

Royal Historical Society

Response to Stern Review of the REF

1. What changes to existing processes could more efficiently or more accurately assess the outputs, impacts and contexts of research in order to allocate QR? Should the definition of impact be broadened or refined? Is there scope for more or different use of metrics in any areas?

It is essential that an exercise such as the REF commands wide support within the academic community and that its conclusions are respected. This is currently a clear strength and would be compromised by wider use of metrics, which simply do not work across the board. A main finding of *The Metric Tide*

<http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title> is that, in contrast to peer review, academics are sceptical of metrics, which are particularly problematic when assessing outputs in the Arts and Humanities. In terms of historical scholarship, there are no current measures which provide reliable data, and this is not likely to change given the broad range of types of publications in which scholars publish quality research, including book chapters, websites, and datasets. History has no established rank order of periodicals and impact factors—as in Humanities more generally—mean very little e.g. <http://guides.temple.edu/c.php?g=78121&p=509794>

There are two additional difficulties. The first is that, for historians, books are of primary importance in disseminating research. This was demonstrated in REF2014 where ‘books and parts of books’ were most likely to receive scores of 4*. There is no way of evaluating this type of output other than through peer review. In a discipline where so many outputs are submitted in book form, either as monographs or as chapters in edited volumes, metrics thus pose a particular problem. Second, the download half-life of journal articles in History—and Humanities articles more generally—is very much longer than it is for the Sciences. This is insufficiently recognized. The point is made in the British Academy report on Open Access <http://www.britac.ac.uk/policy/Openaccess.cfm> which nevertheless severely underestimated this half-life as the report did not include downloads from heavily used archive sites such as JSTOR. The RHS estimates that the true download half-life of a History article is at least 12 years.

The RHS would therefore argue strongly that the quality of scholarship in History, as in Humanities more generally, is not quantifiable by metrics and its full value and impact become apparent over a significantly longer term than a REF cycle. Greater use of metrics in place of peer review would not only fail to capture the nature and quality of world-leading scholarship but is also likely to have a distorting effect on the methods by which historical scholarship is disseminated. As peer review offers the flexibility to assess research in new, minority or unfashionable fields, any downgrading, or substitution by metrics, is also likely to distort subject matter by encouraging publications on well-worn or vogueish topics.

The RHS thus remains strongly committed to peer review, which is widely and routinely used to assess research quality in, for example, employment, promotion and publication decisions. It is, in fact, the *only* expert device we have to assess quality. After consulting REF2014 panel members, the RHS is confident that the workload in terms of the peer review of outputs was manageable, and the process very conscientiously carried out. Panel membership attracts outstanding academics who benefit from the opportunity to survey the field and in

whom their peers have confidence. Maintaining this calibre, and this level of participation, is essential to the REF process.

The assessment of impact was new to REF2014 and here some doubt has been expressed, both over the volume of material to review and the fact that inevitably peer reviewers had less experience of evaluating impact. Academic assessors are trained to assess intellectual quality rather than impact. There was also less time and information available for review.

This issue is likely to diminish as the accumulated experience of impact grows within the research community. The RHS moreover believes that impact has benefitted the historical profession as it has underlined the deep public interest in History and the relevance of our research to various fields, including education, digitization, and policy. Although the definition of impact for REF2014 excluded the kind of broad expertise of a historical field that is evident in much public engagement work—and which should be reflected in how the underpinning research is defined and understood—it is valued by many.

However, research undertaken by the RHS demonstrates that, in terms of authorship, ICSs, are not representative of the wider research community. The ‘impact case study’ is an artificially constructed exercise, but the fact that 75% of identified PIs were men and just under 65% of PIs were professors is of real concern. One simple way of making ICSs more representative of the historical profession would be to make impact portable. There is an obvious logical inconsistency in having outputs transferable and impact not as both rest on underlying research usually undertaken over a number of years. The RHS is clear that it is highly discriminatory against ECRs not to allow them to transfer impact from one institution to another, or to include that based on unpublished research in a PhD thesis. This makes it almost inevitable that institutions will rely on case studies contributed by people at mature stages in their careers.

A further consideration is how the requirement that departments submit one impact case study plus one other for up to 10 researchers has affected very small research clusters, for example, in universities where a department or school might only have 2 or 3 historians. The RHS is concerned that various REF measures put this kind of unit at risk (see the remarks on environment below).

2. If REF is mainly a tool to allocate QR at institutional level, what is the benefit of organising an exercise over as many Units of Assessment as in REF 2014, or in having returns linking outputs to particular investigators? Would there be advantages in reporting on some dimensions of the REF (e.g. impact and/or environment) at a more aggregate or institutional level?

