (Thrift 2002; Ferguson 2003), particularly in non-geography journals.

The evidence of such a wide range of journals, and especially of such a large number with single submitted outputs, may be read in two opposing ways. It can be taken to be clear evidence of a vibrant and interdisciplinary field, willing to present its contributions to research and scholarship across a wide

range of publishing outlets in cognate disciplines as well as in those closer to home, and motivated by the needs to proselytise, to satisfy co-authors and to contribute to other host communities. The less benign opposing interpretation is that not all of these are journals that habitually publish world-leading outputs, and that to some degree the promiscuity on display here may be viewed as something of an indulgence. Of course, both interpretations may contain elements of truth. Sampling the 'singleton' outputs shows that there are clearly examples of world-leading outputs in these journals, but the overall distribution of grades shows a greater percentage (62%) of 1* and 2* articles (and therefore a lower percentage of 3* and 4*) than in the total submission (47%).

Finally, the character of the journal output of UK academic Geography and Environmental Studies, as revealed here, is likely to present a challenge for any form of bibliometric analysis in the REF. One characteristic of such analysis is that it requires some kind of normalisation (van Raan *et al.* 2007) so that academic fields of different size can be judged fairly against each other. Citation rates will be higher in subject areas with large numbers of authors, unless smaller sizes are compensated by longer bibliographies, which is not normally the case. This requires a suitable benchmark, which may involve identification of a set of journals within which the level of citation globally can be used to standardise the citation rates of any selected output. Given the even balance of physical and human geography, with their substantially non-overlapping journal sets, and the fact that the top 50 journals in Table 2 only account for just over half of all journal outputs, it will clearly be difficult to establish appropriate normalisation procedures for Geography and Environmental Studies, unless each output is normalised against the average citation rate for the journal in which it is published – a practice that would immediately invite game-playing. There may also be a need to normalise against the 'age' of a publication, to allow for the variable and non-linear rate at which citations accumulate; this is considered in the next section.

Characteristics of journal publication and its assessment

It is possible to review the potential for bibliometric analysis of the outcomes of the 2008 RAE, by examining relations between peer review and citation counts for some of the submitted work. Citation counts were not used in the RAE, so those reported here had to be identified after completion of the RAE process in

Table 2 The 50 journals in which the largest numbers of articles submitted to the Geography and Environmental Studies panel in the 2008 RAE were published

Rank	No. of articles	Journal title
1	154	<i>Environment and Planning A</i>
2	125	<i>Quaternary Science Reviews</i>
3	105	<i>Transactions of the Institute of British Geographers</i>
4	94	<i>Environment and Planning D: Society and Space</i>
5	83	<i>Geoforum</i>
6	81	<i>Journal of Geophysical Research</i>
7	73	<i>Earth Surface Processes and Landforms</i>
8	70	<i>Geomorphology</i>
9	63	<i>The Holocene</i>
10	60	<i>Journal of Quaternary Science</i>
11	60	<i>Urban Studies</i>
12	56	<i>Progress in Human Geography</i>
13	48	<i>Antipode</i>
14	46	<i>Geophysical Research Letters</i>
15	46	<i>Journal of Historical Geography</i>
16	42	<i>Political Geography</i>
17	41	<i>Hydrological Processes</i>
18	40	<i>Geology</i>
19	40	<i>Remote Sensing of Environment</i>
20	38	<i>Annals of the Association of American Geographers</i>
21	38	<i>International Journal of Remote Sensing</i>
22	38	<i>Nature</i>
23	37	<i>Science</i>
24	37	<i>Science of the Total Environment</i>
25	37	<i>Water Resources Research</i>
26	36	<i>Area</i>
27	35	<i>Cultural Geographies/Ecumene</i>
28	31	<i>Social and Cultural Geography</i>
29	29	<i>Regional Studies</i>
30	28	<i>Earth and Planetary Sciences Letters</i>
31	28	<i>Journal of Economic Geography</i>
32	27	<i>Journal of Hydrology</i>
33	27	<i>Palaeogeography, Palaeoclimatology, Palaeoecology</i>
34	25	<i>Global Change Biology</i>
35	25	<i>Quaternary Research</i>
36	24	<i>Environment and Planning C: Government and Policy</i>
37	24	<i>Journal of Glaciology</i>
38	22	<i>Earth Science Reviews</i>
39	22	<i>Journal of Rural Studies</i>
40	22	<i>Population, Space and Place/International Journal of Population Geography</i>
41	21	<i>Catena</i>
42	20	<i>Health and Place</i>
43	20	<i>Journal of Paleolimnology</i>
44	20	<i>Journal of Urban and Regional Research</i>
45	20	<i>Social Sciences & Medicine</i>
46	19	<i>Journal of Archaeological Science</i>
47	19	<i>Journal of Biogeography</i>
48	19	<i>Sedimentary Geology</i>
49	18	<i>Boreas</i>
50	18	<i>Quaternary International</i>

