

28th July 2016

Independent review of the Research
Excellence Framework (REF):
Synthesis of responses submitted to
the REF Review Call for Evidence and
follow-up interviews

Independent review of the Research Excellence Framework (REF): Synthesis of responses submitted to the REF Review Call for Evidence and follow-up interviews

Synthesis of responses submitted to the REF Review Call for Evidence and follow-up
interviews

Main report on the consultation:

<https://www.gov.uk/government/publications/research-excellence-framework-review>

technopolis |group| July, 2016

Kalle Nielsen

Peter Kolarz

Kristine Farla

Victoria Blessing

Paul Simmonds

Contents

- 1 Introduction1**
 - 1.1 This paper..... 1
 - 1.2 Responses submitted to the call for evidence 1
 - 1.3 Programme of interviews..... 2
 - 1.4 Structure of the report3

- 2 Efficiency and accuracy of the assessment process..... 4**
 - 2.1 Suggested improvements to existing processes.....4
 - 2.1.1 Number of outputs.....4
 - 2.1.2 Staff selection6
 - 2.1.3 Frequency of REF assessments9
 - 2.1.4 Environment assessment9
 - 2.1.5 Other issues.....9
 - 2.2 The definition of impact in the REF10
 - 2.3 Peer review and metrics..... 14
 - 2.3.1 Attitude towards peer review 14
 - 2.3.2 Attitude towards the use of metrics 16
 - 2.3.3 Suggestions for more extensive use of metrics 17

- 3 Level of assessment19**
 - 3.1 Units of Assessment (UoAs) 19
 - 3.2 Linking outputs to individuals 20
 - 3.2.1 Arguments for de-coupling outputs and individuals 20
 - 3.2.2 Arguments for retaining the link between outputs and individuals..... 21
 - 3.2.3 The assessment of individual staff 21
 - 3.3 Aggregation22
 - 3.3.1 Assessment of research outputs at a more aggregate level22
 - 3.3.2 Assessment of research environments and impacts at a more aggregate level.....22

- 4 Information for HEIs 24**
 - 4.1 Current use of information gathered through the REF by organisations.....24
 - 4.2 Suggestions for information that could be more useful to organisations.....25
 - 4.3 Information from REF and other management information.....26

- 5 Information to be collected by government 28**
 - 5.1 Caution against adding additional data requirements 28

5.2	Better use of existing data	28
5.2.1	Better use of data collected through the REF	28
5.2.2	National coordination and use of other existing data	29
5.3	Suggestions for new data to be collected	29
6	Support for interdisciplinary research and collaboration.....	31
6.1	Promoting interdisciplinary research	31
6.2	Collaboration between universities and other public and private sector bodies	32
6.3	Open Access	34
7	Influence on choices and gaming	35
7.1	Influence on choices of individuals and institutions	35
7.2	Other drivers	36
7.3	‘Gaming’ and solutions to gaming.....	36
7.3.1	Types of gaming identified	36
7.3.2	Suggested measures to restrict gaming	37
8	Influence on disciplines and other areas of scholarly activity	39
9	Future plans.....	40
9.1	Assessment of future plans in the REF	40
9.2	The use of QR funding by institutions	40
10	Other comments.....	42
10.1	The purpose of the REF	42
10.2	Burden of the REF	42
10.3	Teaching	44
10.4	Equality and diversity	44
10.5	Continuity and transparency	44
11	Summary of key issues	45

1 Introduction

1.1 This paper

In December 2015, Universities and Science Minister Jo Johnson launched a review of the Research Excellence Framework (REF), chaired by the President of the British Academy, Lord Nicholas Stern. In support of the Review, Lord Stern published a Call for Evidence to explore a range of issues arising from discussions in the early stages of the review.

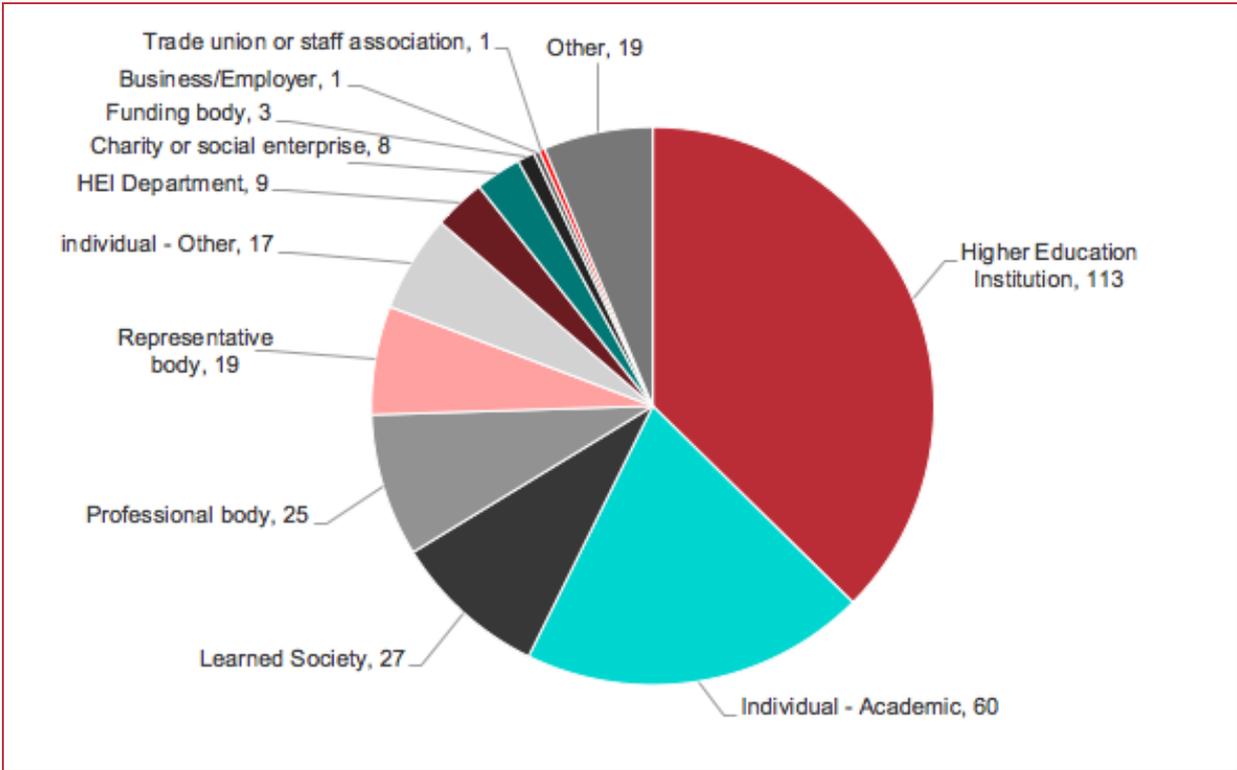
This paper presents a synthesis of the 301 responses provided by the UK Higher Education community to that ‘Call for Evidence,’ combined with the key messages arising from a supplementary programme of 40 qualitative interviews with universities, academics, research users and various intermediaries, from industry associations to learned societies.

What we heard, from both the Call for Evidence and the programme of interviews, has been shared with Lord Stern at various junctures, to support the deliberations of the Review Group as it moved through its own work. This paper brings together that rolling programme of briefings in a single document. The feedback from both aspects of the REF consultation process is presented here in an anonymised and abridged form, following the structure of the Call for Evidence.

1.2 Responses submitted to the call for evidence

The consultation ran from the 27 January to the 24 March 2016 and solicited 301 unique responses from broad range of key actors across the sector. These include individuals, groups of individuals, Professional and representative bodies, learned societies and from the majority of Higher Education Institutions in the UK. Figure 1 provides an overview of responses by category of respondents.

Figure 1 Respondents to the Call for Evidence, by category



Source: Technopolis

1.3 Programme of interviews

The study team carried out telephone and face-to-face interviews with more than 40 people and organisations, in order to get a better understanding of particular issues (e.g. REF and interdisciplinary research) or to plug gaps in the responses to the call for evidence (e.g. the business community). The semi-structured interviews focused on the future development of the REF, tackling many of the issues that had arisen within the Call for Evidence, but also exploring views on potential alternatives to REF, as a thought experiment. The interviews lasted between 30 minutes and 2 hours, and have been presented here in a synthesised and anonymous format. The names of the contributing organisations are listed in the appendices.

Table 1 Interviews by type of stakeholder group

Main stakeholder groups	No of interviews completed
Higher Educations Institutions (HEIs)	18
Businesses and industry bodies	7
Research charities and intermediaries	10
Government departments and executive agencies	4
Academics and other individual experts	3
Total	42

1.4 Structure of the report

The report is structured according to the questions posed in the Call for Evidence. Each section corresponds to one of the nine questions asked and starts with the wording from the call set out in a text box. The last section of the report provides a summary in the form of a table indicating the balance of opinion on some of the main issues discussed.

Within each section, the main issues raised by the respondents are discussed in turn; we focused on presenting the recurring issues raised by multiple respondents, however, we also include a small number of more singular perspectives where those were particularly clearly articulated. The thematic analysis is complemented, where appropriate, by indications of the weight of opinion. A distinction is made between HEIs, individuals (academic and ‘other’ individuals) and ‘other organisations’ where the latter refers to all respondents not included in the first two groups.

We have been cautious in our presentation of the distribution of responses, and have avoided the suggestion that a majority view among respondents is the definitive balance of opinion for the whole community. The respondents are self-selected and cannot be assumed to constitute a representative sample of the sector as a whole. While a large majority of all UK universities and colleges submitted a reply, only a small fraction of the total population of individual academics responded. In addition, the consultation questions are open in nature and invite the respondents to comment on a wide variety of issues. As a consequence, only a subset of respondents will have commented on any one issue, and there might well be a bias in favour of requests for change over support for the status quo. Overall, a count of respondents’ expressed opinion can give an indication of the balance of opinion but should not be taken as a robust measure for the opinion of the sector.

