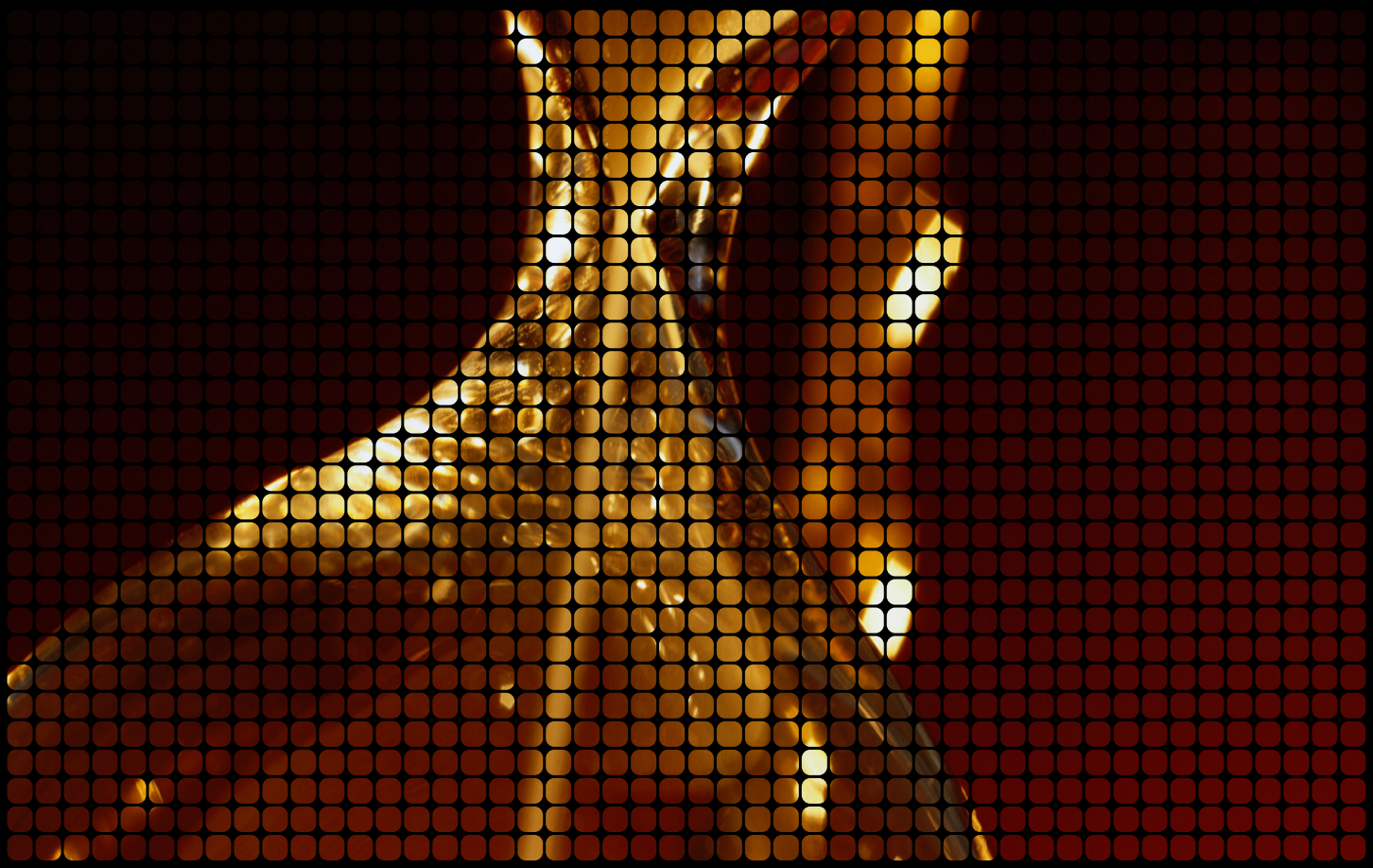

Digital Research Reports

Evidence for excellence: has the signal overtaken the substance?