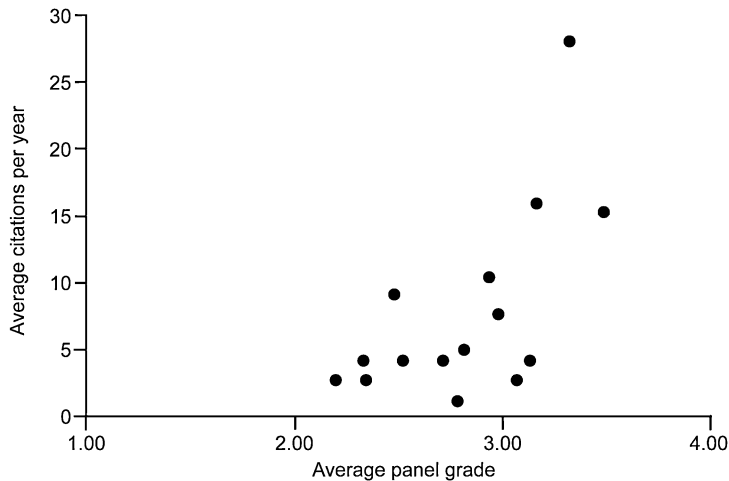


Figure 2 The average number of citations per year plotted against the average panel grade for papers returned to the RAE panel for Geography and Environmental Studies, and published in 16 journals covering human and physical geography, and general science

late 2008. A mixed sample of 16 journals from the top 50 in Table 2 was chosen for this purpose, with a total of just over 700 outputs assessed, using the Web of Knowledge and in one case the Scopus databases. These sources are appropriate for journal articles during the relevant period, although for other kinds of outputs, other databases may be more comprehensive (Meho and Yang 2007). Since one objective was to effect comparison amongst some journals, a minimum of about 30 papers was required in any one. It should be noted that although the RAE assessed work submitted by an author, it did so as individual outputs, independently of authorship, the number of authors, or the position in a list of co-authors in which the submitting author occurred (because there are no consistent and verifiable bases on which to judge the co-authors' contributions).

Journal standing and the ecological fallacy

There is often a presumption that a judgement can be made about the quality of a paper because of the standing of the journal in which it is published. This is a presumption that ignores both the fact that publication is essentially a 'pass-fail' process rather than one involving grades of quality, and the ecological fallacy in which an individual is inappropriately defined by the character of a group. In Figure 2, the limitations of this notion are seen in the relatively weak correlation between average citation rates and average peer grades for the submitted papers published in the 16 sampled

journals. There is a trend, there is a correlation, but both are weak, and mask considerable variability even after averaging for each journal. There is also no obvious pattern in the residuals. Some of the journals at the top right are well-known general science journals, but not all. Some of the low outliers are human geography journals, but there is also an American Geophysical Union journal here. Moreover, by no means all outputs published in presumed 'top' journals – general or specialised – were highly graded; peer review provided evidence of variable quality in most journals for which sample sizes were significant (see below). On this evidence it is hard to conclude that bibliometrics will necessarily be a suitable *alternative* to peer review.

