

Dispelling the myths surrounding the Research Excellence Framework

Paul Trevorrow¹ and Dietrich A. Volmer²

¹Executive Journals Editor, Wiley-Blackwell, Chichester, UK

²Editor, RCM, Universität des Saarlandes, Saarbrücken, Germany

In a recent discussion with a colleague from the UK, we learned that scientists in his department are currently under pressure to submit their work to the highest impact journal they can in order to have the greatest impact in the upcoming Research Exercise Framework (REF) for the UK's academic units. Unfortunately, he pointed out that, under these constraints, *RCM* – as well as *JASMS*, *JMS*, *IJMS* and *EJMS* – does not fit the journal profile to which they were asked to publish. Instead, *Analytical Chemistry* and similar multidisciplinary journals with higher impact factors are favoured by the university.

For us at RCM, this is an interesting and somewhat frustrating situation. Does it mean, for mass spectrometry, that articles from the UK may be re-routed to more generalist journals such as *JACS* and *Analytical Chemistry* rather than the core mass spectrometry journals? And if that is the case, how are authors going to pitch (or reformat) their mass spectrometry articles, so a generalist journal will accept them?

Let's be honest about this, most of the typical *RCM*, *JASMS*, *JMS*, *IJMS* or *EJMS* articles will not find favor with editors of generalist analytical chemistry journals because of the very specialized nature of the research. Of course, this is exactly the reason for the existence of community journals such as *RCM* and other mass spectrometry journals.

Interestingly, among the many articles and opinions on the subject of 'impact' assessment, we came across the following statement by Zwahlen *et al.* who make particular reference to the UK situation, "We share... concern about the epidemic of impactitis, which appears to rage in Germany and elsewhere. One country which appears to be to some extent immune against this disease is the United Kingdom".^[1] Well, this immunity certainly does not appear to be present in every academic institution in the UK.

The above situation has prompted us to take a closer look at the importance of impact factors, modern research assessments tools and the factors and interactions that ultimately determine the 'impact' of generalist, multidisciplinary *versus* community journals, with particular reference to the upcoming REF exercise.

RESEARCH ASSESSMENT EXERCISES

National research assessments (NRAs) have been on the UK calendar for almost three decades. Since the formalisation of the first Research Assessment Exercise (RAE) in 1986,

subsequent evaluative exercises have been instigated in 1992, 1996, 2001 and 2008. The latest and current mutation is the Research Excellence Framework (REF) due for completion in 2014. These NRAs were implemented by the UK's Higher Education Funding Council (HEFCE), a quasi-governmental agency, tasked with the function of awarding public 'block grant' funds to teaching and research in UK universities.^[2] While there are mechanical variations to the exercises, they rely commonly on a 'unit of assessment' (a department or organising unit) and the provision of a defined number of research works for peer-panel evaluation. Since the 1996 RAE through to the recent REF, four research items are required for evaluation, typically in the form of four journal articles. While there are other evaluative criteria, the provision and evaluation of journal articles in the RAE and REF concern us here and form the direction of this perspective.

HOW DO THE RAE AND REF EVALUATE JOURNAL ARTICLES?

Throughout the RAEs taking place between 1986 and 2008, peer-review was used to assess 'research quality', marked by predefined criteria such as originality, relevance and potential impact on society.^[3] Peer-review was conducted by a panel of experts required to evaluate 'unit' research by reading a sample of publications submitted for assessment. In the RAE exercises, the use of journal metrics (citation counts, impact factor etc.) as quality proxies was explicitly ruled out. Prior to the completion of the 2008 RAE, a report from the UK Treasury suggested a more economical approach for the future of funding: the use of metrics.^[3] While the Treasury were keen to allocate funding, based on metrics, this was not adopted **in full** by the HEFCE in their blueprints for the REF. The reluctance of the HEFCE to indulge fully in a bibliometrically measured evaluation process is particularly evident in the HEFCE's guidance regarding the use of bibliometric data:

71. In all UOAs panels will assess outputs through a process of expert review. In doing so, panels may make use of additional information – whether provided by HEIs in their submissions, and/or citation data provided by the REF team – to inform their judgements. In all cases expert review will be the primary means of assessment.