An analysis of journal articles submitted to RAE2008

Jonathan Adams & Karen A Gurney

JUNE 2014

DIGITAL science

About the Authors

Jonathan Adams joined Digital Science as Chief Scientist in October 2013. Previously he was the lead founder of Evidence Ltd (2000-2009) and Director of Research Evaluation for Thomson Reuters (2009-2013). Jonathan led the 2008 review of research evaluation in New Zealand and was a member of the Australian Research Council (ARC) indicators development group for its research excellence assessment (ERA). In 2004 he chaired the EC Evaluation Monitoring Committee for Framework Programme 6. In 2006 he chaired the Monitoring Group of the European Research Fund for Coal & Steel. In 2010 he was an Expert Advisor to the interim evaluation of the EU's 7th Framework Programme for Research (FP7).

Karen Gurney was a co-founder of Evidence Ltd (2000-2009) and led the UK-based research evaluation arm of Thomson Reuters (2009-2013). She is now at Sheffield Hallam University. Karen has managed research evaluation projects for companies and research-funding agencies in Europe, North America and Asia-Pacific. She is an Associate of the Royal College of Science and has a specialist background in biomedical sciences, having worked with Leukaemia Research Fund and the Paediatric Epidemiology Group at the University of Leeds. Karen has published extensively in science journals as well as research policy.

About Digital Science

Digital Science is a technology company serving the needs of scientific research. We offer a range of scientific technology and content solutions that help make scientific research more efficient. Whether at the bench or in a research setting, our products help to simplify workflows and change the way science is done. We believe passionately that tomorrow's research will be different - and better - than today's.

Visit www.digital-science.com

This report has been published by Digital Science, a business division of Macmillan Science and Education.

The Macmillan Building, 4 Crinan Street, London, N1 9XW
info@digital-science.com

ISBN: 978-0-9929477-0-5

Evidence for excellence: has the signal overtaken the substance?

Research assessment and evaluation is used regionally, by the European Commission, nationally, in the UK and Australia, and by a growing diversity of research funders and performers. It is used to optimise current projects and programmes and to track historical trends in performance. It affects the scale and distribution of research investment, it influences research strategy and it may be decisive in research careers. But has evaluation as a well-intentioned audit been overtaken by evaluation as marketing?

We analyse past submissions to the UK's Research Assessment Exercise (RAE) to reveal what people do compared to what they say. Specifically, we look to see if a drive to 'signal' excellence has overtaken the intention to submit 'substance'. We find that researchers submit journal articles in preference to the outputs that elsewhere they say are central to their field, they skew their selection to high-impact journals, and they submit pieces from such journals even when they are not well cited and, sometimes, not even research papers. We suggest that this is because they believe that the 'brand' of a journal known to have high average impact is a better proxy 'signal' in place of real evidence of excellence.

Research assessment, publications and bibliometric indicators

Peer review is the 'gold standard' for research evaluation[i] but does not always produce a perfect outcome.[ii] It is subjective, subject to reputational bias and opaque to non-expert stakeholders. To balance this, quantitative performance indicators have come into widespread use over the last 25 years. These are not a universal solution, however, because research culture varies, as do opinions about the best indicators of 'high quality research'.[iii]

Since 1986, the UK has conducted a cyclical Research Assessment Exercise (RAE, re-launched in 2014 as the Research Excellence Framework). The RAE has been critical to both reputation and resources. It provides an indicator of esteem, through a published grade of achievement at subject level, and it provides a funding factor, when the subject grades feed into allocations to institutions. Its methodology requires peer panels to assess data on funding, postgraduate training and just four publications for each researcher. Universities and individuals should be strongly motivated to select, from the diversity of publications over an RAE census period, those four items that will convince the peer panel that each academic and each group of academics has achieved excellence.

Universities can review what their competitors submitted at past RAEs, draw inferences about what panels thought of this and adjust their own strategy in response. Tactical gaming will inevitably occur, which leads to the question of what academics – after several RAE cycles – now believe is the best signal of high quality research.

