

RSS response to Lord Stern's review of the Research Excellence Framework

The Royal Statistical Society (RSS) is a learned society and professional body for statisticians and data analysts, with almost 8000 members worldwide. As a charity, we advocate the key role of statistics and data in society, and we work to ensure that policy formulation and decision making are informed by evidence for the public good. We are therefore glad to share the Society's recent thinking on the Research Excellence Framework (REF), in light of the Stern Review's stated goals to allocate future university research funding more efficiently, offer greater rewards for excellent research, and reduce the administrative burden on institutions. This response has been informed by our Academic Affairs Advisory Group, and the RSS Working Group on REF League Tables that reported last year.

The RSS also endorses, and has contributed to, a response submitted by the Council for the Mathematical Sciences.

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?

Our first key point is that the evidence necessary to evaluate research and to inform career decisions is of both a qualitative and quantitative nature, and it is necessary for the REF to be appropriately balanced in taking account of both types of evidence. Quantitative (metric based) evaluations of research, research outputs and career decisions synthesise numbers of papers, grants, citations etc. into some measure of quality. Qualitative evaluations involve assessment by peers or discipline 'experts'. The weighting given to quantitative / numerical metrics will differ across disciplines and across situations. In general, we strongly advise that metrics are useful as additional information to assess performance, but should not form the primary basis for assessment. Metrics should complement, but not replace, expert opinion.

Our second key point is that obstacles preventing due credit for interdisciplinary or cross-disciplinary work need to be addressed. The difficulty of fairly assessing research that builds to cross-disciplinary and interdisciplinary applications is a serious problem. Any failure to address it risks driving researchers away from engaging across subject boundaries. We can see this with regard to statistics and its applications. In our response to the Nurse Review of Research Councils, we called for a strong umbrella strategy across research councils to advance the statistical theory and methodology that underpins scientific findings.¹ Similar to known research council silos, the REF has not been well structured to take account of cross-cutting research and its impacts. Methodological advances very often begin to develop before a clear link to a specific area of application has been established. Statistical methods are the building blocks of scientific discovery, and quantitative research methods that are incubated in the mathematical sciences have applications across science and in other academic fields. The

¹ *Royal Statistical Society response to the call for evidence on the Nurse Review of Research Councils* (PDF), April 2015. Available from: <http://www.rss.org.uk/Images/PDF/influencing-change/2015/rss-response-nurse-review-of-research-councils-17-4-15.pdf>

REF needs to take account of cross-disciplinary or interdisciplinary settings in its assessment of the research environment. The REF peer review structure or process also needs to be adjusted to better account for the impact of cross-cutting disciplines.

The Nurse Review recommended that “more consideration needs to be given to highly significant scholarly impact, that is, work which has a major influence on a field, resulting in a ‘paradigm shift’”. We agree strongly with this statement, and conclude that scholarly impact, both within and between disciplines should, in general, be weighted more strongly to ensure the long-term strength of our research base. If the REF does not recognise the importance of fundamental research, institutes will likely prioritise appointments with high prospects of short-term impact. These funding pressures would be to the detriment of the UK’s long-run strength in fundamental research.

Our third key point regards the quality of the quantitative measures or metrics produced by the REF, which are affected by strategic decisions or ‘gaming’. It is generally recognised that any high stakes system that encourages players to optimise their performance will change the nature of the system. The REF is one such system. It is also a major source of public information about universities’ performance, for which the development of more reliable quantitative metrics is needed as a priority. Future REFs should eliminate, as far as possible, the effects of tactical decisions made by universities about their REF submissions.

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?

Provided that there is no systematic bias, aggregate metrics should broadly be favoured as they are generally more robust. At present however, certain aggregate REF metrics face legitimate concern that they are ‘gamed’. Journal impact factors, university rankings and citation counts are three prominent examples that are subject to such concerns, as was recognised in HEFCE’s recent review of metrics.² With regard to university rankings, we have called for the investigation of a ‘quality per researcher’ measure.³

Overall, we unequivocally support UoA-level assessment at the present time.

Section 2

While the primary purpose of REF is QR resource allocation, data collected through the REF and results of REF assessments can also inform disciplinary, institutional and UK-wide decision making.

² HEFCE (2015) ‘The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management’, Available at: <http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/>

³ The Royal Statistical Society (2015) *Report from the working group on Research Excellence Framework (REF) League Tables* (PDF) Available from: www.rss.org.uk/ref

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?

We have no response here as this question is for submitting organisations.

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

We are keen that the administrative burden of future REFs be reduced. This means very careful consideration indeed should be given to any actions that increase the data collected for REF.

Section 3

The incentive effects of the REF shape academic behaviour, such as through the introduction of the impact criteria.

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?