The benefits of organizing returns by unit of assessment are most apparent in the evaluation of outputs, where subject-specific specialists are clearly best placed to conduct peer reviews. History is represented within the great majority of universities as well as in other cultural institutions. Large disciplines need their own UoA; the volume of outputs is substantial and the variety of expertise already contained within the discipline is broad.

In History, as more generally, REF owes its credibility as an assessment exercise to its expert review panels. It is clear from our consultations that the History panel worked very effectively, with a shared understanding of criteria and quality. In contrast, colleagues on

panels that covered a range of disciplines found the task of assessment could be more difficult and even conflictive. Departments within these broader panels, for example Languages, experienced more uncertainty preparing for REF. REF ‘scores’ based on amalgamated disciplines may also be misleading in terms of individual departments or schools and we would certainly resist History’s incorporation into a wider UoA. The international strength of historical research in the UK is reflected in the large number of high performing units, which has been confirmed in all previous REF exercises. We believe it is important to showcase this; any move to amalgamation would occlude the proportion of world-leading historical research for which British universities are responsible.

It is hard, if not impossible, to see how outputs could *not* be linked to individual researchers in History. This means that, while allocated scores—at least in terms of QR—go to institutions—and so, in a sense, it is the headline institutional score that matters—it is not clear how this could be obtained *without* expert peer review, which has to take place at the level of the individual outputs or ICS. There is also a further point, in that the granular detail of REF feeds into, for example, university guides, admissions league tables, and wider research rankings and here the disciplinary picture is crucial. This a particular concern for small, strong units within less-research intensive universities; these are not uncommon in History.

In broad terms, the RHS believes that the current arrangement for outputs (four, with differentially weighted monographs), impact and environment is manageable and effective with outputs as the main weighting. There is some feeling that environment should not weigh more heavily in the process. While, in a ‘bundling’ category such as environment, some form of metric evaluation is conceivable—research income and PGR numbers are two of the very few measures that can be aggregated across all subjects and both relate to environment—we see real difficulties with evaluating environment simply by metrics. REF is designed to recognize and support essential research activities, including a rich academic culture represented by seminars, workshops, conferences etc, participation and leadership in learned societies, editorial work, peer review and collaboration across institutions. Not to assess these vital academic functions would be to undermine them.

Wider use of metrics would also raise real issues of equity even at UoA level. It is clear from the RHS’s analysis of REF2014 that research income and PGR numbers were crucial to success in terms of research environment. Every university in the top 22, bar one, graduated at least 1 PhD per FTE over the REF cycle and 10 more than 1.5; the best predictor of rank in research environment was the number of PhDs per staff FTE between 2008 and 2013. Given the concentration of AHRC funding for doctoral study in a small number of consortia, in which Russell Group institutions predominate, this makes it almost impossible for small units in less-research intensive universities to do well in terms of environment no matter how strong their collective research endeavours. The RHS views this with real concern.

3. What use is made of the information gathered through REF in decision making and strategic planning in your organisation? What information could be more useful? Does REF information duplicate or take priority over other management information?

This is not primarily a question for representative bodies such as learned societies. Indeed, as REF information is provided at aggregate levels, and only every seven years or so, it is hard to see it as a significant source of management information.

The RHS believes that there is a useful purpose to having research, particularly research outputs, evaluated by independent external assessors. However, if, as we believe and as is set out in question 2, REF is a tool to allocate QR then it should be used for this purpose. It should not be for government to suggest or draw on other uses of REF by individual universities and the RHS would resist any move to embed REF as a performance management tool.

4. What data should REF collect to be of greater support to Government and research funders in driving research excellence and productivity?

The RHS is sceptical as to whether data in and of itself can be used to drive research excellence. We are strongly committed to research excellence, and to research publication in all its various guises, but see academic freedom, research time, the availability of funding and an atmosphere of creativity and communication as being far more pertinent to those working in Humanities as guarantees of research excellence and productivity. However, there are areas where data gathering as part of REF would usefully highlight areas of concern, for example historical research conducted in other languages and linguistic skills that need developing and/or support.

It is also the case that the suggestion in the recent Green Paper <https://www.gov.uk/government/consultations/higher-education-teaching-excellence-social-mobility-and-student-choice> that measures of casualisation might serve as a measure of teaching quality could equally be applied to REF. The close link between teaching and research is one of the great strengths of the British university system. This is reflected in the high proportion of staff on full academic contracts, which should be protected. Collecting such data would also provide a clear picture of the employment position and foreground an issue that is of particular importance to ECRs and the future shape of the profession.