In Part 2, the panels set out the following:

- a. Whether they will make any use of citation data in the assessment.
 - b. Whether they require any of the types of additional information listed in 'guidance on submissions' (paragraph 127).
 - c. How they will use any such information to inform their assessments.
72. Those panels using citation data will do so within the framework set out in 'guidance on submissions' (paragraphs 131 to 136). In particular, they will consider the number of times an output has been cited as additional information about the academic significance of submitted outputs. No panel will make use of journal rankings or journal impact factors in the assessment. Panels will continue to rely on expert review as the primary means of assessing outputs, in order to reach rounded judgements about the full range of assessment criteria ('originality, significance and rigour'). They will also recognise the significance of outputs beyond academia wherever appropriate, and will assess all outputs on an equal basis, regardless of whether or not citation data is available for them. They will recognise the limited value of citation data for recently published outputs, the variable citation patterns for different fields of research, the possibility of 'negative citations', and the limitations of such data for outputs in languages other than English. Panels will also be instructed to have due regard to the potential equality implications of using citation data as additional information.
73. Given the limited role of citation data in the assessment, the funding bodies do not sanction or recommend that HEIs rely on citation information to inform the selection of staff or outputs for inclusion in their submissions (see 'guidance on submissions', paragraph 136).^[4]

Excerpt from "Consultation on draft panel criteria and working methods" republished with permission from Ref. [4].

In as much, the REF endorses the use of metrics as an informative mechanism secondary to peer-panelling. Once more, as with the RAE, impact factors and journal rankings are explicitly ruled out of the working framework.

So, are measurement criteria in the REF and previous RAEs independent of impact factors and journal rankings?

The question is a moot point; one of the oft-reported misgivings of peer-review in general is its potential for subjectivity. While impact factors do not feature formally in evaluations by the RAE or REF, it is believed that subjectivity caused by the inherent status ranking of journals is inescapable (informed in part by journal ranking mechanisms such as the impact factor)^[5] which results in papers published in some titles being evaluated more highly than those published elsewhere.^[6] The thinking follows that, for a peer-panel, an article published in *Nature* is perceived as being of different quality than if published in a community journal such as the *Journal of the American Society for Mass Spectrometry*.

One must remember that peer panellists, in this epoch of metricisation, are not blind to the mechanics of publishing or the shortcomings of ranking mechanisms. As Broadbent, a peer panellist for the 2001 and 2008 RAEs in the field of Accounting & Finance and Business & Management comments: "The journal placing was not taken as a mark of the quality per se and the variability of papers within journals was also recognised. Importantly, a paper appearing in a less well known journal was not penalised for its placing. . . not all work in the better known journals was of the highest quality".^[7]

Oddly enough, while impact factor rankings are explicitly written out of the RAE and REF they play a definite role in the selection of publication by researchers and department heads. As revealed by Professor Paul, RAE panellist for Business & Management Studies and Library & Information Management; "it was 'common knowledge' that universities had used ranked lists of journals to decide which papers to submit to the RAE".^[8] Moreover, it is no secret that department heads have advised researchers to publish in high impact factor journals. Anecdotal evidence of this was collected during the 2008 RAE. With the 2014 REF, the missive has been once more recycled, a directive that appears to be relatively widespread amongst UK chemistry departments.

What is interesting with the "Nature or nothing" approach to article submission is the assumed correlation of journal impact factors with the quality of the submitted research. This is a perception that seems to be more prevalent in the research environment, those who are submitting their research, than it does in the evaluative environment of the REF, those that are evaluating the research. The use of the impact factor as a proxy of article quality is a gross misappropriation, a misappropriation against which even the impact factor's founding father, Eugene Garfield, has lobbied in numerous articles.^[9-13]

Similarly, in their excellent analysis of the subject, Smeyers and Burbules point out "As journal editors, we both know that the most-cited articles are not necessarily the ones we consider the best and most important ones we have published".^[14] As editors of RCM, we certainly observe the same for our journal.

And let's face it: most authors of papers in high impact factor journals benefit from the 'halo-effect' of those few outstanding and highly cited articles, as only a small number of papers generates the majority of citations. Impact factors can be viewed as citation ratios for journals over a given census and target period and simply do not reflect the quality of a single paper or a single author, as pointed out by Monastersky.^[15] Monastersky gives *Nature* as an example where, for the year 2004, a quarter of the articles resulted in 89% of the citations, so a vast majority of the articles received far fewer citations than reflected by the impact factor.^[15] In an analysis of all journals in the Spectroscopy category contributing to the 2008 impact factors of their host journals, skewness is illustrated^[16]: 50% of journals citations are attributed to 11% of articles, 96% of the citations are from 50% of the articles with 44% of the articles remaining uncited (Fig. 1). Thus, the danger in applying the impact factor to the value of all articles is highlighted. It accentuates the value of a large number of items that achieve zero citations and grossly attenuates those that are citation rich.