The variable effect of publication date

An advantage of peer assessment over citation counts is that the former does not require normalisation either for size of sub-field or for year of publication. Consider Figure 3, which plots citation counts for the sampled papers published in each year between 2001 and 2007. Unsurprisingly, it is evident from these diagrams that the length of citation history is an important determinant of citation levels; excluding outliers, there seems a general trend in which papers increase their citation counts relatively rapidly for about 4–5 years, then citation rates decrease. This implies the need to normalise according to the age of a paper, but there is no consistency in the pattern of citation accumulation

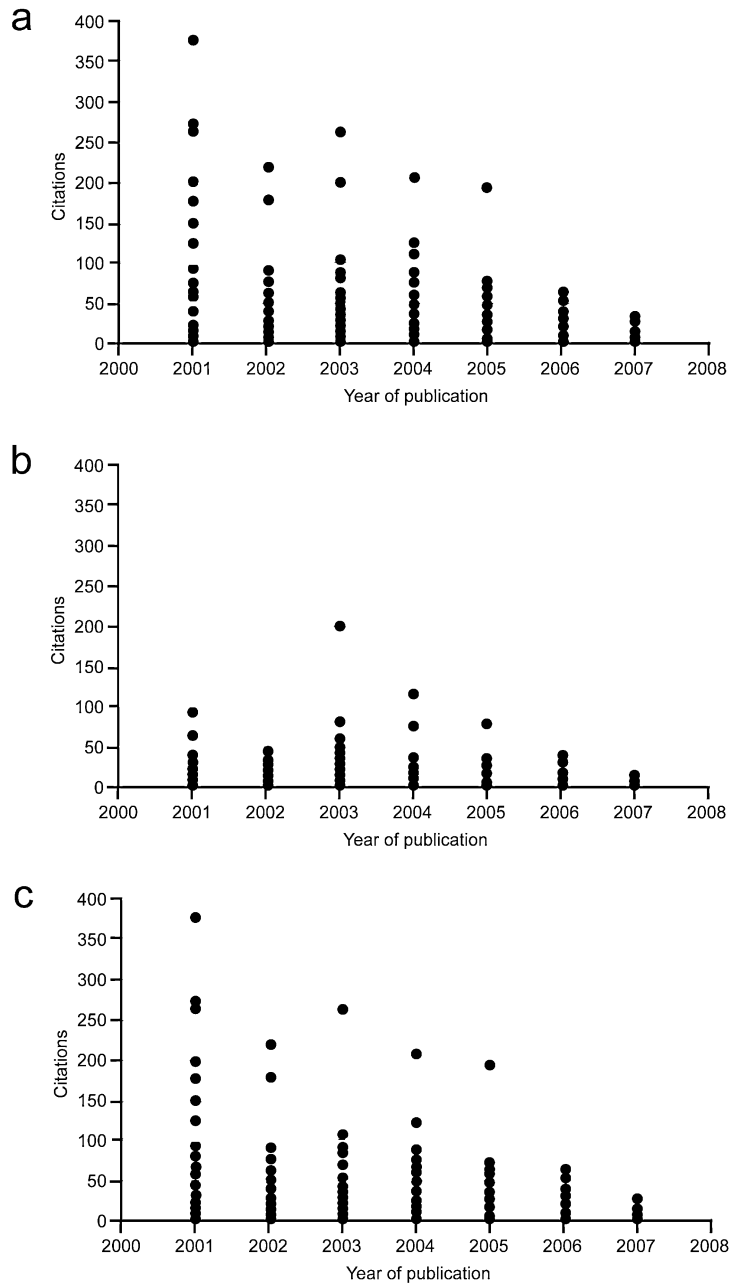


Figure 3 Citation counts for a sample of 710 outputs assessed during the 2008 RAE, published in 12 journals, plotted by publication year. (a) Total sample, (b) outputs in human geography journals, (c) outputs in physical geography journals