Figure 1. The relative volumes of different output types submitted to RAE2008

Bibliometrics, drawing on data on commercially indexed journal publications and citations, are the most widespread group of quantitative research performance indicators. The preferred form of publication varies, however, and while journal papers (research based articles and reviews) have priority in science, other disciplines favour other types of output: engineers have excellent conference series[iv]; the social sciences and the humanities use monographs[v]; and the visual and performing arts cross a range of media[vi]. For this reason, the journals covered by Elsevier's Scopus and Thomson Reuters *Web of Science* are seen frequently to cite books, book chapters, conferences and non-indexed journals.[vii] Most references in well-cited European social science papers are to non-indexed books and reports.[viii]

Bibliometrics are widely used in research evaluation – although they only work well for some parts of research – because they are accessible and produce indicators that seem easily understood. One is the h-index, as a single-number badge for an individual's entire portfolio yet driven less by real quality and more by field citation rates and by career length.[ix] The success of such simplistic citation measures – which hide as much as they reveal – influences the way researchers reflect on the presentation of their own work.

Since the RAE affects both reputation and resources, researchers should select and submit evidence to show their best research in the best possible light. Reports from the ESRC, AHRC, British Academy and HEFCE (referenced above) describe what people say about the best evidence of research performance in their discipline. Do these assertions differ from what they now believe will deliver the best assessment outcome? Our hypothesis is that two, dichotomous views of indicators of excellence have developed.

- What people assert reflects specific valid achievement in their field.
- What people believe actually influences peers carrying out mass review.

To test this, we examine how behaviour has changed over successive RAE cycles and how this has affected the value judgments that were made in submissions to the 2008 RAE.

Changes in submission behaviour

There has been a shift out of conference proceedings in engineering and out of scholarly monographs in the social sciences

Researchers can choose what they submit to the RAE from output types including not only journal articles but also books, book chapters, conference proceedings and patents and other formats such as videos, recordings and installations where these are core to research practice. (Figure 1)

These output types are indexed in the RAE database, by subject-based Unit of Assessment. We collated the number and frequency of outputs for clusters of subjects. This reveals that changes in submission behaviour over three successive RAE cycles (1996, 2001 and 2008) led to a progressive concentration on journal articles as the output type that is most frequently submitted.[x] The relative volume of articles grew from 62% in 1996 to 75% in 2008.

To enable this relative growth, there has been a shift out of conference proceedings in engineering and out of scholarly monographs in the social sciences. In the arts and humanities, however, the preferred mode of output remains both diverse and strongly disposed to books and chapters. (Figure 2)