2 Efficiency and accuracy of the assessment process

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?

2.1 Suggested improvements to existing processes

2.1.1 Number of outputs

The REF 2014 rule requiring HEIs to submit four research outputs for every person submitted was widely written about. Many HEIs and other organisations suggested decreasing the number from 4 to 3, or even 2, in order to reduce the burden on the assessment panels and ensure a more complete and probing assessment of all outputs. Some also argued that this would further emphasise the need to produce quality over quantity and provide more room to pursue longer-term research projects without the need to produce a large volume of research in the short term.

Several organisations expressed support for retaining 4 research outputs per person, and argued against re-basing the figure and changing what has emerged as a community-wide norm. This was seen to strike the right balance between comprehensiveness and feasibility. Reference was also made to earlier consultations conducted by HEFCE in the preparation of REF 2014 in which a proposal to reduce of the number of outputs was rejected by the sector.

Several HEIs also suggested a more flexible system that would allow a different number of research outputs to be submitted from different individuals - between 1 and 4 or 1 and 6, for example – within the frame of an overall target for the unit. This would allow a wider group of researchers to contribute and also eliminate the need to justify special staff circumstances. In combination with a requirement to submit all staff, it could also help keep the number of outputs submitted at a manageable level. Some argued for taking the further step of decoupling the number of outputs from individuals completely, as discussed below under the related issues concerning staff selection (section o) and the link between researchers and outputs (section 3.2).

There were also suggestions from several HEIs and other organisations of a more radical change towards a system of ‘sampling’ based on the submission of the all published academic outputs. It was suggested that HEIs should be able to choose which outputs were to be reviewed in depth, or that it should be determined by statistical methods such as stratified sampling. The remaining outputs would be subject to a metrics-based description. It was also noted that the implementation of the Open Access requirements could facilitate the submission of all research outputs.

Outputs – key messages from stakeholder interviews

Our interviews produced a range of feedback on the question of research outputs, which echoed the written feedback, and included the following:

- Several contributors expressed the view that a future REF would ideally consider all research outputs produced by a given university – for each Unit of Assessment – across the period under review.
- The arguments for moving to such an arrangement were threefold. In the first instance, it would allow quality profiles to be determined based on the totality of a school or department's research outputs and that was thought to be a more robust basis for comparing quality and allocating funding. Secondly, it would remove the need for HEIs to expend quite so much time and effort in determining the set of papers and people judged most likely to produce the highest quality score. Thirdly, it would require the assessment process to embrace the diversity of research outputs and research-active staff.
- Those arguing for the inclusion of all outputs suggested REF would be able to cope with the dramatic expansion in the volume of material to be assessed, by sampling.
- Many contributors thought that a move to the submission of all outputs was impractical and unnecessary, and that it was entirely appropriate to allow universities to be selective.
- Others suggested the '4-outputs' was now a UK norm, and that it was a good balance between the possibly very much higher numbers of multi-authored papers typical of certain disciplines in the sciences and the smaller number of monographs common to other fields.
- There was a strong desire to see future REF's do even more to encourage the submission of the full spectrum of types of research outputs, from papers to performances.
- Most discussion partners took the view that submitting all staff was more important for the robustness of the assessment process than all outputs.
- In several cases, contributors made the case for the introduction of a more flexible system for outputs, with a cap and floor on the number of outputs from each individual to allow HEIs to capture their diversity while also avoiding overwhelming the peer review process.

2.1.2 Staff selection

A large number of the 300 respondents brought up the issue of staff selection and its effect on both the accuracy and efficiency of the assessment process and staff morale. There was no consensus but many responses discussed the idea in length and several other key elements of the REF – such as the burden of the assessment on HEIs and REF panels, the number of outputs to be submitted, the use of metrics, the link between individuals and outputs, the assessment of collaborative work and various types of gaming – were seen to revolve around this issue. Consequently, this was also discussed in relation to the level of assessment (see section 3) and ‘gaming’ (see section 7.3).

Many respondents, including HEIs, individuals and other organisations, argued in favour of removing staff selection. A requirement to submit all staff would produce a more accurate assessment of a unit’s research activities. They also felt it would eliminate some of the negative consequences of the process of staff selection, including the damaging effects on careers of being excluded and the potentially demoralising and divisive effect on departments. It was also argued that eliminating staff selection would go a long way towards reducing the burden of the exercise. It was noted that a significant part of the cost associated with the REF was borne by the institutions, and that the process of selecting staff was a central part this.

All staff employed in a submitting institution with research in their contract of employment should be submitted to the REF. Decision-making around which staff to submit to the REF incurs substantial opportunity costs for academics involved in the selection process, diverting time from research. Universities establish levels of research excellence against which to judge the research of individual staff members. This is either hugely burdensome, involving pilot REF exercises, or done on little more than guess work. The outcome of a such selection process is divisive within an institution, can be damaging to individual research careers, and is potentially discriminatory. In addition, partial submissions of researchers subvert the benchmarking role of the REF.

(Learned Society)

Respondents differed in their definition of ‘all staff’ in this context. Some respondents referred to ‘eligible’ or ‘research-active’ staff whereas others specifically referred to all HESA-returned academic staff, including both teaching and research staff. These suggestions were typically combined with other proposed changes, such as decoupling of outputs from individuals (see section 3.2 below) or sampling (see above).

Many other respondents, including HEIs and other organisations, supported the retention of the ability of institutions to select staff for submission to the REF. It was argued that if staff selection were to be removed, a larger volume of outputs would need to be reviewed thereby further increasing the burden on assessment panels. It was also felt that it would limit the possibilities for less research-intensive universities to develop areas of excellence “without being penalised for a non-research active ‘tail’” (HEI). Several respondents expressed concerns that a requirement for all eligible staff to be submitted could lead institutions to re-

write staff contracts as ‘teaching-only’ as a way to exclude them from the scope of the exercise. This could exacerbate the divisive and damaging effects of selection. The main arguments against submitting all staff are summed up in this response:

On balance, we are in favour of retaining the selective approach to the REF. This provides discretion for universities to make strategic decisions regarding their research profile. While requiring submission of 100% of research staff (or a minimum proportion) has many merits, not least eliminating the reduction in morale amongst those not submitted, this could undermine the ability of some institutions to develop capability and capacity, and nurture pockets of excellence that are important to the overall health, diversity and vitality of the research base. [...] Whilst gaming may be reduced, it could lead to unintended consequences whereby although the counterweight of encouraging careers for teaching led staff would be a benefit, significant proportions of staff are moved into teaching-only contracts; and may undermine the breadth, depth and vitality of the research base by making it more difficult for institutions to support emerging researchers. It is also unlikely to lead to a lower cost and administrative burden without a wholesale revision of the assessment methodology, as the number of outputs would necessarily rise without the introduction of alternatives such as sampling.

(Representative body)

Approximately half of all of the HEIs responding to the call for evidence commented on the issue of staff selection. Research-intensive universities were more likely to raise the issue and more likely to be in favour of removing staff selection. A number of respondents discussed the advantages and disadvantages of staff selection without taking a clear position.

Staff selection – key messages from stakeholder interviews

Stakeholders interviewed for the Review were specifically asked about their position on staff selection. Views were mixed but a small majority among them agreed that HEIs should submit all staff to the assessment process, rather than a selection. The largest, most research-intensive HEIs tended to favour full submission of all staff, where the less research-intensive universities tended to prefer to retain selectivity. These were the main views expressed by the interviewees (HEIs unless otherwise stated):

- For many, submitting all staff would give a more complete view of institutional performance and allow more credible / useful benchmarking across HEIs.
- One contributor argued that the rules on selectivity was the one aspect of REF 2014 that warranted a fundamental review, and that it might reasonably be the single focus for the evolution of the system.
- For many, including all staff would reduce university costs too, at least to some degree.
- Others suggest that the costs would not change dramatically, as HEIs would still be selecting outputs.
- It ought to eliminate some of the more negative consequences of selectivity, relating to the inclusion or exclusion of staff and the sometimes intrusive process by which to determine someone's special circumstances.
- Including all staff by default would be a positive development for early career researchers and those with special circumstances, and would be in line with the Concordat to Support the Career Development of Researchers.
- Most contributors consider that 'all staff' would mean all staff with 'research' in their contract, including researchers.
- Almost everyone commented on the risk that a move to submitting 'all staff' to REF may cause HEIs to move more staff onto teaching-only contracts, which was seen as being unfair on the individuals concerned and detrimental to the important interplay between research and teaching.
- Several suggested one might combat this movement of staff onto different contracts by including teaching-only staff, at least in the demographic data, within the submission. There was also a suggestion that this would also help when it came to considering who was included within the submission for any future Teaching Excellence Framework (TEF).
- Several stated that it would be helpful to break the link between outputs and people, so that one could submit all staff but not overwhelm the peer review with all outputs.
- Several people took a contrary position, and argued that selectivity had helped to identify pockets of excellence across the system and thereby encourage less research intensive universities to invest in developing their research capabilities; there is a concern that a move away from selectivity would favour the big players and hinder the transition of the many smaller players as they progress from teaching only to research and teaching HEIs.
- Several interviewees (HEIs, intermediaries and individuals) emphasised that any change in this direction should be carefully designed and piloted to avoid undesirable consequences.