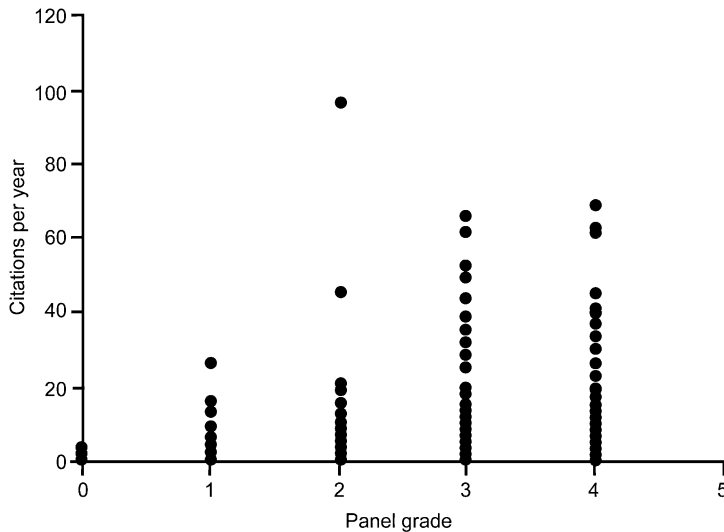


Figure 4 The relationship between citation counts per year averaged between the publication date and 2007, and the RAE grade, for a sample of 710 papers in 12 journals

between disciplines; even in Geography, there are differences in the citation levels and growth rates for human and physical geography (compare Figure 3(b) and 3(c)). By contrast, peer review, while not perfect, is better able to judge certain aspects of the potential standing of an output independently of its type, the journal in which it is published, the date of publication and the size of sub-field. It is also able to identify those cases where high citation reflects critical referencing of a poor paper, where an article has been constructed primarily to publicise research, or where deliberate selection of fashionable key words has occurred in order to attract citation. From this alone, one might conclude that bibliometric data are better interpreted in the form of raw counts as part of a peer review process, rather than being subjected to prior normalisation, relative to citation rates for a selected set of journals presumed to relate to a particular disciplinary field.

Peer review and citation rates

As suggested above, there is a correlation between peer review and citation rates. Figure 4 illustrates this in a different way, plotting the citation counts per year between the date of publication and 2007, in relation to the peer review grade on the five-point scale used in the RAE (with 'U' replaced by '0' on the abscissa). As might be expected, for any peer grade there is a column of citation counts ranging upwards from zero

(common for outputs published in 2007, and difficult to normalise). As the peer grade increases, the height of the column increases, reflecting a tendency for higher citation of those outputs rated more highly by peer assessment. This implies that there is a correlation, but it is very weak and dominated by the effect of increasing variability of citation rates as the peer grade increases. A simple, alternative test of the efficacy of citation rates as a measure of quality is therefore required. This can be achieved by combining the evidence in Figures 3 and 4. Figure 4 shows that 21 per cent of outputs in this sample were rated as 4* by peer assessment (in fact, 20% for the human geography sub-sample, and 22% for the physical geography sub-sample). If the outputs published in each year in Figure 3(b) and 3(c) are ranked by their citation counts, and the top 20 per cent or 22 per cent are identified, we can count how many 4* outputs are within these top citation-based groups (and how many lie outside). The result of this analysis shows that in the physical geography set, 53 per cent of the papers graded 4* are also within the top 22 per cent by citation counts, while in human geography, only 38 per cent of the 4* papers are in the top 20 per cent by citation counts. Citation counts capture some evidence of the peer judgement of quality, and 4* outputs are over-represented in the top fifth by citation counts. However, that top fifth by citation counts misses 50 per cent or more of the best peer-assessed outputs, and

therefore does not do a particularly good job of identifying the highest quality work as defined by peer review.