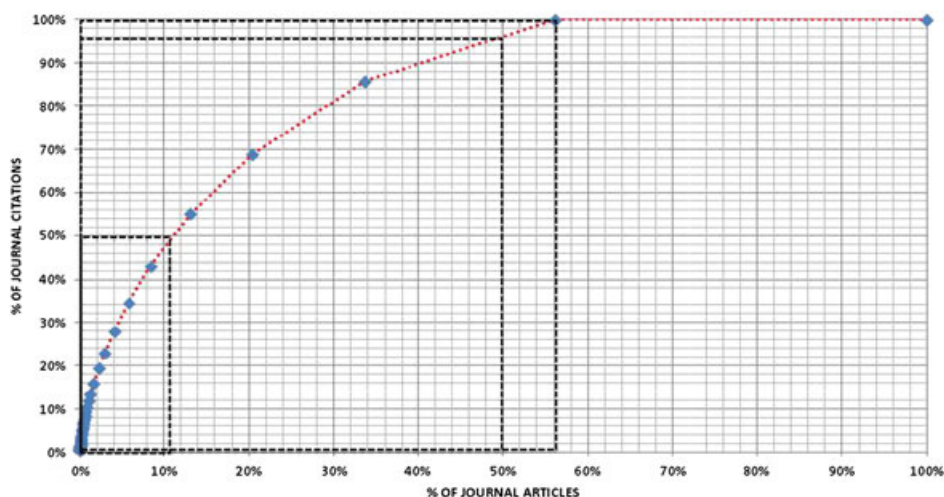


Figure 1. Percentage of citations as a function of journal articles published in 2006 and 2007 and cited in 2008 (15,506 articles) within the Spectroscopy category of Thomson Reuters®, *Journal Citation Reports*®.

A second interesting facet to this cultural phenomenon is that while the impact factor is used to determine journal selection, the HEFCE publicly announces Elsevier as their sole provider of bibliometric data. Elsevier do not publish impact factors! The impact factor remains exclusively a product of Thomson Reuters. The two are mutually exclusive databases with different coverage and inclusion criteria. If rankings were to come into play arguably they would utilise Elsevier’s ranking mechanisms; the SNIP and the SJR. To use a metric in the selection of journals that plays no part in the evaluation of the research submitted to the REF suggests some misunderstanding in the processes. Whether this can be dismissed as misunderstanding or whether the impact factor is being used as an approximate yard-stick for an imagined quality is uncertain; however, the effect of journal streaming upon research submission should be viewed in the context of the measurement apparatus of the evaluating body.

While the use of metrics will not to be used wholesale across every discipline, where metrics are appropriated peer-panellists are advised to “consider the number of times an output has been cited as additional information about the academic significance of submitted outputs”.^[4] As such, the bibliometric influence on the evaluation process will be measured by citations in a given period among the citing and cited network of the Scopus database. To reiterate the point once more, the Scopus database relies on a different network of cited and citing material from the impact factor and from the Web of Science. Measuring one’s article citation using Web of Science as a proxy will yield different citation counts due to differences in coverage.

In terms of receiving citations in a given period (it must be noted that the citation capture window is not stated by the HEFCE) two important considerations should be taken into account: the time in which the article has to accrue citations, and the article’s accessibility and visibility by those who are likely to cite it. As is demonstrated in the citation profile of all research articles in the Spectroscopy category of Thomson Reuters®, *Journal Citation Reports*® (*JCR*) citations to articles tend to peak in the 3rd year following publication (Fig. 2).

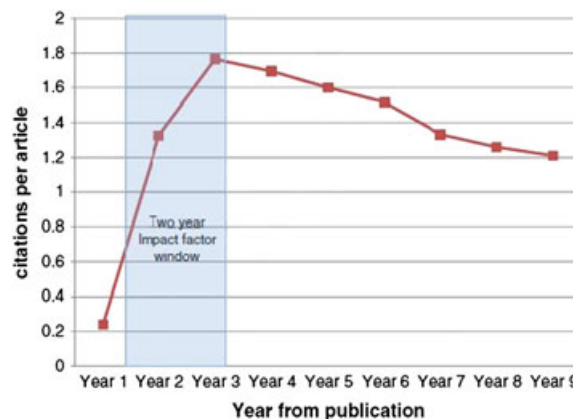


Figure 2. Average number of citations per year for all research articles published in *Journal Citation Reports*® Spectroscopy subject area category over a nine-year period.^[18]

As such, there is a potential advantage in temporal proximity to the 3rd year ‘sweet spot’ at the time of evaluation. That is of course if the evaluation is measured by citations within a temporal period. If “citations to date” is the determinant then advantage will be achieved in publishing beyond the ‘sweet spot’.

Bibliometric indicators are temporally sensitive. Not only do we need to allow time for an article to be published (publication-lag) but we also need to allow time for the information to disseminate. In order to accrue citations, the publication cycle must repeat once more to enable subsequent articles to cite the target and then await the citation linkage within the evaluated database. In such an environment, getting articles to the appropriate market quickly in a citable format and recycling the process swiftly to enable these citation linkages is an advantage – an advantage held by very few journals.