2.1.3 Frequency of REF assessments

Several HEIs and other organisations suggested increasing the interval between assessment exercises, for example increasing it from five years to seven years. In a minority of cases, respondents suggested it might be increased to 10 years. A lower frequency would decrease the cost to the system overall and potentially alleviate some of the problems with ‘short-termism’ discussed below. Others argued that the 5-7 year periodicity of the current system felt about right, and was rather well aligned with the ‘metabolic rate’ of the research base; these contributors stated that a 3-year assessment was too frequent and almost continuous assessment, while a 10-year term was thought to be a little too long, and would lose some of the dynamism of the current arrangements. In one or two cases, people stated that a longer, 10-cycle for the full assessment could be combined with a light-touch interim assessment based more heavily on metrics, in order to strike a balance between the desire for economy and dynamism. Other contributors anticipated this suggestion and argued against the idea of alternating full and light-touch assessments, on the assumption that the latter would need to reduce the central role of peer review and HEI-led submissions, and rely to a greater extent on metrics.

2.1.4 Environment assessment

Many respondents, especially HEIs and learned societies, felt that the environment statement could be reworked and simplified.

There was a general sense that the narrative part of the template could be reduced and that metrics could be used to a larger extent. This would reduce burden of writing and assessing qualitative content and arguably increase the accuracy as the scope for unverifiable claims, described by some respondents as ‘creative writing’, was reduced. Many respondents were in favour of including the impact statement (but not the impact case studies) as part of the environment statement.

As discussed under sections 6, 7, and 8 below, some also supported an increased use the environment template to incentivise interdisciplinary and collaborative work, provide staff data to help address gaming and to enhance the forward-looking element of the assessment.

2.1.5 Other issues

A large number of other issues were brought up, for example:

- **Grading:** Some individuals, HEIs and other organisations commented on the issue of grade inflation, pointing out that an ever increasing proportion of submitted work was graded four star. As a consequence, some felt that the grading system was now unable to discriminate sufficiently between “the very best world-leading research and the rest” (HEI). Others asked for more specific guidance between 3 and 4 star ratings or more transparency in application of definitions across panels. A couple of respondents suggested a more granular grading scale, for example with half points
- **Types of outputs:** Some respondents, especially individuals and other organisations, commented on the position of monographs in the assessment of outputs. A clarification of the rules was requested and there were calls to allow a broader range of outputs that contributed to the development of disciplines, including textbooks, review articles and the reproduction of previous findings

- **Weighting:** Some HEIs, particularly post-1992 universities, called for the weighting for impact to be increased while a number of HEIs and other organisations supported keeping the current level.

Panel assessment and grading – views from stakeholder interviews

Several of our interviewees were concerned about the phenomenon they referred to as grade inflation, in particular between RAE2008 and REF 2014, but for some it was considered to be a longer-run issue.

- For most, the improving scores were a reflection of HEIs' improving abilities to identify and submit only that sub-set of their total research output they judged would be scored as being internationally outstanding. There was scepticism that the system overall is improving at quite the same rate as successive RAE and REF quality profiles suggest. People argued that the utility and credibility of the profiles was being weakened, and that it made benchmarking harder and especially for the international community.
- Monographs and impact case studies and environment statements were also mentioned by different respondents in this context, with a concern about the ease with which panels can work confidently with a measurement scale devised for individual scholars and articles, when looking at other forms of material; how does one judge a narrative based case study of a particular socio-economic impact as being "... world-leading in terms of originality, significance and rigour"? Several people argued that panels were insufficiently discriminating about research impact, awarding 4-star scores to too many impact case studies. One suggestion was to introduce a mechanism for post-calibration scores. Some people also suggested that more panels could adopt the practice of using a more fine-grained scale to inform debates among assessors, in order to agree on a final score that would then be reconciled with the established 5-point scale.
- The use of the grade point average (GPA) for league tables by the press also raised concern and it was argued that using rankings based on 'Research Power' or submitting all staff (see above) would help address this problem.

2.2 The definition of impact in the REF

Generally, contributors reported that the impact element was well received. A number of benefits were cited including increased internal recognition of impact, enhanced dialogue with users and more effective showcasing of the benefits of research.

Some commented on the increased workload associated with the preparation of the impact case studies and there were some suggestions to reduce number of case studies or allow re-submission of existing cases where new impact had occurred.

There was general support for a broad definition of impact that could capture impact in a variety of disciplines and socio-economic contexts. A significant number of respondents, including many individuals and other organisations as well as some HEIs, suggested a broadening of the current definition of impact. Some of the recurring suggestions were to include:

- Impact on teaching
- Impact on academia, science and disciplines

- Impact at the local or regional scale, without global ramifications.
- Impact of public engagement and dissemination

Several respondents commented that the underlying REF impact model failed to account for the non-linear nature of impact. Specifically, the requirement that impact case studies should be underpinned by specific pieces of excellent research, rated at least 2-star by REF and undertaken within the submitting unit, was seen to exclude a range of impactful research activity, including research-based consultancy and co-production of knowledge:

The current definition of impact supports a very limited (and linear) view of the use of university knowledge outside academia. Cultural, economic, social and health progress in society is normally a multi-disciplinary activity, informed by what we know, some of which is direct product of a specific piece of work. Just as research when done well answers a well formulated research question, so we have impact when we are addressing a specific societal need - that brings together all knowledge, not just the very small fraction of all relevant knowledge that was product of our research.

(Higher Education Institution)

The time between research and impact was also discussed. A number of respondents argued that the timeframe as currently defined was too short and failed to capture the long-term nature of some types of impact.

A mixed group of respondents also requested clarification and guidance on definitions and assessment criteria. Many observed that REF2014 impact case studies had employed a range of different measures of impact, which made comparison more difficult and left more scope for embellishment. Some therefore suggested some standardised measures, for example for economic impact, which could be developed through analysis of REF 2014 impact case studies. Suggestions were also made to align the definition impact in the REF with that used by the research councils.¹

A number of HEIs and other organisations argued in support of keeping the current definition and assessment format for impact case studies. It was seen as sufficiently broad and flexible, and it is argued that changing it would lead to further burden of readjustment. Stability would facilitate further learning and embedding of impact within institutions and academic communities.

¹ RCUK defines 'Academic Impact' as "the demonstrable contribution that excellent research makes to academic advances, across and within disciplines, including significant advances in understanding, methods, theory and application", and 'Economic and societal impact' as "The demonstrable contribution that excellent research makes to society and the economy." See <http://www.rcuk.ac.uk/innovation/impacts/>.

The impact agenda – views from stakeholder interviews

There was general support for the continued inclusion of a research impact module within future REFs, operating in a broadly similar manner to the arrangements used in 2014, however, this support was not universal. The following bullet points bring out the main points from our discussions, and as is the case elsewhere they do not come together in a singularly neat position; there are strongly contrasting opinions:

- Most of our discussion partners remarked that the impact element of REF 2014 had worked surprisingly well overall and that the manner in which it had been set up meant it turned out reasonably well across all disciplines. Paradoxically, having been a source of grave concern for many when it was first proposed, research impact is no longer a source of great anxiety for academics. Moreover, it has captured the imagination of the wider community, in particular in policy circles.
- Several contributors noted that the impact element had been the biggest single cost item for universities in REF, and that the next REF ought to try to reduce that overall cost.
- Larger universities suggested the numbers of case studies might reasonably be reduced, possibly fixing the numbers based on a lower coefficient (e.g. 1 to 10 or 1 to 15 staff) and judged against staff numbers at the level of the main panels rather than the individual sub-panels. One contributor suggested submitting the impact case studies to the Main Panels for assessment. Others argued that the nature of and pathways to impact are markedly different even between sub-fields and that this requires a level of domain knowledge and sensitivity that requires the case studies to be looked at sub-panel level.
- Research users were unanimous in their positive view of the process overall, and would welcome more emphasis being given to impact next time round. They remarked on the importance of retaining (ideally strengthening) a user perspective to ensure assessment panels understand the difference between outcomes and impacts: for example, where a scientist is invited to give evidence from their research to a high-level review group or advisory committee, that is not a policy impact in itself, and should arguably be rated as ‘unclassified’ rather than 3-star or 4-star.
- People believe the process will be easier next time round, as the concept is clearer and, critically, most HEIs have put in place institution-wide systems for collecting impact-related material continuously. There was however a keen interest in learning more about the rules for impact in the next REF, as early as possible, and in particular around the treatment of case studies that were submitted originally in REF 2014. Impact is cumulative, and contributors believe that a proportion of their ‘old’ case studies will have continued to mature and that the story may be even more impressive in 2021, as compared with 2014. Will it be possible to reuse case studies, and how will the incremental improvement be demonstrated and tested?
- Some interviewees criticised the impact element for being ‘misguided,’ with a multiplicity of factors affecting the nature and extent of any wider impact; often there is only a weak link between success in the wider world and the excellence of specific research activities. These individuals considered research impact to be an unnecessary additional cost and complication for a national assessment process designed primarily to determine the allocation of QR funding among institutions.