It is also evident from the outliers in Figure 4 that some outputs with high citation rates are not judged to match the criteria of the highest grade category ('world-leading'; Table 1). It is therefore useful to consider what the qualities of these papers might be. They may sometimes be papers that cross a boundary between being very good reviews and presenting the findings of original research work; if reviewing, they may do so in a summative, backward-looking fashion, rather than looking forward to set new agendas based on rigorous inference from past research results; or, sometimes, they may be published in journals that encourage pithy presentations lacking details of research methods and wider implications. In all cases, these may be very good or excellent papers, but not necessarily 'world-leading'; yet their qualities are reflected in high levels of citation. An implication of these possible distinctions is that peer review is at least capable of looking forward, while by definition, citation counts look back. This raises significant strategic questions both for the Funding Councils and for authors. In the former case, it may mean that bibliometrics will tend to be historic indicators and will increase the tendency to reward past successes; in the latter, it suggests that, sadly, there are ways of writing papers to maximise citation counts, implying that game-playing will perhaps be even more apparent if bibliometric assessment is used.

Some journal comparisons

Comparisons can be made between journals, although they are inconsistent, and their interpretation has to be treated with caution. For example, the *Transactions of the Institute of British Geographers* and the *Annals of the Association of American Geographers* both appear in the top 50 journals listed in Table 2. The former has an impact factor almost 40 per cent higher than the latter, but the average panel grade for the 105 papers published in the *Transactions* was 10 per cent below that of the 38 papers published in the *Annals*, and 16 per cent of those in the *Transactions* received 4* grades compared with 40 per cent of those in the *Annals*. This may reflect a tendency for UK-based authors to offer outputs they consider their best to an 'international' journal, with the RAE and its emphasis on 'international quality' in mind; but of course, the different sample size also has a significant effect. It could also reflect the diversity of material published in these general journals. By contrast, the differences in

average panel grades between two specialist journals, *Remote Sensing of Environment* and the *International Journal of Remote Sensing*, were more consistent with the journals' impact factors (although only the former showed a strong grade-citation pattern similar to that in Figure 4).

However, when comparing other specialist journals in UK-US pairings, such as *Earth Surface Processes and Landforms* (UK-based) with *Geomorphology* (US-based), and *Hydrological Processes* (UK) with *Water Resources Research* (US), there are no consistent differences between impact factors, citation rates and panel grades. The more specialist human geography journals also show that within these sub-disciplines, bibliometrics fail to capture quality successfully. Thus, while the *Journal of Economic Geography* performs better than *Social and Cultural Geography*, *Political Geography* and the *Journal of Historical Geography* in terms of impact factor, average panel grade and citation rate, the *Journal of Historical Geography* produces an average panel grade that belies its low impact factor and citation rates, since it is clearly publishing work whose quality is conventionally not reflected in citation. The general science journals also have different profiles. All publications in *Nature* were deemed to be of international standing (grades equal to or better than 2*), while 10 per cent of papers in *Science* did not meet this standard. *Science*, however, performed better in terms of the proportion of 4* papers (48%, compared with 39%). *Global Change Biology* outperformed both in terms of average panel grade, although its impact factor and citation rates are considerably lower. These comparisons show that the differences between rankings based on peer review and those based on bibliometrics can be quite variable, and reflect a range of influences including the publishing strategies of authors (partly being tailored to suit an assessment process they know they will face). However, it appears that peer review and citation rates are really measuring different things; perhaps, they are feeling their way towards something that might be 'quality' in the former case, and 'impact' in the latter.

Monographs

As noted above, authored books constituted a significant component of the submitted outputs, and were notable for their disproportionately high quality. There continues to be concern about the future of scholarly authored volumes deserving of the monograph label, whether this is defined by sole authorship or by detailed focus on a specific field. This concern is

