



# Assessing impact submissions for REF 2014: An evaluation

Executive summary

Catriona Manville, Susan Guthrie, Marie-Louise Henham, Bryn Garrod,  
Sonia Sousa, Anne Kirtley, Sophie Castle-Clarke and Tom Ling



EUROPE

This executive summary is an extract from *Assessing impact submissions for REF 2014: An evaluation*, by Catriona Manville, Susan Guthrie, Marie-Louise Henham, Bryn Garrod, Sonia Sousa, Anne Kirtley, Sophie Castle-Clarke and Tom Ling, RR-1032-HEFCE, 2015 (available at [www.rand.org/t/RR1032](http://www.rand.org/t/RR1032)).

The research was conducted by RAND Europe and funded by the Higher Education Funding Council for England, the Higher Education Funding Council for Wales, the Scottish Funding Council and the Department for Employment and Learning, Northern Ireland.

Published by the RAND Corporation, Santa Monica, Calif., and Cambridge, UK

**RAND** is a registered trademark.

© Copyright HEFCE 2015

RAND Europe is a not-for-profit research institute whose mission is to help improve policy and decisionmaking through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

#### Support RAND

Make a tax-deductible charitable contribution at  
[www.rand.org/giving/contribute](http://www.rand.org/giving/contribute)

[www.rand.org](http://www.rand.org)  
[www.randeurope.org](http://www.randeurope.org)

## Overarching summary

In 2014 the impact of academic research outside of academia was for the first time assessed as part of the Research Excellence Framework (REF) in the UK, determining future funding for institutional research. Impact was assessed and graded by academics and research users. Guidance and support was provided by the higher education funding bodies. In total 6,975 impact case studies and 1,911 impact templates were assessed and graded. This report explores the strengths and weaknesses of the assessment process and identifies unforeseen issues and their management. It is based on focus groups, interviews, survey and documentary analysis. It details the process that took place, examines the perceptions of academics and research users who were involved in the process, arrives at overall judgements about the conduct of the impact element of REF 2014, and discusses the implications for future assessments.

### S.1. Overview of the assessment process

The impact assessment process undertaken in REF 2014 comprised six main stages, as shown below in Figure S.1.

The assessment of impact was made on the basis of two different types of document submitted by higher education institutions (HEIs): impact case studies (REF3b) and impact templates (REF3a). Case studies provide

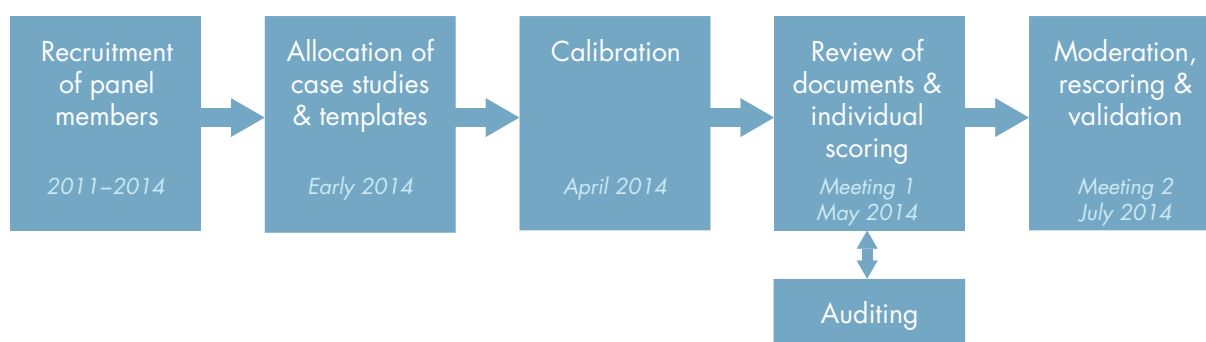
examples where research has led to impact, and templates set out the wider strategy around facilitating the translation of research into impact.

Submissions are made at the level of the sub-panel (i.e. each HEI will make multiple submissions, one per unit of assessment (UOA) that is relevant to their research). Each submission consists of one impact template, plus a number of case studies (where the number depends on how many staff are submitted by the institution in that UOA, approximately one case study per 10 staff).