(The impact agenda – views from stakeholder interviews, continued)

- Others criticised the assessment process for lacking robustness, pointing on the one hand to the highly particular and often ‘messy’ nature of the case study narratives, and, on the other, to the ease with which panels were able to work with a measurement scale designed for grading research outputs. According to one interviewee, an analysis of the results of the assessment of impact failed to find any correlation to HEI’s quality scores, which may reflect a fundamental problem with the link between good research and good impact, or more likely, the immaturity and subjectivity of the impact assessment the process. There was also concern that the novelty of impact assessment and the absence of calibration of impact scores across broad fields, meant that some units of assessment had found it very much easier than others to come forward with self-evidently ‘significant’ examples of impact, where other disciplines had had to work very much harder.
- Several people remarked on extent of the differences in scoring between medical research (Main Panel A) and the other Panels, which they considered to be wrong in principle: all things being equal, a UK-wide assessment of research impact in the humanities should largely mirror the national profiles for impact in the social sciences, where the assessment is being carried out by field-level specialists. People recommended that more work needs to be done on the detailed guidelines for preparing and assessing impact case studies.
- Two people remarked on the risks associated with calling for impact to be demonstrated across all disciplines and fields, and the potential for this new focus to feed forward into HEI strategies and favouring larger departments with a much stronger emphasis on applied research. In many areas, more fundamental and theoretical research produces impact in a more diffuse fashion in time and space and will have gestation periods for some of the most profound contributions that run in to many decades. The current arrangements are likely to mean HEIs must exclude some of their staff’s most significant contributions to wider social or economic outcomes.
- There was widespread interest in loosening the REF 2014 rules that require HEIs to link each impact case study to specific pieces of (excellent) research. Several universities argued that some of their institution’s most notable social impacts relate to bodies of work, which accumulate over time, and cannot be reduced to a single piece of science. There was also interest in exploring ways in which universities’ research infrastructure and data holdings – built through research and for research – might be eligible in future.
- While nowhere near universal, there was wide-ranging support for giving more weight to impact within the overall assessment scores, with many contributors – from different stakeholder groups – suggesting impact might reasonably be used to drive 25% of the total allocation next time round, with the additional 5% switching from research excellence rather than environment.
- One academic argued that far more weight ought to be given to impact – within the overall REF assessment process – as this is the ultimate test of the relevance, quality, rigour of academic science. The individual concerned argued that impact is testable in a way that science is not, in particular fields at least, where there is a growing crisis of replicability (where scientific experiments prove difficult or impossible to replicate, even where that work was highly cited and widely influential).
- Others disagreed strongly with the suggestion that impact ought to be used to determine the distribution of even more of the total QR pot, arguing on the one hand that there are too many external factors at play (in determining the scale of success of any impact) that lie outside the influence of HEIs and, on the other, that the assessment process is still in its infancy and somewhat subjective. It was thought to have had an unduly large influence on the final QR allocations for REF 2014.

2.3 Peer review and metrics

2.3.1 *Attitude towards peer review*

A clear majority of respondents expressed support for keeping peer review at the heart of the assessment process. It was seen as the only robust method available, and the only one to command the respect of the academic community.

Some individual academics criticised the way in which peer review was run in the REF. For example, it was argued – with evidence from at least one sub panel comparing citation scores with REF panel assessments – that REF peer reviews were biased with respect to discipline sub-areas (e.g. theoretical work favoured over empirical work), types of institutions and gender (male researchers favoured over female researchers). Several respondents suggested ways to counter such bias, for example through blind submissions, vetting and anti-bias training of panel members.

Concerns were also raised about the limited time available to panel members to assess each output. One respondent pointed out that most research outputs were peer reviewed prior to publication and questioned the value of reviewing the outputs a second time for the REF. As a way to alleviate the burden on panel members, it was therefore suggested that REF coordinate with journal peer reviewers to avoid reviewing the same work twice.

Peer review – views from stakeholder interviews

Interviewees from all groups agreed that the RAE and REF had been right to insist on retaining peer review at the heart of the assessment process and to refuse to switch to a system based on metrics only. There was also a strong sense across most contributors that HEFCE's commitment to retain the HE community at the centre of the process – submissions and assessments – had been central to the level of confidence engendered and the very low levels of disagreement voiced as regards the correctness of individual assessments. There were several further reflections, which were of interest:

- While everyone agreed that peer review should be retained as the primary means of assessment, a significant proportion of all interviewees went on to note that the REF peer review process operates at a level that is challenging and will typically require panels and individual reviews to devise strategies to cope with (i) the large volume of material to be assessed and (ii) the diversity of subject matter under consideration: it is not peer review as understood from grant applications or journal reviewers. Several panel members suggested that this can lead to certain biases, whether that is to favour one kind of research over another or to rely on reputational factors of whole departments or journals. Many people suggested that reviewers were often falling back on journal impact factors, contrary to their briefings.
- Some interviewees expressed concerns about the ability of panels to cope with the volume of papers. Some HEIs suggested reorganising the peer review process: the wider community could be brought in with a first assessment by members of a national college and a second, more in-depth assessment for those where there is a high degree of variance between reviewers or those at the margins of the scoring threshold.
- While there were criticisms of the limitations of the Peer Review process within REF, there were suggestions that the process works rather more robustly and consistently for research outputs as compared with the assessments of the Environment Statements or Impact Case Studies: a wide-ranging statistical analysis performed by one individual interviewee suggested that the peer review assessment of research outputs was more robust than the assessment of environment and impact assessments. The environment assessment appeared to be strongly correlated with size and the impact assessment widely scattered, possibly favouring the panellists' home institutions. Another contributor has analysed the published REF results in computer science and informatics, which he suggests reveals biases in the peer review process for research outputs, favouring theoretical over applied research submissions, research-intensive over other institutions and even gender.
- Others noted that the move to a smaller number of sub-panels, for REF 2014, had led to more cases where panels brought together experts from widely differing academic traditions, which created some difficulties with intra-group dynamics but also meant there were effectively two or three separate pools of peer reviewers within the REF subpanel which made it harder in practical terms to run the evaluation process.

2.3.2 Attitude towards the use of metrics

Metrics were not generally supported as the main method of assessment. Frequent references were made to the Metric Tide report² to argue that metrics would be unable to supplant peer review at this point.

A few HEIs argued that metrics could help improve efficiency and remove some of the biases inherent in the peer review system but a broad cross-section of respondents were critical. The respondents identified a number of problems with the use of metrics for assessment purposes:

- Metrics are seen to be generally unreliable, insufficiently developed, they do not measure quality and are compromised by international collaboration
- A large number of respondents warned that increased use of metrics would disadvantage certain disciplines, particularly humanities, arts, social sciences where research is often published in forms that do not feature in bibliometric databases. In smaller disciplines, the fact that research communities have relatively few members overall means that citation counts would be comparably low, even for excellent research.
- Metrics promote certain types of research and privilege mainstream research or 'hot topics' where high citation counts are more likely over risk-taking and exploration of new areas.
- Metrics used on their own would promote game-playing behaviour, such as 'citation clubs'.
- Some warn that any new metrics introduced for the REF would be used by institutions for unrelated purposes such as staff management.
- It is also argued that exclusive reliance on metrics could lead to dramatic fluctuations in results from one assessment to the next.

2.3.2.1 Potential role for metrics within a system based on peer review

A large number of respondents saw a, possibly increased, supporting role for metrics within the current peer review system.

In the assessment of research outputs, there was widespread support among HEIs for using metrics to support peer review, potentially increasing its role compared to the REF2014. It was generally not seen as appropriate to impose a uniform set of metrics across all panels because of problems with coverage and relevance across disciplines. For the purpose of assessing outputs, any use of metrics should be determined by the panels on the basis of what was deemed appropriate within the disciplines concerned. Some respondents asked for more transparency in the way metrics are used by the panels and that the use within HEIs and REF panels of journal impact factors as a proxy for quality should be actively discouraged.

Many saw a larger scope for using metrics within the environment section, as discussed above. Some argued that metrics are more reliable than narrative statements in the context of the environment submission, and support assigning greater weight to quantitative data in the

² Wilsdon, J., et al. (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/Title.104463.en.html>).

assessment. Suggestions for new metrics to be included in the environment statement included data on collaboration, co-authorships and other bibliometric indicators. This issue also relates to the suggestion of providing aggregate environment submissions at the institutional level, and with the widespread wish to see closer coordination of data collection and use between national funders (e.g. ResearchFish, HESA, and RCUK data).

As discussed above, some respondents also called for a larger role for metrics within the impact case studies, for example in the form of standardised evidence (quantitative or otherwise) to improve comparability. This was more controversial, however, as many argued that this would narrow the scope of impact considered and favour quantifiable economic impact over other types.

2.3.3 Suggestions for more extensive use of metrics

A small number of respondents submitted outlines of assessment system models reliant on more extensive use of metrics. One type of system was related to the issue of sampling (see above). It would combine a metrics-based assessment of all outputs with qualitative peer review of selected outputs. Another suggestion was to cut costs by increasing the time between peer-review-based REF assessment and add a metrics-based interim assessment. This option was discussed above (section 2.1). Finally, some respondents expressed the hope that future metrics will overcome current shortcomings and be able to be adopted more widely.

Increased use of metrics – views from stakeholder interviews

Overall, the interviewees supported the conclusions from the Metric Tide report (see above) and insisted on REF using peer review as its primary assessment methodology and that the use metrics must be done with full recognition of their strengths and weaknesses:

- There was general support for increased use of metrics to inform the peer review process among HEIs, intermediaries and research users, whereas individual academics remained deeply antagonistic to the idea of metrics in any form. Several argued that bibliometrics had become less robust in recent times – where it had been predicted they would become very much safer over time – as individuals learn to game the system (e.g. by citation networks and new players in other parts of the world flood the system).
- Many contributors argued that the progress with information systems and performance metrics should enable REF to benefit from greater use of a variety of performance metrics and ratios (KPIs) within each pillar of the assessment process, including for example PGR completion rates (environment statement), field-weighted citation impact (research excellence) and sales and employment (impact).
- Several interviewees argued that the next REF should insist on panels making fuller use of data and metrics in their assessment process, to reduce subjectivity and challenge preconceptions. People believe it should be possible to prepare standard metrics (analytical briefings) for all sub-panels, and that Panel chairs should require members to consider these different data when carrying out their individual and group assessments.
- People acknowledge there is a need to allow flexibility to cope with the very different dissemination and communication models used across fields (e.g. performing arts, or practice-based research as compared with say experimental physics).
- Bibliometrics were seen to be less useful in certain disciplines (e.g. arts and humanities). Still, it was seen as important to keep all disciplines on the same terms. A dual system whereby metrics are used for STEM subjects but not for arts and humanities could lead to a loss of status for the latter disciplines.
- There was a sense that these very real differences had created something of a vacuum as regards when and how to make use of metrics and paradoxically that there were various instances where reviewers were using metrics in ways that most specialists would argue were unsafe.
- Several people also argued that bibliometrics could help with the sampling of papers for peer review, albeit there would be a need to think about how best to deal with the issue of time lags (the lack of time to accumulate citations for recently published outputs).