An important criterion enabling community journals to function is peer-ship. The mass spectrometry community enjoy the offerings of *RCM*, *JASMS*, *JMS*, *IJMS* and *EJMS* because the journals are among peers, among like-minded

people. For communities such as ours this gives us several advantages. It offers us immediate access to expertise – the editors are known entities whom we trust to disseminate our research for review by peers who can swiftly validate a submission or offer invaluable improvements that maximise the ‘impact’ of the final article. The peer’s proximity to the literature and the ability to glean a first glimpse is advantageous in that it introduces an ‘early-view’ effect to the citation network. In more generalist journals the network of peer-reviewers is diluted among various disciplines with a smaller concentration of expertise. That is not to say that the peer-review process is any more or less stringent between journals with differing impact factors, as Molinie and Bodenhausen colourfully point out:

“If asked to review a paper, we do not pay more attention when the request comes from *Science* than from the *Journal of Magnetic Resonance*. On the contrary! Since the likelihood that a paper actually will be accepted in *Science* appears slim, it is all too tempting for a referee to deliver a superficial review. Worse, the ‘generalist’ editors of non-specialist journals do not know whom to ask. As a result, it is not rare to read papers in *Science* that are as muddled in their argumentation as spectacular in their claims. Such papers would never be accepted by the *Journal of Magnetic Resonance*! In fact, many articles that are published in *Science* are very ephemeral, while more fundamental long-lasting papers can only be found in the specialized literature”.^[17]

It is disheartening to think that the specialism that we offer, collectively as a cohort of community mass spectrometry journals, adds buoyancy to a measurement system in which we are discounted.

We conclude this perspective with our opinion that the journal impact factor undoubtedly has its roles and uses but the utility of this metric as a proxy for article quality in an evaluation system that is measured among different citation criteria is imprudent and incongruous to the parameters of the framework in which it is placed. With the REF, and in terms of bibliometrics, departments might benefit by looking more closely at the mechanics of how the citation networks function and determine the characteristics that enable articles to maximise times cited within the evaluation window. A publishing strategy based on the perception of a single metric may not be the best criterion to maximise one’s performance within the most recent of UK national evaluation exercises.

REFERENCES

- [1] M. Zwahlen, C. Junker, M. Egger. Commentary II – The journal impact factor in the evaluation of research quality: villain, scapegoat or innocent bystander? *Soz. Präventivmed.* **2004**, *49*, 19.
- [2] D. Otley. Research assessment in the UK: an overview of 1992–2008. *Aust. Account. Rev.* **2010**, *20*, 3.
- [3] J. Taylor. The Assessment of research quality in UK universities: peer review or metrics? *Br. J. Manag.* **2011**, *22*, 202.
- [4] Anon. Consultation on draft panel criteria and working methods. (Ed: H.E.F.C. England). London, **2011**.
- [5] D. Northcott, S. Linacre. Producing spaces for academic discourse: the impact of research assessment exercises and journal quality rankings. *Aust. Account. Rev.* **2010**, *20*, 38.
- [6] R. Johnston. Research quality assessment and geography in Australia research quality assessment and geography in Australia: can anything be learned from the UK experience? *Geogr. Res.* **2006**, *44*, 1.
- [7] J. Broadbent. The UK Research Assessment Exercise: performance measurement and resource allocation. *Aust. Account. Rev.* **2010**, *20*, 14.
- [8] Z. Corbyn, in Times Higher Education Supplement, London, **2008**.
- [9] E. Garfield. How can impact factors be improved? *Br. Med. J.* **1996**, *313*, 411.
- [10] E. Garfield. The meaning of the Impact Factor. *Int. J. Clin. Health Psychol.* **2003**, *3*, 7.
- [11] E. Garfield. The Impact Factor and using it correctly. *Unfallchirurg* **1998**, *101*, 413.
- [12] E. Garfield. The history and meaning of the journal impact factor. *J. Am. Med. Assn.* **2006**, *295*, 90.
- [13] E. Garfield. Evaluating research – do bibliometric indicators provide the best measures - introduction to a review of bibliometric and other science indicators and their role in research evaluation (reprinted). *Curr. Contents* **1989**, *14*, 3.
- [14] P. Smeyers, N. C. Burbules. How to improve your impact factor: questioning the quantification of academic quality. *J. Philos. Educ.* **2011**, *45*, 1.
- [15] R. Monastersky. The number that’s devouring science. *The Chronicle of Higher Education* **2005**, 52.
- [16] P. O. Seglen. The skewness of science. *J. Am. Soc. Inform. Sci.* **1992**, *43*, 628.
- [17] A. Molinie, G. Bodenhausen. Bibliometrics as weapons of mass citation. *Chimia* **2010**, *64*, 78.
- [18] P. Trevorrow, M. Thevis. Drug Testing and Analysis’ one-year impact factor. *Drug Test. Anal.* **2012**, *4*, 1.