

Independent review of the role of metrics in research assessment: Call for evidence

Summary of areas for advice

1. This template provides a summary of the areas on which the steering group is seeking advice. You may use this template to respond with your advice if you would find it helpful.
2. Please send responses to metrics@hefce.ac.uk by **noon on 30 June 2014**.

Independent review of the role of metrics in research assessment: Call for evidence	
Name: Professor Nigel Vincent	Job title: Vice-President for Research & Higher Education Policy
Organisation: The British Academy	
E-mail: nigel.vincent@manchester.ac.uk	Phone: 07932 770508
Identifying useful metrics for research assessment:	
<ul style="list-style-type: none">• What empirical evidence (qualitative or quantitative) is needed for the evaluation of research, research outputs and career decisions?• What metric indicators are currently useful for the assessment of research outputs, research impacts and research environments?• What new metrics, not readily available currently, might be useful in the future?• Are there aspects of metrics that could be applied to research from different disciplines?• What are the implications of the disciplinary differences in practices and norms of research culture for the use of metrics?• What are the best sources for bibliometric data? What evidence supports the reliability of these sources?• What evidence supports the use of metrics as good indicators of research quality?• Is there evidence for the move to more open access to the research literature to enable new metrics to be used or enhance the usefulness of existing metrics?	
<i>This response is submitted on behalf of the British Academy and the disciplines which it represents, namely the whole spread of Humanities and Social Sciences (henceforth HSS). Some of the arguments and suggestions developed here apply across the full</i>	

spectrum of STEM and HSS disciplines, some only to HSS and some to subsets of H and/or SS disciplines.

*In this connection the first thing to say is that **there is a consensus across all disciplines that the most reliable way to assess research is by means of peer review**. This was made clear in submissions to earlier consultations by a range of national bodies including both the British Academy and the Royal Society and is reflected in the way the criteria have been drawn up across all four main panels in REF 2014. There is no subsequent evidence that we are aware of which undermines this cross-disciplinary consensus. It is true that peer review is not without its weaknesses; criticisms commonly made are that it can be conservative and subjective. However, a system of evaluation by panels, as in RAE and REF, involving multiple readers and the possibility of discussion in contested cases goes a long way towards neutralising these potential weaknesses. In any case, the risks posed by peer review are far outweighed by those inherent in bibliometrics, which include:*

- *undeservedly high citation scores for poor work which has been the object of widespread criticism*
- *lack of availability of comparable data for material published in languages other than English*
- *the danger that disciplines where citation occurs over longer timespans may be discriminated against*
- *lack of metrics for work published in book form, whether monographs, collections of essays or critical editions*
- *the fact that high citation scores can be due to the popularity of a given topic or sub-field as much as to the quality of the item in question.*

The recent moves towards open access (OA) provide an additional complicating factor. It is rightly argued that OA can be expected to increase download figures and citation scores. However, since the move to OA will vary in kind ('gold' vs 'green') and in speed of adoption across disciplines, it is likely that in the short to medium term bibliometric scores will be distorted, to the disadvantage of HSS disciplines.

*There is no doubt that, to different degrees in different disciplines, some metrics may be of use, but even so the various metrics will need to be balanced against each other and this will require human intervention. In short, metrics, even where they are of relevance, are a form of evidence and evidence always needs to be weighed and assessed by juries and judges. **Metrics may inform the assessments of specialist panels but they cannot substitute for them.***

How should metrics be used in research assessment?

- What examples are there of the use of metrics in research assessment?

- To what extent is it possible to use metrics to capture the quality and significance of research?
- Are there disciplines in which metrics could usefully play a greater or lesser role? What evidence is there to support or refute this?
- How does the level at which metrics are calculated (nation, institution, research unit, journal, individual) impact on their usefulness and robustness?