3 Level of assessment

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?

3.1 Units of Assessment (UoAs)

Most of the respondents who answered this question, across different groups, expressed support for keeping the current UoAs or, at least, not reducing the number of UoAs any further.

Particularly for the assessment of research outputs, it was argued that fewer, larger panels would be too broad and inadequately able to perform ‘expert’ reviews that reflect the research culture and methodology of scientific disciplines. Further, the UoA-level assessments provided important benchmarks at the discipline level. It was also argued that keeping the UoA structure relatively stable would allow for better longitudinal comparison.

Some respondents commented that the current UoAs had failed to capture the structure of university departments or disciplines. In some cases, it was felt that the aggregation of panels in REF2014 compared to RAE 2008 has disadvantaged smaller subjects which no longer have a dedicated UoA and led to “the amalgamation of some non-cognate disciplines” (HEI respondent). On that basis, some argued that there was scope for adjustments, while a few respondents requested an increase in the overall number of UoAs, for example to the situation in RAE2008.

In contrast, several respondents argued that a further consolidation of UoAs would be welcome, as broader panels would be better able to accommodate interdisciplinary research and would reduce the need for tactical considerations by HEIs about where to submit research.

Level of assessment – views from stakeholder interviews

Overall, the interviewees supported the arguments put forward in the written evidence.

- People agreed that research excellence must be tackled through peer review and that this needs to be closely aligned with the relevant cognate disciplines. The UoA is the appropriate level of analysis, and in some cases is arguably already too granular.
- The great majority of people believe the environment statement can be tackled at a higher level than a UoA, possibly at the level of the Main Panels. In several cases, people suggest it might be tackled at the institutional level – with one statement for each HEI – reflecting the origins of much of the content. Others argued that this would miss important points of difference as regards institutional strategies for the research environment in say the arts as compared with engineering. They argued that the modes – and norms – for things like researcher training, international cooperation, knowledge transfer, etc. are quite distinct across broad disciplines, and that while every panel may be content to have a copy of a university’s institution-level environment statement, they would also want to see a more specific presentation of the particular strategies of for example the medical school (but shorter and possibly more heavily metricised).
- Most contributors took the view that the Impact Template was of limited use as constructed, and might be omitted next time round or slimmed down and included as a sub-section within the environment statement.
- Most contributors also took the view that the Impact Case Studies were appropriately located at the level of the UoAs, give the potential differences in impact pathways and impact types across disciplines, although several people thought the case studies might easily be looked at by the Main Panels, as a workable compromise.
- Three universities stood apart from the mainstream view and argued that the next REF ought to try to get back to basics, and should run all assessments at an institutional level as the default, given the purpose is primarily to determine the distribution of QR by institution. Other contributors believe such a granular approach would obscure the pockets of excellence in the long tail of smaller, less research intensive universities, favouring the big players and derailing others’ efforts to strengthen research across the community.

3.2 Linking outputs to individuals

3.2.1 Arguments for de-coupling outputs and individuals

This issue was related to the question of selectivity discussed above and attracted a number of responses.

Many respondents, mostly HEIs and other organisations, suggested de-coupling outputs from individuals. Instead of requiring four outputs per individual, this would require the submission of a defined number of outputs for the department or group as a whole (e.g. based on the number of FTE at the census date or as an average over the REF period).

De-coupling staff and outputs would, in most suggested models, eliminate the costly process of staff selection and provide everyone an opportunity to contribute, including individual academics without four relevant outputs to submit. It would remove the ‘in-or-out’ nature of the exercise. It would also eliminate the need to justify individual staff circumstances, as institutions would have the flexibility to accommodate them within their overall quota.

It was also argued by some HEIs that it would give a more accurate picture of the unit as a group and give a better basis for comparing across units. Finally, it was argued that de-coupling outputs from individuals could facilitate the recognition of collaborative work as focus would be on the outputs rather than on the individual author. Individuals would not be limited to a single UoA, as different outputs could be submitted to different panels.

Rather than a complete de-coupling, some suggested models which would lead to a partial de-coupling, with a minimum number of outputs to be submitted from each individual combined with a total number for the unit.

3.2.2 Arguments for retaining the link between outputs and individuals

In favour of keeping the current link between outputs and individuals, it was argued by some respondents from all groups, that the link incentivised researchers to aspire to excellence. The current system was deemed to strike the right balance between quality and quantity and had been proven to work.

Many respondents were also concerned that a de-coupling of outputs from individual researchers would lead to a situation in which fewer researchers would be allowed to account for a higher proportion of the outputs submitted by a unit. The REF would then give a less representative picture of the broad group of researchers and lead to further division between research-active and non-active staff. It could also exacerbate the problem with 'poaching' as the potential value of a highly productive individual would increase.

In addition, some argued that determining the number of outputs to be submitted on the basis of FTE alone could lead to the underrepresentation of certain groups of staff or disadvantage units with a high proportion of staff with special individual circumstances.

3.2.3 The assessment of individual staff

A moderate number of respondents also commented on whether or not individual results of the assessment of outputs should be made available. A small number of institutions argued that this would be useful management information (see also section 0 below) and eliminate the 'guessing game' about individual scores. Other respondents, including both HEIs and individuals, argued against it on the grounds scores could be used to penalise individuals and have a demoralising effects on departments.

People and outputs – views from stakeholder interviews

Overall, the interviewees supported the arguments put forward in the written evidence and had little to add.

- People agreed that the assessment should be of institutions / departments and not individuals.
- They also agreed (see above) that overall, and in principle, it would be beneficial to include all staff in any HEI's submission.
- Breaking the link between staff and outputs would allow HEFCE to include all people without including all outputs, which may be unmanageable (without some additional step to randomly sample submissions). It also holds out the promise of reducing some of the personnel challenges related to early career researchers and special circumstances. One must avoid creating a situation however where a small proportion of a department's staff emerge as the super researchers responsible for the lion's share of research output and QR income; this would be divisive.

3.3 Aggregation

3.3.1 Assessment of research outputs at a more aggregate level

There appeared to be limited appetite for a wholesale aggregation of the assessment of research outputs at institutional level. The main arguments put forward in favour of an all institutional-level assessment were that it would reduce the cost of the assessment and that it would benefit interdisciplinary research and collaboration if the need to fit into discipline-based units of assessment was removed. It was also argued that it would reduce game-playing as universities would no longer need to attempt to maximise their scores through the choice of UoAs.

A large number of respondents across all three groups took the opposite view. Among other things, it was felt that further aggregation would prevent 'pockets of excellence' from being recognised and rewarded as the scores of high-performing units could be diluted within less research-intensive institutions. Thus, aggregation could privilege large, research-intensive institutions over others. The aforementioned benchmarking at the UoA-level, valued by many respondents, would also be lost.

3.3.2 Assessment of research environments and impacts at a more aggregate level

With respect to the impact and environment dimensions, some respondents saw assessment at a more aggregate level as a realistic option.

In favour of aggregating these dimensions at the institutional level, it was argued that it would reduce the burden of preparing the statements and avoid duplication as UoA-level templates already now tended to contain many of the same elements across institutional departments. Although not all respondents agreed, some also argued that assessment at an aggregate level would be more accurate as research environments and impact support systems were shaped by factors that cut across individual departments. For example:

... current Units of Assessment do not reflect our structures; this must be true of many institutions. The presentation of facts about the research environment on a wider scale – perhaps the whole university – would be both simpler and a better reflection of the way in which resources are shared. It is certainly arguable that the [university] research environment’ is more critical than any individual faculty’s or department’s environment (let alone an artificial Unit of Assessment).

(Individual respondent)

Several intermediary models for aggregating environment and impact elements were suggested:

- Environment and impact templates could be reported at the level of the one of the four main panels,³ which would give more room for interdisciplinary and collaborative research.
- Impact case studies could also be submitted to main panels to provide efficiency gains and allow for a reduction in the number of impact case studies per FTE without disproportionate effects on smaller UoAs (see also the discussion on gaming below).
- A ‘hybrid’ model would see certain aspects reported at an aggregate level and others at the UoA level, for example, “a metrics-based environment assessment at institutional level, supplemented by short narratives on environment at UoA level. These narratives should allow for the highlighting of otherwise obscured pockets of excellence.” (Higher Education Institution).
- Suggestions were also made to set up dedicated panels to assess interdisciplinary and collaborative research (see also section o below).

³ In REF2014, 36 sub-panels (Units of Assessment) conducted the detailed assessment of submissions. The role of the four main panels was to provide leadership and guidance. For details, see: REF 01.2010, “Units of assessment and recruitment of expert panels”, p. 4 (available at: <http://www.ref.ac.uk/pubs/2010-01/>).

4 Information for HEIs

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?

4.1 Current use of information gathered through the REF by organisations

According to institutional responses, primarily from HEIs but also some other organisations, REF information was useful for a variety of purposes. In this context, ‘REF information’ was used to refer both to information prepared by the HEIs in preparation of REF as well as the results from the REF assessment itself.

REF information informed strategic planning in most institutions. Many saw this as a positive influence that helped give a broad picture of institutional research and drive a focus on performance and quality. The REF could provide framework for and a periodical external check on internal performance monitoring. In recent years, the REF had also helped institutions integrate issues such as diversity, equality, open access and research impact.