5. How might the REF be further refined or used by Government to incentivise constructive and creative behaviours such as promoting interdisciplinary research, collaboration between universities, and/or collaboration between universities and other public or private sector bodies?

This is in some ways a curious question: how can REF be used in support collaboration between universities when by its very nature it makes them compete? However, the RHS is confident that collaborative mechanisms—which are often informal—are strong and that historians collaborate with as much or more vigour as ever. There are also examples of strong regional partnerships that encourage and fund research and doctoral studentship collaborations, for example the White Rose. These are, however, not directly related to REF.

Interdisciplinary research is stimulating and valuable but it is also important to defend multi-disciplinary collaborations and the single-discipline scholarship on which these rest. Promoting interdisciplinary research is not in itself a guarantee of improvements in research quality and the RHS would encourage the review to address this directly.

In terms of REF, impact has clearly encouraged collaborations with other private and public sector bodies and there may be scope to adjust output criteria to acknowledge this. The arrangements for cross-referencing interdisciplinary work to other panels and interdisciplinary specialists appear to have worked effectively, another strength of the

flexibility provided by peer review.

6. In your view how does the REF process influence, positively or negatively, the choices of individual researchers and / or higher education institutions? What are the reasons for this and what are the effects? How do such effects of the REF compare with effects of other drivers in the system (e.g. success for individuals in international career markets, or for universities in global rankings)? What suggestions would you have to restrict gaming the system?

There is widespread concern over the potential for departments to score highly in REF by only submitting a small proportion of their staff. There is a division of opinion among historians as to whether this should be addressed through a 100% return of eligible staff, given that a possible response would be to increase the use of teaching contracts and so alter the proportion of staff employed on full academic contracts (see the response to Qu. 4 above). However, there is a clear desire to prevent—or discourage—institutions from gaming the system. Not only can this have detrimental consequences for those omitted for strategic reasons but it can also lead to bad management decisions, as the 2/3 borderline is by far the most difficult to predict in internal assessment exercises. The RHS would therefore strongly favour restoring the previous system of requiring UoAs to identify the proportion of staff submitted.

It is hard to see how the other identified drivers affect historians or, indeed, most individual academics. Only a handful of scholars operate strategically in ‘international career markets’ while ‘global rankings’ are of little relevance below institutional level and then only to a select group of universities.

7. In your view how does the REF process influence the development of academic disciplines or impact upon other areas of scholarly activity relative to other factors? What changes would create or sustain positive influences in the future?

There is no doubt that REF has affected scholarly behaviour in History and that it will continue to do so. The introduction of impact, discussed above, has encouraged the academic community to be more outward facing and has also diversified the understanding of academic merit. Promotion criteria, for example, now commonly reflect the importance of external engagement. Less happily, the pressure for outputs has downgraded the status of the book in several disciplines—especially, but not exclusively, in the social sciences—where such publications used to be common. The differential weighting of monographs has prevented this in History and other Humanities subjects, and the RHS sees this as essential both to prevent distorting the research process and to reflect the research and scholarship that goes into producing such a substantial piece of work.

There is some suggestion that triple or quadruple weighting of monographs, and other very substantial outputs, would be beneficial, removing what may be seen as a perverse incentive to produce lower-quality article-length outputs to meet the output requirement. A single-authored 80-100,000-word monograph—the norm in our discipline—represents greater productivity than that required in other fields where team based research is the standard mode. Against this is the point that book authors seldom write nothing else and the greater complication it would give to panel deliberations. The RHS notes that the panel accepted 97% of double-weighting requests for REF2014, but this might change if the weighting rises.

One final concern is that REF pressure has downgraded the value of producing synthetic books and articles, which have an interpretative function and are primarily used for teaching purposes, including postgraduate teaching and training. Though pedagogically important, these are unlikely to be included in TEF and so run a real risk of falling between two stools. It would be helpful to have the value of these outputs recognised.

8. How can the REF better address the future plans of institutions and how they will utilise QR funding obtained through the exercise?

This is not a question that a learned society is well suited to answer but we are quite sure that our Fellows would resist adding bureaucratic requirements rather than taking them away. There is also a concern that the ‘strategy’ sections of the environment and impact statements were the least valuable parts of all the submissions: they are brief, rhetorical, and impossible for the panel to check.

9. Are there additional issues you would like to bring to the attention of the Review?

The changes to the last REF were very profound and the RHS believes that there is a strong case for little or even minimal change to allow the present system to bed down, particularly in terms of the new emphasis on impact. Institutions are already undertaking extensive planning for a future REF and there is no doubt that radical adjustment would be disruptive and costly. As stated above, REF affects both individual and institutional choices; as research planning should look to the medium and long term, this is all the more reason not to keep moving the goalposts.