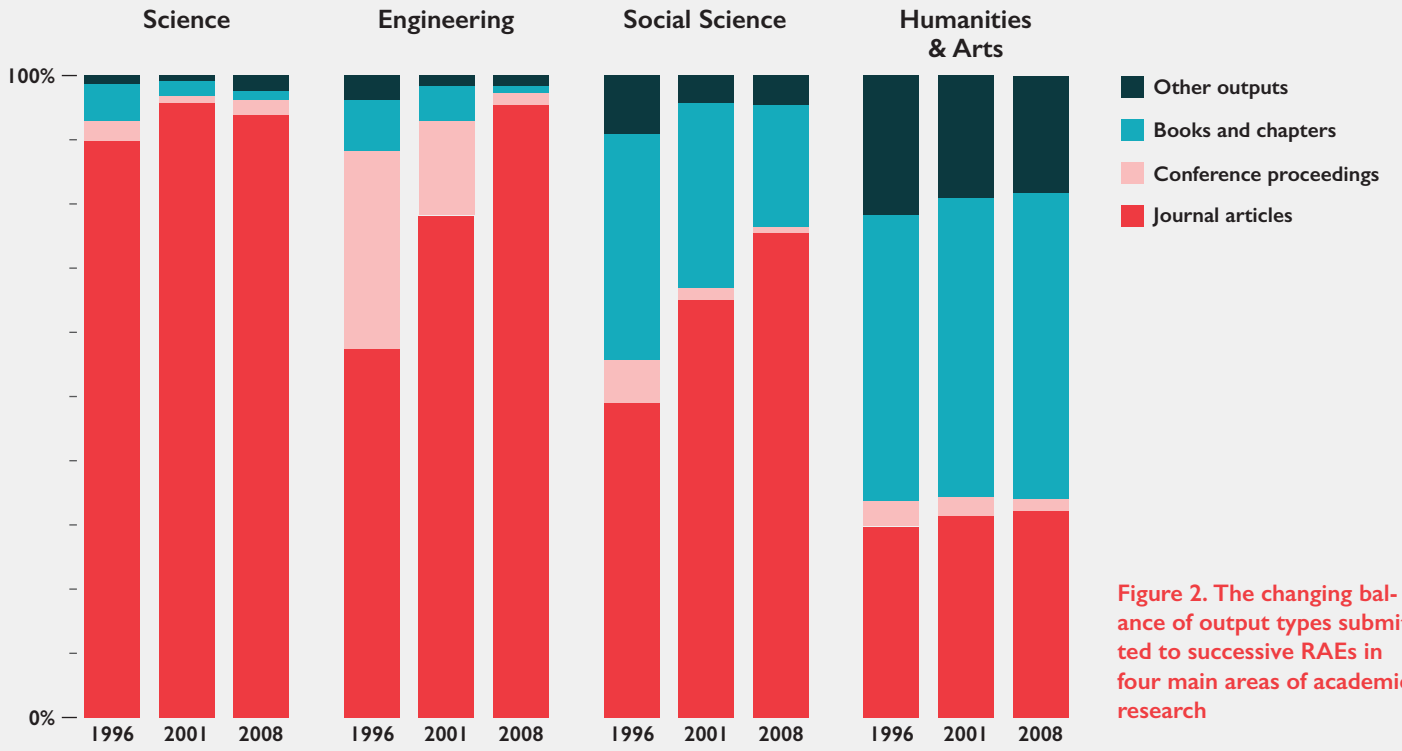

Figure 2. The changing balance of output types submitted to successive RAEs in four main areas of academic research

Conference proceedings may be less significant among engineering disciplines, but they have increased from 8% of items cited by papers in the early 1980s to over 10% now. They still play a key role in information and communication science and technology, accounting for half the material submitted for UK assessment and 20% of the references.[xi]

If it is unlikely that Figure 2 is capturing coincidental changes in research culture in engineering and social sciences. So is it instead a sign of changing researcher belief about what is most likely to be seen as evidence of achievement? If so, why would papers come to dominate?

High-quality, commercial journal citation indices may be the problem. These sources mean we can readily tally the citation counts for individual papers and calculate an average citation counts for all papers in a journal (which leads to a journal 'Impact Factor'[xii]). These counts - for a paper and the average for the journal - are widely quoted as indicators of 'academic impact'.[xiii] Similar databases are still only superficial for conference proceedings and for books.

Researchers changed what they select not what they do

Perhaps the change in behaviour over RAE cycles is evidence that numbers inexorably overcome real cultural preferences. An item from a journal of known impact becomes a simpler signal than other work that requires reading to verify the substance.

What journals did researchers submit?

The trend across RAE submissions suggests that researchers and research managers increasingly believe that journal articles, rather than conference proceedings or books, will catch the peer assessor's eye. This picture is confirmed by publication behaviour within each RAE cycle. We used data from Thomson Reuters *Web of Science* to track the annual tally of papers with UK authors. First, while the overall output of the UK is rising this is driven mainly by international collaboration and the domestic output (with UK-only authors)

remains more or less constant. Second, there are peaks in domestic productivity that are associated with the RAE censuses (Figure 3).

What do the RAE data tell us about value judgments made about the journals from which papers selected for assessment are drawn? Each paper has two potential indicators of comparative research excellence.

- We can look at the 'impact factor' of the journal in which it was published.
- We can look at the number of times the paper itself was cited, comparing that with other papers published at the same time and in the same field.

And, for each journal, we can compare the relative frequency of UK authorship with the relative frequency with which its papers were submitted to the RAE.

How much choice did UK authors have about what to submit?

RAE-eligible staff could select from a large pool of papers

Over the 2001-2007 assessment period for RAE2008, there were about 546,000 papers with at least one UK author among the journals indexed by Thomson Reuters *Web of Science*. Most of these had multiple UK authorships and, even before we add papers for RAE-eligible staff recruited from overseas, that creates many possible submission opportunities.

Around 50,000 research-active staff were submitted by their employing universities for RAE2008. Each is expected to enter up to four research outputs. They could all submit papers, in which case they would need up to 200,000 unique items, but many choose to submit other outputs (see Figure 1). In fact, at RAE2008, fewer than 165,000 papers were submitted.