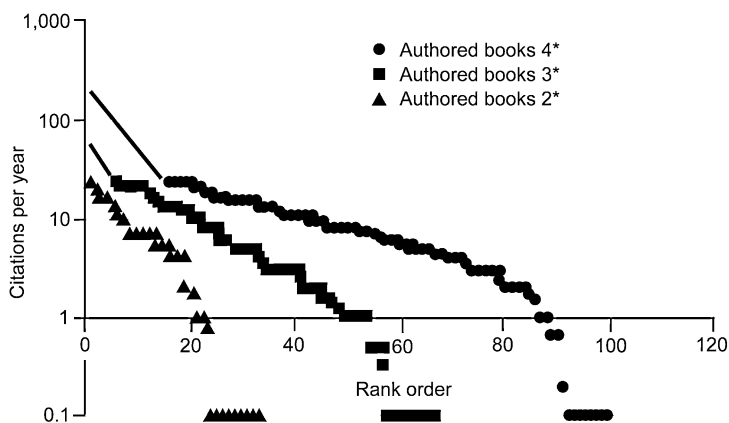


Figure 5 Authored books ranked by Google Scholar citations within peer-review grade classes (2* to 4*). The data points for the highest ranking cases have been omitted and replaced by a generalised trend, to prevent inferences about gradings of individual outputs

accentuated by the relatively poor coverage of books in the databases used for bibliometric analysis, with the attendant risk that authors cease to produce such outputs simply because their quality and impact cannot be measured. However, since books are produced to be sold, are advertised by publishers and booksellers, and are reviewed in the public domain, there are different ways in which they gain currency and presence. It is nevertheless possible to use a resource such as Google Scholar as a first approximation to a citation index for books, and Figure 5 presents some results using this source to assess citations for the books submitted to the Geography and Environmental Studies RAE panel.

Figure 5 plots the citation count against a book's rank on this criterion for each of the quality classes, from 2* to 4*. What this shows is that a very wide range of total citation occurs in each quality class, partly a reflection of different publication dates. There are some books not included because they were published in 2007 and lacked any citation (and zero cannot be plotted on the logarithmic scale). It is clear that the citation ranking for each of the lower grade classes overlaps that of the class above to a considerable degree. For example, approximately the top third of the 3* books have citation levels higher than those for the 4* book at the median rank for that class; and the top half of the 2* books have citation levels above those for the median 3* book. This suggests that even if a reliable estimate of citation numbers does become available for books, the mapping of this information onto peer assessment will be far from exact. This will

probably reflect the fact that peer review is seeking to assess the quality of research and scholarship in a book, while citation counts will also be measuring something, but not necessarily the same thing as that being referenced by peer review.

Implications

What we have attempted to do in this brief summary is to draw attention to the diversity of types of research output submitted to the Geography and Environmental Studies panel in the 2008 RAE, and to contribute to discussion about the procedures for its assessment in the light of proposals for a more metrics-based future Research Excellence Framework. Since it deals with the outputs in this context, this paper does not offer insights into the detailed processes and practices of publication, selection for submission and peer assessment. There can be no ethnography of a panel's decisionmaking processes beyond the claim that all outputs were given equality of treatment, with no prior assumption about quality based on the type of output. Furthermore, the conclusions drawn may be deemed to be mainly relevant to the UK, given the ways in which those detailed processes have been affected by the rôle of the RAE since 1986, and will continue to be affected by the REF. However, it is clear that comparable quality-related research assessments intended to drive funding in higher education take place in other countries, and the UK experience can thus be of wider interest (cf. Johnston 2006).

There are some issues on which we have been unable to comment, most notably the assessment of policy-related outputs (explicitly applied work as opposed to the merely applicable, of which there is a substantial amount). This is partly because of the apparent relative paucity of such work. Such a lack could be because its authors perceive an absence of appropriate mechanisms for assessing applied work, and decline to submit it, however much it may be explicitly welcomed by the RAE. The quality and success of applied work could be assessed by indicators of impact – for example, on practice – but this is not readily measured, even by citation data; how it could be measured remains an open question. However, there is anecdotal evidence that the best applied work is innovative, questioning of current practice, and agenda setting, and does not have to appear only in grey literature, but could be submitted as scholarship or research to a journal, for assessment under the same rules as other output. We think, in fact, that the few items explicitly submitted as ‘Research reports for external bodies’ simply represent the thin end of a considerable applied wedge that is submitted as other forms of output. Similarly, there is little that we can say about the issue of interdisciplinary research, other than that it is reviewed as any other output and treated on its own (often considerable) merits. Finally, the disproportionately high number of authored books in the 4* category is a very important finding in relation to proposals to introduce greater emphasis on metric data in future research assessments, as bibliometrics are currently weak in capturing citation of ‘Authored books’.