Many institutions also used REF information to inform decisions about resource allocation and to identify future areas of investment. In several institutions, the internal allocation of QR funding was closely linked to REF results, for example as a function of the number of staff submitted or the score obtained.

The REF results were also valued by many institutions as an external benchmarking used to compare performance against previous assessment exercises and against other institutions, at the institutional and departmental levels. As a national standardised benchmark, this was seen as valuable, independently of the QR funding attached to it.

Assessment results were used by many institutions for marketing and promotional purposes. For example, improved scores are advertised and the impact case studies are used to showcase the impact of research. Among other things, these strategies were used to attract overseas post-graduate students and to prove credentials to potential external collaborators.

Some institutions also used REF data for staff management purposes, including hiring, promotions and the allocation of research time. Several other HEIs stated specifically that they refrained from using it for promotion cases and staff performance reviews. Many individual academics felt this to be an inappropriate level of ‘micro-management’, especially since assessment results of individual outputs remained unknown to institutions. REF information – described by one respondent as “half-guessed interpretations of the REF results, which themselves are highly error prone and potentially biased at this level of detail” (individual academic) – was not seen as a legitimate basis for such decisions.

Some institutions commented that REF information has limited use, citing limitations with respect to international comparability and assessment criteria used.

Information for HEIs – views from stakeholder interviews

Overall, the interviewees restated many of the arguments put forward in the written evidence. These are some of the main points:

- Universities stated that the RAE and REF had provided UK HEIs with the incentive to become better at managing their research, and that most now have the management information systems and decision-making structures necessary to be more strategic – if they so want – and generally slicker and sharper in their commitment to carry out worldclass research.
- HEIs do not rely on REF to provide them with this kind of strategic intelligence, as the national assessment exercise runs too infrequently for that. For most, the principles and assessment criteria derived from submitting to REF (and making applications to other research funders) are now embedded in their institutional research management, HR and QA systems. They generate many of the REF data continuously and monitor / report on those KPIs on a regular basis, and not only when required for the REF. In many cases, HEIs will also be benchmarking themselves against national and international ‘competitors’ on an annual basis. Paradoxically, there is a much heavier reliance on metrics within these institutional M&E systems.
- The individual REF results are used by most HEIs both to calibrate internal perceptions’ of their overall / UoA strengths and weaknesses, which may be cause for celebration or further inquiry. The press-compiled rankings are studied closely too, but the tendency to use GPAs reduces the positive value (benchmarking) and this tends to be looked at as a possible threat to one’s reputation that needs to be addressed through additional communication efforts. In most cases, universities will use their results and their rankings selectively to promote the university or the department.
- The impact case studies are also widely reused in publicity materials, on web sites and in grant applications. It is not clear they are being used by prospective students or industrial partners to benchmark or compare HEIs; indeed the research users interviewed said the case studies had helped give them an overview of their wider (historical) engagement with the HE sector, but were not particularly relevant to spotting who to collaborate with. They know that already.
- People agreed that the assessment should be of institutions / departments and not individuals. They also agreed (see above) that overall, and in principle, it would be beneficial to include all staff in any HEI’s submission. Breaking the link between staff and outputs would allow HEFCE to include all people without including all outputs, which may be unmanageable (without some additional step to randomly sample submissions). It also holds out the promise of reducing some of the personnel challenges related to early career researchers and special circumstances. One must avoid creating a situation however where a small proportion of a department’s staff emerge as the super researchers responsible for the lion’s share of research output and QR income; this would be divisive.

4.2 Suggestions for information that could be more useful to organisations

Many institutions commented that the feedback provided by REF panels could be more helpful if it was more detailed and of better quality. The feedback was felt by some to be insufficient considering how much time goes into preparing the submissions. Specifically, several respondents asked for feedback at a more disaggregate level, for example at the level

of sub disciplines or research grouping. Some also suggested that the scores of individual outputs and impact case studies be made available to the institutions on a confidential basis (see also section 3.2.3)

It was also suggested by a few respondents that some REF information could be updated on a more regular basis. While REF information was important to many institutions, its usefulness waned with time between assessments. Regularly updated information could be based on less resource-intensive methods using metrics but there were no clear suggestions as to how this might be done. The related idea of a metrics-based interim assessment between full REF assessments is discussed below.

As discussed below in the following section, it was also felt that more could be done to use the information generated by the REF, and to align REF data requirements with other existing sources to improve efficiency and consistency.

International benchmarks – views from stakeholder interviews

HEIs and intermediary organisations agreed that the REF ‘grades’ lacked comparability with any international indices or scoreboards, and are therefore of limited utility to any UK institutions looking to climb the international rankings. However, this was not a particular concern for stakeholders and it was questioned whether anything could or should be done to bring REF scores in line with international metrics.

- It was also pointed out that international benchmarks already exist – the QS World University Ranking was mentioned – and that many individual HEIs fund their own targeted benchmarking, comparing individual department’s performance with their preferred benchmark departments globally.
- People also remarked on the extent to which the RAE has been widely studied and adopted internationally, albeit there is a belief that most give rather more weight to metrics (for cost and efficiency reasons) and that the primacy of peer review within the UK’s assessment process marks it out as one of the better and most robust assessment systems.
- Most took the view that the inclusion of international experts within the main panels had provided invaluable guidance and calibration for sub-panels. It was also reported that the international peers had found the scoring of papers too ‘hawkish,’ suggesting that the quality ratings awarded tended to be lower than the equivalent work would have achieved internationally. This suggests to people that grade-inflation is more likely to be the product of good selection, as opposed to unduly favourable or biased domestic appraisal.

4.3 Information from REF and other management information

There was little sense that information from the REF *duplicated* existing institutional information and several respondents stated that they would have needed to collect similar information were it not already done for the REF.

...we feel the information gathered for the REF provides quality data to inform future strategies for research and innovation, course development and resource allocation, and we would wish to gather this data even if there was no REF.

(Higher Education Institution)

From institutional responses, it was not clear that information from the REF generally took priority over other existing information as a basis for decision-making at HEIs, although a few institutions confirmed that this was the case. Rather, as noted by one respondent, “Over time institutions have embedded internal assessment mechanisms that build on the characteristics of RAE/REF.” (Higher Education Institution).

Some respondents, especially individual academics, expressed concerns about the degree to which REF information was used, some arguing that REF information was ‘driving’ rather than ‘informing’ institutional strategies. In addition to the issues discussed above, this was seen as a risk because of the uncertainty about future REF assessments and the lack of information about its internal workings.

5 Information to be collected by government

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

5.1 Caution against adding additional data requirements

Not all respondents provided an answer to this question and there were relatively few responses from individual academics. Several respondents objected to the premise of the question, arguing that government should refrain from attempts to ‘drive’ research and that funding decisions should be taken at arms’ length.

Many institutions and several individuals believed that the data currently collected through the REF were sufficient, enabling the governments and research funders to hold universities accountable and to make informed decisions about future investments. The UK was seen a very efficient system that does not need ‘fixing’. Institutions were concerned that the burdens imposed on them by addition data requirements would outweigh any benefit flowing from it. For example, one HEI wrote:

Caution should be exercised in adding data requirements to the REF to fulfil a diverse ‘wish list’, given that each subsequent RAE/REF has become more complex and burdensome.

(Higher Education Institution)

It was also urged that any new data requirements should be published as soon as possible to allow time for institutions to adjust their systems and processes.

5.2 Better use of existing data

Rather than collecting new data, many HEIs and other organisations saw scope for making better use of existing data.

5.2.1 Better use of data collected through the REF

The REF2014 impact case studies have already been analysed and information from them made available in a variety of formats. There were suggestions that more REF data should be made publically available, for example more granular data on research outputs and the scores of impact case studies. One purpose would be to further showcase results of research in various disciplines. More could also be done to promote the use of REF information by industry:

Industry partners and companies have not widely used the information from the REF exercise [...] To get more value from the REF process its outputs must be used more widely. To do this the REF must be better publicised and thus the REF could help industry, professional bodies, research funders and Government to better understand the research landscape.

(Professional body)

It was further suggested that REF data should be used to guide government policy. It could be used to map strengths and weaknesses and to identify areas in need of additional investment from other public sources, e.g. the research councils.

5.2.2 National coordination and use of other existing data

Many respondents suggested using other existing data sources, such as HE-BCI, Researchfish and ORCID, instead of adding new requirements to REF. A further step, suggested by many, would be to align and/or ultimately integrate REF data with other sources, such as HESA and research councils. Ultimately, implementing common standards at a national level was seen as a way to make research information systems more interoperable and potentially reduce the burden of the exercise significantly.

With respect to impact, there were suggestions that economic impact should be traced more consistently (see under metrics above).

5.3 Suggestions for new data to be collected

Respondents across all groups made a number of suggestions about new data to be collected, but there was no clear consensus on any one specific indicator:

- **Academic activities and outputs:** Some suggested that the REF is used to collect data on wider range of academic and outputs. For example, it was argued that more should be done to capture work in academic societies and journal editorships. This could, for example, be reported as part of the environment template.
- **External funding:** Many suggested that funding sources should be added to REF submissions in a more standardised and searchable format. This would assist research councils in future planning.
- **'Value for Money':** Several individuals as well as some institutions suggested that a measure of 'Value for Money' is introduced. This should relate the production and quality of research outputs to the funding (inputs) employed to produce them.
- **Collaboration and engagement:** Several suggestions were made for a more systematic collection of data on different types of collaboration and public engagement in addition to what is already contained in the environment and impact templates.
- **International indicators:** Some respondents suggested a more international dimension to the data collected. This could include data on international esteem or collaboration, or data on UK institutions in a format that is comparable internationally. It was noted that

the validity of the comparative international performance indicators currently collected by BIS is compromised by cultural differences in citation dynamics worldwide.