In REF 2014, the case studies and templates were allocated for assessment to sub-panel members and impact assessors by the sub-panels, who were asked to review and score them each individually. They were provided with some guidance to do this, and there was a calibration session early in the process to allow them to calibrate their scoring to the rest of the sub-panel. Each case study and template was reviewed by more than one individual and exact practice on the allocation and scoring processes differed between main panels and sub-panels.

Once the case studies had been scored individually, there was a process of moderation of those scores. Initially there was some moderation at the sub-panel level (though exact processes for this varied), and then at the main panel level. There was also some limited validation between main panels.

Figure S.1: Overview of the impact assessment process



An additional step alongside moderation was audit, a process by which panel members could request checks on the case studies and templates to check their eligibility against various criteria, or to access the sources of evidence referred to in the case studies. Audit was not conducted systematically across all case studies, but rather on the basis of sub-panel requests. Some 74 per cent of institutions had at least 5 per cent of their case studies audited.

Finally, scores for case studies and templates in each submission were combined to produce an overall impact ‘profile’. The profile indicates the proportion of the submission that is at each level from unclassified, to 1 star, 2 star, 3 star or 4 star (the highest rating). Ultimately these profiles are combined with the output and environment profiles to produce an overall profile for each HEI submitting to each UOA. Funding will ultimately be allocated by the funding bodies on the basis of the REF outcomes, with 3 star or above being the requirement for funding allocation.<sup>2</sup>

## S.2. Key messages

### S.2.1. By a large majority, panellists felt the process enabled them to assess impact in a fair, reliable and robust way

*I've been struck all the way through by the emphasis on fairness in the efforts being made on the structure of the exercise to ensure that there was really a fair and proper assessment.<sup>3</sup>*

*Above all, there was extremely intensive and carefully considered discussion of the application of the criteria as a whole in practice, reviewing and revisiting until we were clear collectively that the process was absolutely fair and reliable.*

Panellists felt there was reasonable consistency in the scores given by individuals assessing case studies. This agreement was within main panels and between the different panellist roles. It was recognised that this element of the assessment was new, and although there were areas for improvement for future exercises there was much that had gone well. As the first attempt at assessing impact on a large scale it was seen to be successful, though panellists recognised there were areas

for improvement. One panellist stated: ‘There is much to commend [it] although there are improvements to be made and much to be learnt’. Consistency was supported by various steps in the process described in the following paragraphs.

Many panellists commented positively on **calibration** as ‘creating a shared understanding of impact’, and allowing the case studies to be assessed fairly and reliably. Where a minority of impact assessors were not involved in calibration, there was concern that their understanding was not aligned with those who had been through the process. Following calibration, there was **continued discussion** that was viewed by participants to be inclusive and fair. If a consensus was not reached it was escalated to the whole sub-panel, or from the sub to the main panel for resolution. In sub-panels with fewer submissions (such as Agriculture, veterinary and food science (UOA6)) all case studies were discussed by the full sub-panel and a score agreed. Larger sub-panels only discussed cases where a consensus could not be reached among the initial reviewers. The scores were then **moderated** within both the sub-panel and the main panel to ensure case studies were graded equivalently. On occasions where scoring was thought to be out of line, sub-panels were asked to review their decisions by the sub-panel chair or a member of the relevant main panel. The sub-panel chairs sat on the main panel, further facilitating communication and reinforcing comparability across sub-panels. In addition, the main panels had a number of research user members, as well as international advisors who floated between sub-panels to support sub-panel members and impact assessors in their deliberations. In particular panellists from areas of Main Panel A noted how useful their international advisors were in supporting them to be fair to the impacts they were reviewing.

There was less confidence from panellists about the consistency between main panels. This was most often seen by sub-panel members who were not involved in the process across the main panels; those engaged at the main panel level were considerably more satisfied with the process. This could imply that the issue is one of transparency and the information the sub-panel level received. On the other hand, confidence in the process was enhanced by how case study **allocation** was managed. In the majority of sub-panels each panellist reviewed case studies with a variety of other panellists, allowing learning to be shared. Overall, panellists felt

<sup>2</sup> For further information on the process, see the following chapters, and overview reports published by each main panel (REF 2015).