1. **Journal impact factors:** *Whatever other conclusions are reached about various kinds of metric, there is one metric where there is a clear and unequivocal international consensus among researchers of all disciplines: **the journal impact factor should never figure in any kind of research evaluation whether of individuals or of institutions.** This is one of the principal recommendations of the San Francisco Declaration on Research Assessment (DORA), whose signatories include not only HEFCE itself but also The British Academy, The Royal Society and The Wellcome Trust, as well as a broad range of international academies and research funders. It is noteworthy that in the 2014 REF all four main panels make clear that in evaluating outputs they will not have recourse to journal impact factors. We strongly urge that this policy should be continued in future research evaluation exercises.*
2. **Citation indices:** *these are also of limited value in many HSS fields, particularly those which make extensive use of monographs and book chapters rather than journal articles as publication types, since, as noted above, these types of output typically do not enter into bibliometric calculations. RAE 2008 data suggest that in a number of H disciplines, including literary, historical and cultural studies, more than half the work judged to be excellent was in the form of a book or a chapter. It will be interesting to see what the comparable results are from REF 2014. However, informal surveys among submitting institutions suggest that they are unlikely to be significantly different for the good reason that these forms of output constitute the internationally defined standard of excellence. If excessive reliance on citation indices within the UK were to force scholars away from these modes of publication, this would have the inevitable consequence of reducing the international standing of UK-based research in these fields.*
3. **Individual indices:** *the reservations set out above about indices at the level of outputs also apply within the HSS fields to those metrics such as the H-index or GoogleScholar which are constructed over the lifetime of an individual. Since such metrics are constructed by aggregating output level data, they only serve to compound the problems identified above and are therefore to be avoided.*
4. **Institutional indices:** *there are two difficulties here. First, such indices are in the main constructed out of the indices associated with the individuals employed within an institution so that the problems noted above about the creation of individual indices are simply aggravated. Second, institutions have distinctive disciplinary profiles. Since metrics are, as we have argued, not uniform across different domains of research, it follows that it will be difficult, if not impossible, to compare whole institutions unless they happen to have an identical internal configuration of disciplines. Attempting to construct indices at the institutional*

level for the most part creates a spurious sense of comparability.

'Gaming' and strategic use of metrics:

- What evidence exists around the strategic behaviour of researchers, research managers and publishers responding to specific metrics?
- Has strategic behaviour invalidated the use of metrics and/or led to unacceptable effects?
- What are the risks that some groups within the academic community might be disproportionately disadvantaged by the use of metrics for research assessment and management?
- What can be done to minimise 'gaming' and ensure the use of metrics is as objective and fit-for-purpose as possible?

It is important to distinguish 'gaming' and strategic use of metrics. If the metrics are good and reliable ones for particular disciplines, then it is highly likely that institutions and individuals will develop strategies around them. If for example grant capture or numbers of doctoral students are regarded as significant indicators within some disciplinary communities, then it is hard to see how, or indeed why, the development of appropriate strategies should be discouraged. Two things however must be avoided at all costs:

- The extrapolation of metrics from fields where they are appropriate to those where they are not. Although there is of course room for large projects in fields such as Classics, History and Literary Studies, they do not dominate the field in the way they do in the natural sciences, and it would be wrong to develop evaluation measures that would force such fields to evolve in an inappropriate direction.*
- Gaming, which is a deliberate attempt to misrepresent the extent of an individual's or a department's activities through artificial inflation of for example citation indices. Existing bibliometric methods already seek to deal with excessive self-citation or citation cartels. Purely online resources such as Google scholar are by contrast much easier to manipulate and much harder to guard against. The only sure way is, once again, to place the final decisions in the hands of panels, who can judge when one metric is out of line with other evidence and thus decide to discount it.*

There is a further danger that excessive use of metrics may disadvantage particular groups. Studies have identified younger scholars, women and ethnic minorities as being particularly at risk in this respect. These risks are exacerbated in fields where the burden of research and publication falls on the individual rather than a team.

International perspective:

In addressing the issues and questions above, please include relevant evidence and examples from outside of the UK, where appropriate.

In relation to the international dimension of metrics there are two logically distinct sub-cases which need to be considered separately:

- 1. The first of these concerns disciplines where the professional community is international and recognition within that community is achieved through an individual's or a group's presence at international conferences and in journals which are read and consulted all around the world. This is generally the case in STEM disciplines and some HSS fields such as Economics and Psychology. Here there is undoubtedly a role for metrics but only as moderated and balanced by peer review. And even then, as we have said, some types of metric, in particular journal impact factors, should be excluded.*
- 2. More problematic is the second case of disciplines in which not only is most publication in the form of books and chapters, as discussed above, but in addition, even when publication is in article form, it is frequently in very specialist journals published outside the main UK-USA axis and often in languages other than English. Fields that are affected in this respect include the study of literary traditions other than the Anglophone ones, Classical languages and literatures, archaeology, plus some areas of linguistics, philology and history. These are all fields in which the best UK scholars and departments are held in very high regard around the world. This high esteem is due to the fact that fellow specialists have read the outputs in question and have formed positive judgements about them. In other words they have conducted peer review and will continue to do so. UK national based research assessment methods should do likewise.*

Would you be interested in participating in a workshop/event to discuss the use of metrics in research assessment and management? Yes