- Impact on teaching and early career researchers: Several respondents argued for more data to be collected about the impact on teaching, e.g. curriculum innovation, support and development of research students and early career researchers through research activities linked to the REF as well as the career destination and impact of research students and early career researchers.

Information for government – views from stakeholder interviews

The great majority of contributors had very little to say on this point, and many argued that REF should not be used as a means by which to generate additional data and intelligence just for policy makers. The main points were:

- For most people, there is no justification for adding cost and complexity to an already costly and complex process, in order to benefit other parties.
- Research funders, perhaps unsurprisingly, take a different view and have used RAE and REF derived data and information to inform their own thinking about a wide range of important topics, from vulnerable subjects to challenge-based multidisciplinary research through to institutional capability building. They would like to see the next REF make a very much better job of separating out and dimensioning various topics of interest, particularly within the research environment and research impact pillars. They have an interest in increased standardisation and inter-operability, through use of common identifiers and points of reference. There is also an interest in knowing more about HEI strategy, and how they are explicitly tackling a wide range of research policy priorities, from doctoral training to gender / diversity to inter-sectoral mobility to international collaboration. There was a suggestion that the additional effort required to report on a longer list of important dimensions might usefully be offset by a switch to reporting this material at an institutional level rather than looking to every school or department to try to construct such a document.
- Several universities and intermediaries attempted to pre-empt this request for more information for government, arguing that this was a distraction from the primary purpose and that while any one piece of additional information would have a tiny cost, the accretion of multiple nice to know, can we have data requests does have a material effect on compliance costs.
- Several panellists, research users and universities noted that there is a lot more information being collected by HEIs – and others – and returned to HESA and other data holders, which could be re-used (with care) to provide additional structured data of value to assessors (and maybe government) without further burdening institutions.

6 Support for interdisciplinary research and collaboration

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?

6.1 Promoting interdisciplinary research

A number of respondents, including individual respondents, HEIs and other organisations, reported that interdisciplinary work was disadvantaged by the REF. In particular, the disciplinary ‘silos’ embodied in the panel structures were seen to inhibit the recognition and reward of interdisciplinary work. Others stated that the main problem holding back interdisciplinary work in the context of the REF was the misperception among individuals and institutions that it was poorly received and marked lower than other research.

A number of potential solutions were suggested. For example:

- Use the environment template to better report and reward interdisciplinary work
- Use the ‘interdisciplinary’ tag on submission forms more consistently
- Appoint ‘interdisciplinary champions’ or individuals with experience from interdisciplinary research to assessment panels
- Improve the mechanisms for cross-referral
- Enlarge panels to accommodate interdisciplinary work, for example assessment of impact case studies by main panels
- Set up dedicated panels for interdisciplinary research
- Introduce score enhancement for interdisciplinary work

Interdisciplinary research (IDR) – views from stakeholder interviews

IDR was an issue of great interest for most if not all interviewees, however, no one had a clear view on how REF could overcome the challenges once and for all. It was an area where people were rather tentative about their suggestions for improvements, in most cases noting that their ideas were only half-formed and would need careful development and testing before they might be implemented safely. Some of the suggestions are listed below:

- The Environment Statement could include a section on IDR and section that explains how / why / what IDR is in hand, and possibly introducing the separate scoring of that element.
- The Main Panels include a good cross-section of very senior / experienced people that together cover an incredibly broad spectrum of disciplines; they have the domain breadth and wisdom to tackle the assessment of IDR, and they might be given the task formally of assessing all IDR outputs.
- Others suggested dedicated, challenge-based IDR panels to work in parallel with discipline-based panels. Others were sceptical of this idea, based experience with REF2014 panels with something of an IDR quality (e.g. Sub-Panel D27, Area Studies). Such panels became the destination for papers that were ‘not elsewhere classified’ and it was not felt that these papers were treated well.
- IDR should be encouraged through better communication and refinement of existing processes, not through the introduction of more rules.
- Our non-academic contributors were especially exercised by the idea that REF may be causing HEIs to exclude their IDR work on the grounds that it is harder to predict how it will be graded and by implication encouraging researchers to prefer monodisciplinary, non-collaborative work. This was considered to be particularly problematic for medical research, which has a strong tradition of inter- and multi-disciplinary research carried out between universities and clinicians. Industrialists (and government departments) also expressed a desire to see the next REF do more again to encourage the submission of IDR and complex, challenge based work.
- Interviewees from all groups saw the difficulty of assessing IDR as an issue for all funding systems, not just the REF. HEIs pointed out that RAE and REF had already done a lot through bigger UoAs and specific processes. Some individual researchers also felt that IDR should happen ‘bottom-up’ and that positive discrimination for IDR would produce trivial work at the interfaces of disciplines from researchers simply attempting to find an easier route to securing better appraisals and more QR funding.

6.2 Collaboration between universities and other public and private sector bodies

Collaboration between universities, or even between units within the same university, was seen by some to be discouraged by the competitive nature of the REF. The focus on individual outputs was also seen as a barrier to collaboration. Several respondents suggested simplifying the rules for co-submission of impact case studies as one avenue to facilitate collaboration.

The introduction of the impact dimension in REF2014 was seen to have provided an impetus for collaboration: REF has been a driver for collaboration with non-academic institutions; it has created a cultural shift inside academia and provided a strong incentive to collaborate with partners from outside academia. Several respondents pointed towards other drivers for

university-industry collaboration, not least specific EU and government funding for industry collaboration.

Several respondents also drew attention to the issue of inter-sector mobility, making reference to the Dowling Review.⁴ Again, it was suggested that the environment template could be used to draw attention to the issue. It is also suggested that staff moving into academia from another sector could be allowed to submit a reduced number of outputs.

Several respondents cautioned not to put in place too many administrative (data) or legal (IP) requirements, as it would risk dissuading companies from engaging with HEIs. More generally, respondents felt that collaboration as well as interdisciplinary research should grow naturally without targeted incentives. A relatively large group of respondents including both institutions and individual academics argued that the government should refrain from trying to incentivise certain behaviours and instead leave it to the research councils and other funders provide targeted grant funding.

⁴ The Dowling Review of Business-University Research Collaborations', Department for Business, Innovation & Skills, July 2015 (available at: <https://www.gov.uk/government/publications/business-university-research-collaborations-dowling-review-final-report>).

Collaboration with research users – views from stakeholder interviews

Research users from government and industry commented on the surprisingly impressive organisation and rigour of the REF assessment process, as witnessed directly through their involvement in the assessment. They were very pleased to have the opportunity to participate in the assessment exercise. The following were the main points:

- Concerning the role of users in REF impact case studies, some users saw references to engagement with their organisation that they did not recognise and as a result consider there needs to be a requirement for a more fulsome explanation of the nature of user engagement in future case study templates.
- Commercial confidentiality was also a concern for some companies, which prevented evidence of impact from ongoing or recent projects to be submitted, and our interviewees suggested that this would continue to be the case going forward. This means a proportion of some of the most noteworthy impact case studies cannot be submitted, and that those research groups – even particular fields – will appear to have low levels of engagement. The Impact Template and Environment Statement perhaps provide a place within the submission where the scale and importance of this commercial work might be reported, using more aggregate statistics that are non-disclosive but nonetheless impressive and meaningful to panellists reviewing an institution’s overall achievements on research impact.
- According to intermediaries, collaboration with partners outside the HE sector was important but not particularly encouraged by the REF. From the point of view of industry, researchers in science and engineering were good at collaborating with industry. The main barrier was found at the institutional level, with university knowledge transfer offices imposing unrealistic contractual terms on collaborative agreements.
- Medical schools reportedly struggled with how to present themselves, because of the preponderance of work carried out in conjunction with hospitals. This was not easily credited and is causing a more inward focus. The definition of Category C staff is important to medical research and other areas of inter-sectoral collaboration but it does not solve the problem of how to support permanent wide-ranging collaboration between academics and practitioner researchers.

6.3 Open Access

Many respondents, mostly HEIs and other organisations, welcomed the new Open Access policy but also raised concerns about the costs associated with compliance in the current plans. Several respondents argued that costs could be mitigated if the reference data for Open Access requirements was the point of publication rather than at the point of acceptance:

We would note that HEFCE’s requirement that publications should be deposited within three months of acceptance to qualify as open access is already causing a major administrative burden in the sector, and is likely to cause more as we run up towards the next REF. One way in which HEFCE’s Open Access Policy could be simplified and the costs reduced would be to align it with that of RCUK, and be based on publication dates rather than acceptance dates.

(Professional body)

A few institutions argued that the policy was too costly with no significant benefit.

7 Influence on choices and gaming

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?

7.1 Influence on choices of individuals and institutions

Many HEIs described positive influences on institutional choices. For example, the REF had helped drive a focus on research and quality within institutions and the introduction of impact in REF2014 had driven a culture change within the communities and increased investment in impact from the institutions. Some also saw advantages in increased focus on fostering productive research environments and increased staff mobility.

Many responses, particularly from individual academics but also from HEIs and other organisations, brought up negative influences as well. Survey evidence⁵ referenced by one respondent suggested that a plurality of researchers saw the influence of the REF as negative. Many noted that the REF negatively influenced individuals' choices in terms of standardising their choice of topics, inducing short-termism, dis-incentivising collaboration, placing a time-burden on academics and often also impacting negatively on personal wellbeing, particularly where the REF is used directly as a source of threat around career prospects. One individual academic wrote:

I think the REF exerts a very negative effect on the work of individual researchers, and especially those who are early in their careers. Risky or experimental forms of scholarship are discouraged, as is writing for non-mainstream journals or other outlets, or in other (non-English) languages, or in unconventional forms. This results in a homogenisation and overall impoverishment of academic work.