<sup>3</sup> Quotations used throughout this report are taken from the focus groups and survey responses.

that the **rules and guidance** helped them to fairly and reliably assess case studies.

### **S.2.2. There was variation in the way the process was conducted**

*Plenary discussions of scoring disparities between individuals were possible for small sub-panels but impossible for the largest ones.*

Due to the devolved governance of the process and the understanding that one model for assessment would not suit all disciplines there was variation in the process across sub-panels. This occurred at multiple points in the process – from number and combination of reviewers, to the application of eligibility criteria and the use of corroborating evidence. The flexibility this provided reflected disciplinary differences but may have led to unwarranted variations.

For example, individuals in our focus groups reported differing interpretation of guidelines when identifying the thresholds for the eligibility of a case study, such as the 2-star quality of underpinning research, and when assessing the link between research and impact. We note that this problem of identifying thresholds is not unique to assessing research impact.

Variation in interpretation was also seen with the use of corroborating evidence, and confirming the statements made in impact case studies. It was unclear to panellists the extent to which they should use prior knowledge and the point at which an audit query needed to be raised to check details. This led to differing practices between panellists.

Panellists used their judgement to select the appropriate level of validation for a case study depending on content, and carried out differing levels of checking. This is understandable given the different levels of complexity and different kinds of claims being made. However, participants also reported differing interpretations of how checking of evidence should be conducted, and what was allowed. Consequently, different practices occurred, within and across sub-panels. In general there was a desire from many panellists to have unmediated access to evidence, although the additional burden this would entail was acknowledged. Overall, there was a view from panellists that the process of corroboration was fair and added rigour but at the same time there was an anxiety that there may have been an unknown number of case studies where submissions were over-claiming which may have ‘slipped through the net’.

Panellists in Main Panel D stressed the need for a level of expertise in order to fully assess impact and evidence claims. However it was also recognised that this might introduce unfairness if individuals on the panel had varying levels of knowledge. Examples were given in areas of Main Panels A and C where those who knew about specific case studies ‘were told [by sub-panel chairs or other panellists] they weren’t allowed to use that information’. This was frustrating, as panellists believed it was this very knowledge that partly justified why they had been selected for the panel.

As described in the following section, all main panels used ½ stars to create more granularity. In addition, areas of Main Panel A developed a framework of up to 8 stars when structuring their discussions about scores. The rationale behind this was that there were different levels within each star, and it ensured that 4-star impacts were not downgraded due to comparisons with exceptional examples (a problem discussed below in Section 1.2.3). In Main Panel A, when awarding the scores, case studies scoring 4-8 were all graded as 4-star.

Our study did not review the case studies to examine the scores awarded, so we cannot assess whether variation in interpretation produced variation in scores. However, we can report that the process aimed to limit the impact of variation through calibration at all levels; through creating a shared understanding at a sub-panel and main panel level; and moderating scores at a main panel level and between them. For example, it is important to note that when a sample of Main Panel A impacts were assessed by the other three main panels they also scored highly.

### **S.2.3. Panellists felt they were able to differentiate between submissions in more detail than the scoring process allowed them to express**

*One challenge we had, which makes us appear generous but we weren’t, was that sometimes we found an exceptional 4 that was really off the scale, which made it hard to see that others could still be a 4.*

Several panellists said that there were two different parts to calibration: consistency and level. It was felt that the former was easier than the latter. The difficulty in awarding levels, based on a four-point scale, is reinforced by the fact that all panels used a more granular scale for awarding levels. Sub-panels across all four main panels used ½ stars in their scoring. The difference described however, is not reflected in the scores.

The scores for the impact element are very high, with nearly 40 per cent of all impact sub-profiles rated as 4 star and over 80 per cent considered 3 or 4 star.<sup>4</sup> This apparent ‘bunching’ of scores suggests consistently high-quality case studies and impact templates when graded against the level criteria.