We have concerns about this Review question as it suggests that the government may intend to measure and incentivise behaviours that aren't always or necessarily related to the quality of research. In our view the key purpose of the REF should be to measure the quality of research across the system, to see how the UK is performing in research, and to provide fair assessment of quality no matter what type of working arrangement produced the research. It is likely that incorporating more indirect or behavioural incentives would weaken the REF's focus on the strength of key research outputs from the UK research base.

Promoting interdisciplinary research is specifically mentioned in this question. There is a difference between removing obstacles to interdisciplinary work getting due credit and actually incentivising it. We do not think that interdisciplinary working should be the focus of the REF, however the product of it should be as readily submitted as single-discipline research. Researchers might be driven away from interdisciplinary work if it appears to be more administratively complex or have a lower chance of recognition. We therefore think that the REF needs to improve its procedures for evaluating cross-disciplinary or interdisciplinary research.

Section 4

Previous studies have focused on the costs of REF with respect to the time and resources needed for the submission and assessment processes. The Review is also interested in views and any associated evidence that the REF influences, positively or negatively, the research and career choices of individuals, or the development of academic disciplines. It is also interested in views on how it might encourage institutions to 'game-play' and thereby limit the aggregate value of the exercise.

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?

A major limitation of university rankings produced from the REF has been that institutions were permitted to select which researchers to include. Institutions differ a great deal in their selectivity. To address this we have called for the development of a more accurate 'quality per researcher' metric, which in the 2014 exercise did not exist. Under this metric, every REF-eligible researcher would have to be included, resulting in a "quality" score like the Grade Point Average (GPA), but the GPA weights of 4:3:2:1:0 for REF quality levels would be replaced by the weights that determine QR funding.⁴

The submission of all research staff in such a way to the REF would remove a strategic dilemma otherwise faced by universities (in deciding which staff and how many staff to select), as well as providing a more direct interpretation of the universities' rankings.

Citation metrics are also widely used in the REF, and on this we have advised that adaptations are needed to 'scale' citation metrics according to what the norms are in a given discipline or sub-discipline. In mathematics and statistics, there is a strong feeling that metrics are perhaps less useful than expert opinion. This is for reasons including that certain sub-disciplines of mathematics have lower citation rates, and bibliometric sources do not accurately reflect the citations in these sub-disciplines. Adler et al. (2009) and subsequent discussants have criticised the use of journal impact factors and h-indices, and in particular the use of comparisons across disciplines and sub-disciplines.⁵ There is no reason that publication in a high-impact journal necessarily means the article is of higher research quality. The use of the mean as a summary measure is also problematic, as journal citations have a long 'tail' and are not well summarised by the mean.

The individual article citation rate has the advantage of at least being directly related to the output concerned. There are however differing rates of citation in different disciplines and in different types of publications (review articles are cited more frequently, for example). More recent

⁴ Letter to *THE* from Prof Peter J Diggle, President of the Royal Statistical Society (PDF), 11 May 2015. Available from: <http://www.rss.org.uk/Images/PDF/about/press-releases/PJD-letter-to-THE-11-05-2015.pdf>

⁵ Adler, R. Ewing, J. & Taylor, P. (2009) *Citation Statistics A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)* (PDF) Available from: http://arxiv.org/pdf/0910.3529.pdf%3Forigin%3Dpublication_detail

publications have had less time to be cited compared to older publications. An often raised issue is also that of 'poor quality' papers being extensively cited. This is sometimes dismissed on the grounds that 'on average' it has little effect. However it is important to address even a very small number of such extreme misgradings, as they affect the integrity of research assessments. It is also important to be aware of the flaws with current providers of bibliometric data. Several leading publishers lead the field in terms of visibility and adoption of their data. However these sources each have a specific set of journals in their core set that do not cover all of disciplines such as mathematics, and thus do not always accurately reflect the citation rates of papers.

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?

We would like to see the methodology of the REF scrutinised and improved on a pragmatic basis and think that this would have a positive influence for the future. Any greater reliance on REF metrics that are not methodologically sound would threaten the credibility of the system. We therefore welcomed *The Metric Tide* report's recommendations and in particular the Review's caution that "no metric can currently provide a like-for-like replacement for REF peer review".⁶ We agree that quantitative metrics should be used selectively and improved as a complement to peer review, forming a more reliable basis for public information e.g. for the comparison of institutions, and for allocations of funding.

Section 5

Much of REF focuses on the retrospective analysis of success achieved by institutions either through output or impact. Yet the resources provided anticipate continued success based on that track record. Are there means of better addressing forward-looking institutional plans and priorities, and how these might feed in to national policy?

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

No response to this question.

Final thoughts

The Review is keen to hear of creative ideas and insights and to be open in its approach.

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

No response to this question.

Submitted by RSS' Press & Policy Officer, 22 March 2016

⁶ HEFCE (2015) 'The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management', Available at: <http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/>