Considering the evidence reported here in the context of the use of metric data in the Research Excellence Framework, it is clear that measuring quality of scholarship and research using the criteria-referencing defined in Table 1 is in any case answering a different question from measurement based on bibliometrics. As noted in the Introduction, it is now emerging that all kinds of different questions about intellectual output can be assessed using new and different forms of data. Mere counting of the citations chosen for a public and political act of referencing, for example, cannot achieve an assessment of the true lineage of a piece of research in the way that private download sequencing from the Internet can. The hundreds of thousands of such click-sequences can be used to map the structure of knowledge in ways that could not have been imagined only a very few years ago (Bollen et al. 2009), and that demonstrate how the social, natural and environmental sciences are deeply intertwined. The evidence

presented above of journal-scale grade averages relative to citation rates and impact factors; of the imperfect capture of grades by citation data for individual outputs; and of the overlap of citation counts between grade classes for authored books all emphasise the differences between peer assessment and bibliometric counting, and show that one cannot simply replace the other. Indeed, it is likely that peer assessment does provide a stab at measuring some attributes of ‘quality’, while bibliometrics capture something more akin to the ‘impact’ of a publication – which are in no sense the same thing. This in turn implies that the Research Excellence Framework may need to adopt both methods and to clarify their rôles and what they measure in order to improve the overall assessment of the UK’s university-based research effort, or it seriously risks throwing the intellectual baby out in a swill of numerical bathwater.

Acknowledgements

We are grateful to Ed Hughes (HEFCE) for helpful comments on an early draft of this paper, and to Philip Stickler (Department of Geography, University of Cambridge, for the publication-quality versions of the diagrams.

Notes

- 1 Where the analysis required access to specific outputs and panel grades (for example, to collect citation data), it was completed prior to 10 December 2008, when all records of the assessment process were destroyed to ensure confidentiality. This also ensured consistency of the base period for citation data.
- 2 All the outputs and their classifications are listed at <http://submissions.rae.ac.uk/submissions/>
- 3 However, there was also some submission of human geography outputs to other social science panels.

References

- Bollen J, van de Sompel H, Hagberg A, Bettencourt L, Chute R, Rodriguez M A and Balakireva L** 2009 Clickstream data yields high-resolution maps of science *PLoS ONE* 4 (3) e4803 doi:10.1371/journal.pone.0004803
- Ferguson R I** 2003 Publication practices in physical and human geography: a comment on Nigel Thrift’s ‘The future of geography’ *Geoforum* 34 9–11
- Harvey D** 2006 Editorial: the geographies of critical geography *Transactions of the Institute of British Geographers NS* 31 409–12
- HEFCE** 2005 *Guidance on submissions* Higher Education Funding Council for England, Bristol RAE 03/2005

- Johnston R** 2006 Research Quality Assessment and Geography in Australia: can anything be learned from the UK experience? *Geographical Research* 44 1–11
- Meho L I and Yang K** 2007 Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus, and Google Scholar *Journal of the American Society for Information Science and Technology* 58 1–21
- RAE sub-panel H-32** 2008 *Subject overview report* (<http://www.rae.ac.uk/pubs/2009/ov/>) Accessed June 2009
- RAE sub-panel H-32** in preparation *The current state of UK Geography*
- Thrift N** 2002 The future of geography *Geoforum* 33 291–8
- van Raan A, Moed H and van Leeuwen T** 2007 *Scoping study on the use of bibliometric analysis to measure the quality of research in UK higher education institutions* Centre for Science and Technology Studies, Leiden
- Ward K ed** 2009 The future of research monographs: an international set of perspectives *Progress in Human Geography* 33 101–26