For the highest score there is a quality range which goes from just achieving the threshold up to a theoretically infinitely high quality. To address this, some areas of Main Panel A, as noted above, framed their discussions using a framework of up to ‘8-stars’, awarding case studies thought to be 4-star or above as 4 star, i.e. so that the very best should not determine the threshold between a 3 and 4-star score. The use of this conceptual framework was described by one panellist as helping them ‘to get the calibration right, rather than setting the very best ones as a 4, which would have made 3 a severe judgement. This encouraged us to think that there were lots of 4s’.

Panellists from Main Panels B and D stressed an awareness of the funding implications of the boundary between 2 and 3-star.

Panellists stated that they found it difficult to verify the link between the underpinning research and the impact and sometimes the links were more tenuous than they would have liked. One suggestion from Main Panel A was that attribution could be used as an additional criterion to differentiate where there were multiple submissions claiming the same impact.

#### **S.2.4. There were particular challenges in assessing the impact templates**

*It felt more aspirational than anything based on reality.*

Panellists felt that the impact templates were more difficult to assess than the case studies. In particular, there was a concern that the impact template was not an effective way to assess how institutions support the translation of research to impact, and many questioned its value in the submission.

The lack of requirement to evidence the claims made meant that the quality of writing had a large effect, and this made it difficult to discriminate between templates. In addition, it was felt that it was too easy to retrofit a strategy that may not have been in place throughout the period of assessment. On occasion panellists were

uncomfortable with their judgement as they felt that the template encouraged plans for the future that could be ‘a figment of the [template author’s] imagination’. However, there were some who felt the impact template was a ‘vital document’ to HEIs and to assessment, as it helped panellists to better understand what institutions were doing to support their staff in encouraging and facilitating impact.

There was a range of suggestions for how to improve this element in future assessments. Some suggested it should be completely removed as the environment could be evidenced by the case study examples whether or not the HEI had a strategy. Others suggested combining the impact and the environment template, as the impact environment is a subset of the overall research environment. However, there were those who felt the environment within which impact is fostered should continue to have a separate submission. In particular there were examples of where impact assessors found it helpful to review the impact template to understand the context within which impact had occurred and this might be more difficult if it was subsumed within the environment template (REF5). Alternatively, to reduce the influence of the quality of writing on the submission, some suggested greater use of narrowly factual information. Whatever the future for the impact template, there is a need for guidance to HEIs to ensure greater clarity.

#### **S.2.5. Through involvement, research users built useful networks, but the burden of involvement was a significant challenge**

*[One of the benefits was the] opportunity to make new contacts and strengthen [my] existing network in academia and industry.*

The most cited benefit among impact assessors (mentioned by 44 out of 74 respondents) was networking and collaborative working. This included the enjoyment of engaging with academics (and the academic sector more broadly) as well as networking opportunities with the academic sector for further professional activities. This is closely linked to the second most-cited benefit (listed by 37 out of 74 respondents), which was learning about and gaining an understanding of academic research taking place across the UK. Respondents noted they had gained an awareness of different types of research, had been exposed to new research ideas and

<sup>4</sup> A sub-profile was formed of the impact element for each submission, taking into account the levels awarded to each impact document.

had a better overview of academic disciplines of relevance to their work, among other things.

These benefits, however, came at a cost for those involved. Burden was by far the most frequently mentioned challenge of involvement in the assessment, being mentioned by 45 out of 74 respondents. Based on the quantitative data collected through the survey, the median amount of time spent on the process by impact assessors was 11 days. There was, moreover, variation in the commitment made by individual impact assessors – the interquartile range of the estimates of time spent by impact assessors from the survey spanned from 7 to 15 days.<sup>5</sup> Despite seeing influencing and networking as benefits, the burden was particularly significant for those not employed by an HEI, since time away from work may have been given to cover the time attending meetings but it is likely that the assessment of impact documents was completed outside their normal working hours.

### **S.2.6. Engagement by academics in the process offered benefits for academic careers and institutions**

*It will be useful for me in preparing future impact case studies for the next REF.*

Academic sub-panel members reported direct benefits for careers and institutions resulting from participation. By far the most frequently cited benefit (mentioned by 174 out of 396 academic sub-panel respondents) related to their improved understanding of the assessment process. Specific examples included: learning how the process of assessing impact worked in practice; understanding how other people judge impact; developing a deeper understanding of the rules and how they should be applied; gaining reassurance of the robustness of the process; and learning how to approach the exercise for future assessments – including how to present good case studies and templates. This reflects the increasing importance of impact assessment in academia more widely.