So, UK researchers could be selective. They produced more than three papers in 2001-2007 (over 546,000) for every one that was submitted to RAE2008 (under 165,000). What they did submit compared to what they could submit must reflect what they thought would be compelling evidence of achievement.

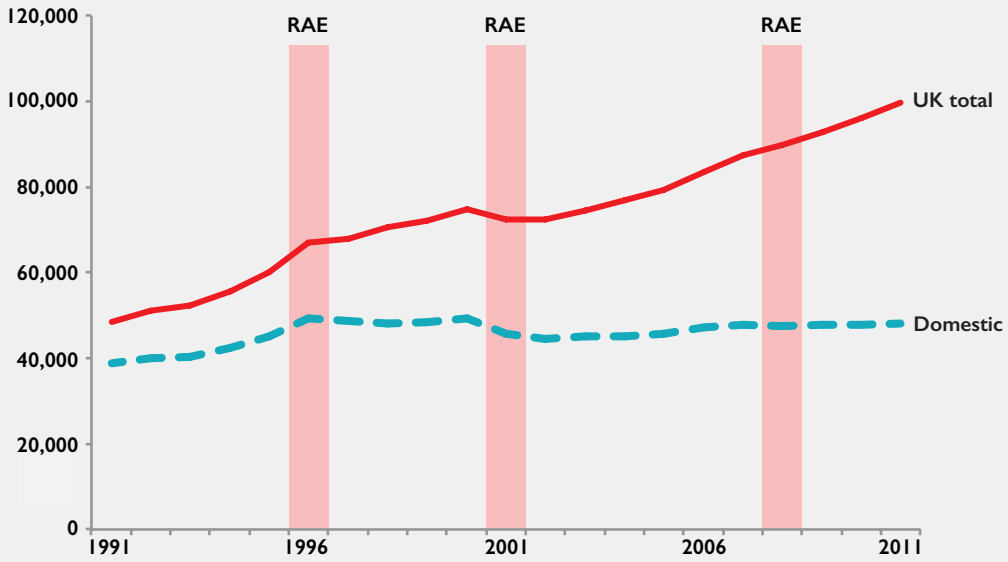

Figure 3. Annual count of papers with UK addresses - domestic papers have no international co-author. Data source: *Thomson Reuters Web of Science*

How frequently did each journal appear in the RAE outputs database?

Some journals publish more often than others and UK researchers publish relatively more in some journals than others. Some outputs are co-authored within the UK. Not every journal item is a research-based article or review, so some researchers could submit editorials and other ephemera if they had a shortfall.

The frequency of Nature papers with a UK address makes the UK second only to the USA.[xiv] During the RAE2008 census period of 2001-2007, Nature was published every week and contained 18,876 items of which 2,752 (14.6%) had at least one UK address and 1,266 of these were research articles. By comparison, during the same period, the Journal of Animal Ecology – a leading serial in zoology and ecology – was published bimonthly and carried 815 items of which 298 (37%) had a UK address and 286 of these were research articles. So, all other things being equal, while the Journal of Animal Ecology is a key serial with a greater UK-focus, we expect more records among the outputs submitted to the RAE to be from Nature. Note also that many UK Nature 'items' were not research articles.

RAE submissions concentrated in a few - often highly cited - journals.

To test for selectivity across journals, we compared RAE frequency to the available pool. To do this, we collated journal publication records in the RAE2008 database by journal title.

- 669 journals each with 44 or more records collectively accounted for half the RAE total (80,829 articles) after aggregating all variant titles that occurred five or more times.
- 14 frequently submitted journals each had more than 500 RAE2008 output records, accounting for 8.5% of journal outputs.
- The remaining records occurred in excess of 15,000 other journals at lower frequency including very many singletons of which a high proportion were untraceable.

Three of the most frequent journals (Nature, The Lancet and Science) were present with a greater number of records than there were unique UK-addressed papers (i.e. articles and reviews). One Nature paper had 'maxed out': it had authors from 12 different UK universities and each institution submitted that same paper once.