Other frequently mentioned benefits included learning about the range of academic research taking place across HEIs (mentioned by 99 of 396 respondents) and learning about the diverse range of impacts emanating from research (mentioned by 77 of 396 respondents). The benefits are both for academics' own careers and for their institutions more widely, in terms of building

an understanding of impact and its assessment, and an inside perspective on what other academics and institutions are doing in this area – effectively giving participants a 'competitive advantage'.

In terms of challenges, the burden of engagement is the most frequently listed (mentioned by 86 of 396 respondents), but several other response categories are mentioned almost as frequently, notably the challenges of using evidence (mentioned by 79 of 396 respondents) and of comparing and fairly assessing different types of impact (76 of 196 respondents). Unlike impact assessors, the issue of burden, though still important, is not so dominant for this group, which also reflects on the challenges of the process itself rather than challenges for themselves in engaging with the process. This re-emphasises the more direct engagement that this group necessarily has with the process and its outcomes, which directly affect their sector. While understanding the process and gaining an 'inside' perspective is beneficial for this group, the challenges and limitations of the process have direct impacts for this group and as such are of significant personal concern to them.

### **S.2.7. Bringing together different perspectives of academics and research users was seen to be successful and valuable**

*It was interesting to hear remarks from the panel about how the REF had tempered the behaviour of the academics; it also tempered our behaviour as well [...] and it was a stroke of genius to get people together to get that consensus generated.*

A significant feature of the impact element of REF 2014 was that it included and synthesised the views of both academics and research users in the assessment process to a greater extent than other elements of the assessment process. There was a possibility that since these two groups operate in related but different worlds that they might struggle to form a shared understanding and build sufficiently trusting relationships to work well together in the assessment process. We raised these issues with both academics and research users in separate focus groups and also conducted one-to-one interviews with representatives of each group. Our survey data also helped inform our understanding of how well this worked.

Both academics and research users in the focus groups and the interviewees highlighted the value of including

<sup>5</sup> The interquartile range is a measure of statistical dispersion, being equal to the difference between the upper and lower quartiles.

both perspectives.<sup>6,7</sup> It was widely agreed that the two perspectives moderated each other and added to panelists' confidence in the process. There was a widespread sense that the exercise was inclusive and that views were listened to equally. This is echoed in our survey which shows that there was a high percentage of satisfied and very satisfied respondents, with the highest levels of satisfaction regarding all members being listened to equally.

### **S.3. Conclusions**

Confidence in outcomes, and in the processes used in the research impact assessment, were relatively high following completion of REF 2014 among those most involved. Based on the evidence presented in this report, this confidence was due in large part to: the formal guidance and processes used; an academic-led and inclusive approach to leadership; and the high level of commitment shown by academics and research users. The introduction of an impact element in REF 2014 might have been expected to generate concerns because of the relative novelty of the approach and because of obvious difficulties in measurement, but in general it has succeeded.

There are a number of incremental improvements which could be made to the process. These include: access for panellists to underpinning research and corroborating evidence; potential use of closed questions detailing eligibility; the stringency with which to apply the rules around the format; and clearer guidance to HEIs and panellists (on how to present eligibility information within the case studies, and the requirement, or not, for impacts to demonstrate both reach and significance).

In addition, we have flagged some areas for further consultation. These include: revisiting how to manage the risk of random variations having excessive effects in small units of assessment; how to avoid the risk of unsubstantiated and false claims being made; how to clarify the processes for assessing different kinds of impact; and how best to capture the information pertaining to the wider HEI environment for nurturing and developing impact.

---

<sup>6</sup> Although a handful of academics worried that users were not suitably equipped to assess impact.

<sup>7</sup> In addition, main and sub-panel members from areas of Main Panel A highlighted the value of their international advisers, and the perspective they brought to the process.