Citation impact of frequently submitted journals

Some journals had more RAE records than UK-authored papers

A common factor for the journals with an exceptional RAE/UK submission ratio (where records submitted exceeded papers published) is their average citation impact: in each case the Thomson Reuters 'Impact Factor' exceeds 30. For other journals with over 500 RAE output records, with the exception of the British Medical Journal at 14, the Impact Factor is high but did not exceed 10.[xv]

There were 68 other journals submitted at a high ratio (> 1.0) compared to the availability of UK authored papers and 29 had more than 100 RAE records. In addition to the three with over 500 records (Figure 4, top) the other 26 fell into two distinct groups: thirteen science and medicine journals of which nine had Impact Factors greater than 10 (Figure 4, middle); and nine business, management and economics journals which also had high Impact Factors relative to their field (Figure 4, bottom).

There are four other journals with a high submission ratio, falling across diverse subjects: The Historical Journal (133 RAE records); Theoretical Computer Science (154); Regional Studies (184); and Human Relations (188).

Citation impact of papers from frequently submitted journals

It might seem to make obvious sense for researchers to choose to submit papers from journals that had particularly high impact. But the peer panels are not supposed to be assessing journal quality. Instead, they are supposed to assess the quality of the individual outputs.

The calculated citation impact of a journal is an 'average' in a very skewed distribution: it is nowhere near the middle or 'median' of the citation impact of individual papers. Typically, about two-thirds of the papers will have less than the average citation count because of a smaller proportion of exceptionally highly cited items. For journals frequently submitted to the RAE, it is unlikely that all submitted items exceed the average citation impact.

Not all RAE records could be matched to original research papers

Nature provides the test example. With 1,510 RAE Output Type D (journal article) records and 1,266 UK-addressed articles published in the census period it would be possible for every eligible UK paper to be submitted to the RAE. In fact, we can only match 740 unique RAE papers to a Nature article or review, which means that 500 eligible UK Nature papers were not submitted. On the other hand, some of the 740 papers that were submitted appeared several times, so they cover 1,092 submission records.

What are the remaining 418 RAE2008 records drawn from Nature, if they cannot be matched to UK papers? They are either authored by researchers recruited from outside the UK – which is possible but perhaps unlikely for all – or they are not research papers at all. They include some of the 1,486 letters, editorials, corrections and other pieces published by Nature between 2001-2007 that had UK addresses.

Not all RAE records had above-average citations

Turning back to the 740 matched Nature papers, we find a wide spread of citation rates. More than a dozen are cited over 1,000 times and clearly mark outstanding pieces of research. But, even in Nature, not everything can be above average. At the other end of the scale, there are over 100 papers that

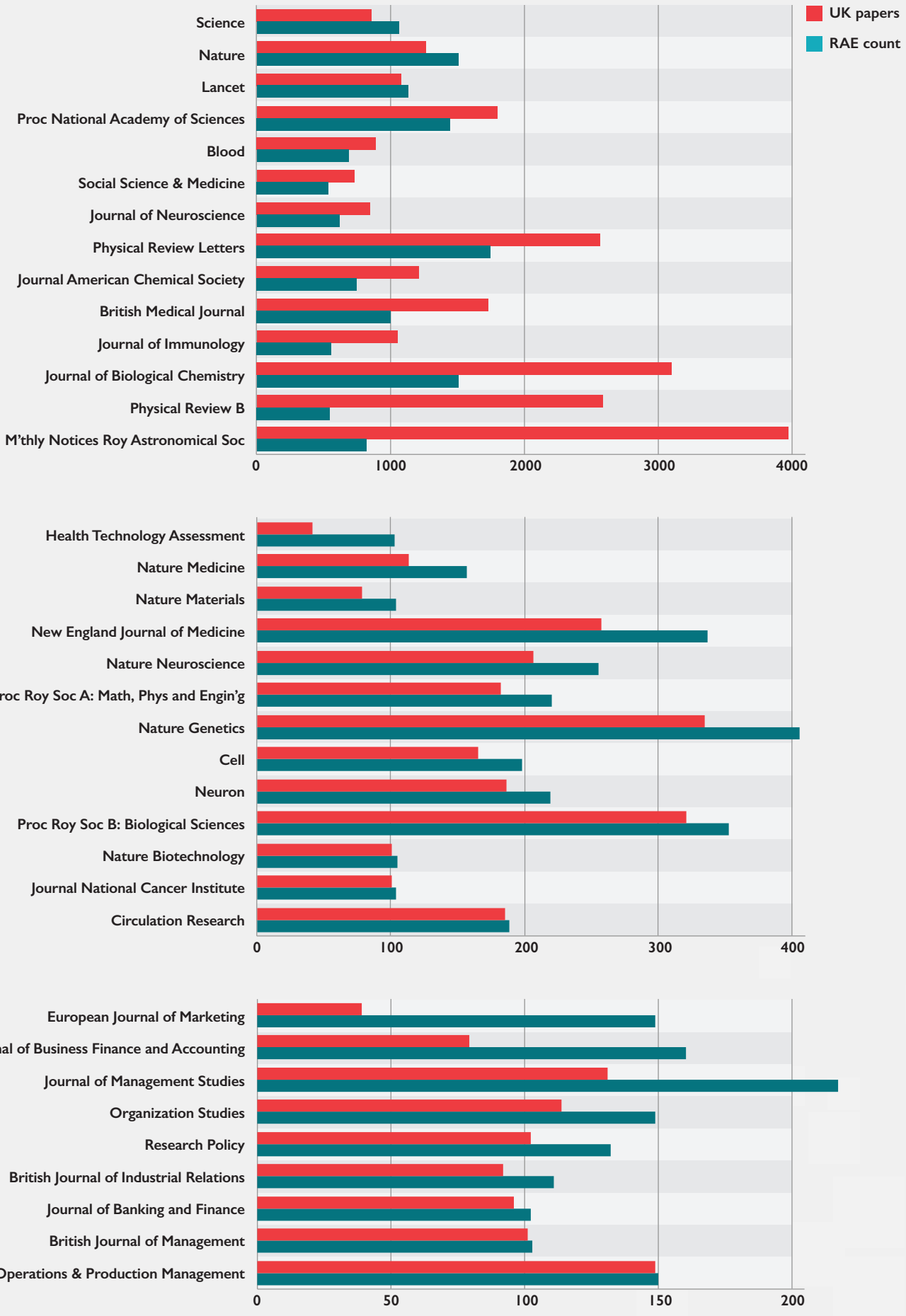

Figure 4. UK papers (2001-2007) vs RAE paper counts for journals with over 500 RAE records (top) and journals submitted more frequently than unique UK records in science and medicine (middle) and in business and management (bottom) Data source: UK papers, *Thomson Reuters Web of Science*; RAE data, HEFCE/RAE Manager.

were cited by less than one-quarter of the average number of citations for the Nature volume in which they appeared.

When we check further, we find that some researchers submit low-cited Nature papers although they do in fact have eligible papers from other journals that had more citations and a better relative impact for their discipline. Nonetheless, they chose not to deploy these in RAE2008.

Conclusions

Researchers prefer to submit items from high impact journals, ahead of well-cited and non-journal outputs

What researchers actually do under assessment differs from what surveys say they believe about the signals of research excellence. When it comes to the RAE, with the exception of the humanities, academics prioritise journals over other publications, they accelerate publication rates at RAE time, they favour journals with high average citation impact and among those journals they are persuaded that a high Impact Factor beats a convincing individual article.

The UK's RAE, rooted in traditional peer review, has evolved over almost thirty years alongside the emergence and acceptance of quantitative research evaluation based primarily on bibliometrics, which seem attractive because they lead to simple numbers. Many academics argued that journal bibliometrics do not work for them. In RAE1996, one third of engineering outputs were conference proceedings and one third of social science submissions were books.

By RAE2008, almost all engineering outputs and over two thirds of social science outputs were journal articles. That would be a massive cultural shift in barely a decade: it looks much more like a change in behaviour, not in what was being written but in what was being offered for assessment. The real substance of what academics thought was the best marker of research excellence was displaced for review purposes by outputs that gave the simplest signal of achievement.

The articles submitted in RAE2008 are a concentrated selection from journals with relatively high citation impact – either globally or more specifically in their field. For some journals, such as Nature, there were more RAE records than papers with UK addresses. Some researchers submitted editorial, letters and other ephemera – often not cited at all – instead of substantive research contributions. Why? Because the kudos of the well-cited journal was a marketing signal outweighing the individual item and outweighing even the opportunity to submit better cited papers from less prestigious journals.

UK researchers seem convinced that journal brand alone can tick the excellence box. Even if their colleagues value papers from other journals more highly, and value other output routes above journals, they behave as if signal beats substance every time when it comes to mass review. This presents a challenge for the Research Excellence Framework. In 2014, can the REF panels rise to the challenge of deep-drilling into the treasury of research publications? For the future, can the REF managers come up with a methodology that convinces the academics that it is real research achievement that wins, not the version that falls out of simplistic indicators?

i. British Academy (2007). Peer Review: the challenges for the humanities and social sciences. A British Academy report. British Academy, London.

ii. House of Commons (2011). Science & Technology Committee report on peer review.

iii. Moed, H F (2005). Citation Analysis in Research Evaluation. Springer, Netherlands. ISBN 1402037139; Butler, L. and McAllister, I. (2009), Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science. Political Studies Review, 7, 3–17.

iv. Kling, R and McKim, G. (2000). Not just a matter of time: field differences and the shaping of electronic media in supporting scientific communication. Journal of the American Society for Information Science. 51, 1306-1320; Borgmann, C L and Furner, J. (2002). Scholarly communication and bibliometrics. Annual Review of Information Science and Technology. 36, 3-72; Borner, K, Chen, C M and Boyack, K W. (2003). Visualizing knowledge domains. Annual Review of Information Science and Technology, 37, 179-255.

v. ESRC (2004) http://www.esrc.ac.uk/research/evaluation-impact/research-evaluation/bibliometrics.aspx; Archambault, E. & Vignola Gagné, E. (2004). The use of bibliometrics in the social sciences and humanities. Science--Metrix, Final Report. Prepared for the Social Sciences and Humanities Research Council of Canada (SSHRCC).

vi. Moed H F, Luwei M, Nederhof A J (2002). Towards research performance in the humanities. Library Trends, 50, 498-520. Engels, T C E, Truyken, L., Ossenblok, B and Spruyt, E.H. (2012). Changing publication patterns in the Social Sciences and Humanities, 2000–2009. Scientometrics, 93, 373-390.

vii. Professor Caroline Wagner (Ohio State University) has shown that many citations in articles from journals indexed by Thomson Reuters Science Citation Index are to work outside the index. See also: Butler, L and Visser, P. (2006) Extending citation analysis to non-source items, Scientometrics, 66, 327-343.

viii. Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: a review. Scientometrics, 66, 81-100.

ix. Hirsch, J E (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Science, 102, 16569–16572. Bornmann, L and Daniel, H D (2005). Does the h-index for ranking of scientists really work? Scientometrics, 65, 391-392.

x. Adams, J. (2007). The use of bibliometrics to measure research quality in UK higher education institutions, pp. 44. Universities UK, London. ISBN 978 1 84036 165 4

xi. Lisée, C., Larivière, V. and Archambault, É. (2008). Conference Proceedings as a Source of Scientific Information: A Bibliometric Analysis", Journal of the American Society for Information Science and Technology, 59, 1776-1784

xii. Pendlebury, D A and Adams, J. (2012). Comments on a critique of the Thomson Reuters journal impact factor. Scientometrics, 92, 395-401. (Thomson Reuters calculates the journal Impact Factor (IF) as part of the Web of Science. Elsevier calculates the Source Normalised Impact per Paper (SNIP) as part of the Scopus database.)

xiii. Research also has economic and social impact but indicators are difficult to determine.

xiv. Nature Publishing Index (2012).

xv. The numeric value of Thomson Reuters journal Impact Factor (IF) changes in response to disciplinary citation cultures. The median IF in Immunology and in Biochemistry & Molecular Biology (disciplines with high citation rates) is around 3 whereas in Chemical and in Mechanical Engineering (where citation rates are lower) it is around 1. Mainstream physics and chemistry lie between these, around 1.5-2. IFs above 10 are always at the high end of the range