(Individual Academic)

There was no clear consensus concerning the influence of the REF on early career researchers. Some noted that institutions had been incentivised to foster young talent but warned that this could be abused by institutions, employing them on short term contracts and so allowing them to submit their publications (see section on 'gaming' below) or even to require them to give their unpublished work to senior researchers who are returned to the REF. Others argued that the REF 'cycle' is particularly damaging to early career researchers, who might find themselves under pressure to publish unfinished work for short term REF requirements instead of developing a long-term research stream.

⁵ "The culture of scientific research 2014", online survey carried out for the Nuffield Council on Bioethics, available at: <http://nuffieldbioethics.org/project/research-culture/survey/>.

Teaching was often cited as an area that has suffered as a result of the REF, owing to the lack of rewards and the distinction between research active and non-research active staff.

7.2 Other drivers

Relatively few respondents commented on the issue of other drivers for behaviour besides the REF. Several mentioned other funding sources such as RCUK, BIS and the EU, but it is generally felt that the REF exerted a significant (positive or negative) influence on behaviour which could be difficult to isolate. For example, it was argued that it was “difficult to disaggregate the effects of the REF and other drivers in the system, since past and prospective REF performance has become embedded into a wide range of higher education processes...” (Learned Society).

7.3 ‘Gaming’ and solutions to gaming

7.3.1 *Types of gaming identified*

There was no apparent agreement among the respondents about what constituted ‘gaming’ in the context of the REF. Generally, it was used about strategic behaviour resulting in what was considered ‘artificial’ or negative outcomes. Some argued that HEIs making strategic choices to maximise institutional scores was distinct from ‘gaming’ and a potentially positive influence. Instead of trying to prevent strategic behaviour, it was argued, the REF should strive to provide the right incentives.

Respondents described four main kinds of behaviour as ‘gaming’:

- **Fractional or short-term contracts** centred around the census date was identified as gaming by many respondents. This practice allowed institutions to submit staff who did not have a genuine, long-term connection to the research at the submitting department. One variation of this type of gaming was the recruitment of established ‘superstars’ to work at 0.2 FTE contracts, including staff who were based at institutions abroad and played a very limited role in the UK institutions for which they were submitted. In another variation, institutions would hire staff, for example early career researchers, on short-term contracts immediately before the REF census date without the intention support their longer-term development.
- **‘Transfer-market’ or ‘poaching’** was also considered to be gaming by a large number of respondents. In the lead-up to the REF census date, an increased level of staff mobility was observed, as many institutions attempted to bring in new staff with portfolio of REFable papers. As a consequence, the hiring institutions could potentially return staff and outputs to the REF that had been developed at a different institution and did not reflect the research activity at the department or UoA during the assessed period. In turn, this could create a disincentive to invest in the development of researchers and drive up salary costs for hiring or retaining staff. While some argued that this allowed elite institutions to raid high quality departments at lower-ranking universities, others argue that increased staff mobility was good for the sector.
- **Highly selective submission of staff** was described by many respondents as a type of gaming aimed to achieve what was seen as artificially high Grade Point Average (GPA)

scores and rankings.⁶ Thus, some institutions engaged in a strategic game balancing staff volume and associated QR funding against a high GPA ranking with the related reputational advantages. Not all respondents considered this to be ‘gaming’, but most agreed that it could have negative consequences such as distorted league tables and exclusion of good research staff who don’t fit the HEI selection strategy.

- **REF impact case study thresholds** also gave rise to institutional behaviour described by some as gaming. This type of behaviour, referred to as ‘cliff-hanging’ by one respondent, consisted in restricting the number of staff submitted to exactly one FTE below the threshold mandating an additional impact case study. This was seen by various organisations as an arbitrary criterion for selecting and excluding staff, with negative consequences for individuals excluded.

The discussion about metrics also prompted many respondents also warn against the risk of increased gaming if metrics were adopted more widely.

7.3.2 *Suggested measures to restrict gaming*

Respondents proposed a range of solutions or counter-measures. The most common were:

- **Limiting ‘portability’ of research outputs** was seen by many HEIs as way to curb the practice of ‘poaching’. The suggestion was to restrict the ability of researchers to bring previously published REFable research outputs with them when they change institution (‘portability’). Different variations were suggested. Many suggested simply attributing outputs to the institution where a researcher was employed at the date of acceptance or publication. Another proposal was to proportionally reward an academic’s current and previous institutions according to the length of employment over the REF assessment period. Finally, it was suggested to allow both current and former institutions to submit the same outputs. Several respondents emphasised the positive aspects of staff mobility but stressed the need to protect institutions that invest in the development of researchers. It is also noted that the rules for research impact already work in a similar fashion.
- **Changing eligibility of staff** was suggested by HEIs and other organisations as a means to counteract ‘poaching’ and the use fractional contracts. A number of suggestions were made about how to ensure that staff members returned to the REF play a genuine role at the institution. Some suggestions concerned the ‘length of service’, for example requiring that the member of staff joined the institution at a cut-off date prior to the REF census date and/or that their contract extended for a defined period of time afterwards. It was also suggested to raise the minimum requirement for fractional contracts from 0.2 FTE to 0.5 FTE. An alternative suggestion was requiring institutions to report on staff turnover as a way to discourage excessive staff movements. Among other things, it was argued that such changes would help increase institutional stability.
- The suggestion to **eliminate staff selection and require institutions submit all eligible staff** was discussed above with respect to the accuracy and efficiency of the

⁶ GPA scores are calculated by multiplying the percentage of research in each grade by its rating (0-4), adding them all together and dividing by 100 (see: <http://ref2014.leeds.ac.uk/definitions/>). GPA scores do not take into account the volume or relative proportion eligible staff submitted. GPA scores are not provided as such by the funding councils but are commonly used as the basis for rankings as a simple measure of the average quality of research in a unit or institution. See for example: <https://www.timeshighereducation.com/features/ref-2014-results-by-subject/2017594.article>.

assessment but was also raised by a large number of respondents in relation to gaming. Several institutions in favour of eliminating staff selection suggested a compromise whereby a minimum threshold would be introduced, requiring institutions to submit at least 75 to 90% (suggestions varied).

- A number of institutions argued against the elimination of staff selection because, instead of solving problem of gaming, it would simply move it into a new arena. Specifically, concerns were expressed that it would lead institutions to perform staff selection through other means, redrawing staff contracts to put staff deemed unsuited for submission to the REF on teaching-only contracts. Some supporters of complete staff submissions argued that such practice would likely be prevented by legal and human resource management standards. Others argued that 'all staff' should include teaching staff as well as research staff, and that REF and the future TEF should be coordinated to prevent gaming based on staff categories.
- As an alternative to requiring all staff to be submitted, some respondents suggested introducing a requirement in the environment template to submit figures describing the proportion of staff submitted. GPA scores should be prevented from forming the basis for league tables and only be published in conjunction with weighted data. It was argued that this would provide a more accurate picture of the research in the submitting departments and dissuade highly selective staff returns aimed at maximising GPA scores alone.

Finally, some respondents suggested that greater clarity and continuity of rules would help reduce the amount of gaming.

8 Influence on disciplines and other areas of scholarly activity

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?

Responses to this question reviewed many of the issues covered in previous questions, including issues surrounding the structure of review panels (see section 3.1), the issues of collaboration and interdisciplinary research (section o) and the effect of the REF on researchers (Section 7).

Some of the main issues reported to influence disciplines were:

- As discussed above, a large number of respondents report that REF tends to incentivise ‘safe’ choices, and to have a ‘normalising’ or ‘narrowing’ effect on research.
- Interdisciplinary research, smaller disciplines as well as new, emerging fields often did not fit comfortably within the REF UoA structure and can consequently be overlooked. One suggestion was to address emerging fields in the environment template. Interdisciplinary research was discussed in section o above.
- Activities underpinning the development and sustenance of disciplines were said to go unrewarded. This could be work for Societies and professional bodies or journal editorships. Many forms of outputs were ineligible for the REF such as textbooks, and respondents still reported that monographs – which are important in some disciplines – were poorly received in the REF.
- The effect of the REF on teaching was a recurring issue. As REF did not recognise teaching, there was a tendency to prioritise research over teaching, or to separate the two from each other. Several respondents looked to the future TEF to rebalance the relationship between teaching and research.

Among the positive influences, some institutions believed that the REF has driven quality and capacity in their disciplines and has provided support for Open Access and early career researchers.

9 Future plans

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

9.1 Assessment of future plans in the REF

Many HEIs and other organisations argued against an enhanced assessment of future plans as part of the REF. Two main arguments were advanced in support of this position. First, it was argued by a large number of respondents that the role QR funding was to fund research on the basis of a retrospective assessment of research results. It was the role of the other part of the dual support system, the research councils, to provide prospective grant funding.

Secondly, it was noted that an assessment of future plans would lack robustness and rely too heavily on unverifiable ‘creative writing’. To the extent that future plans were to become a more important part of the REF assessment, several respondents advised that it should be based on auditable criteria, and that strategies put forward for one assessment cycle, should be audited as part of the following review. Reference was also made to Scotland, where forward-looking outcome agreements already exist. According to one typical response:

REF is now sufficiently mature that, in the vast majority of cases, the past is a good predictor of the future. Research councils already make awards based largely on forward-looking measures. For this reason, it is a very good thing that REF focuses on retrospective analysis to maintain a balance. The retrospective nature of the REF provides an important element of stability and in part guarantees a level of effective infrastructure essential to sustainability. QR enables institutions to respond flexibly to pressures over time and to support innovation; it should not become part of a national strategic planning exercise.

(Higher Education Institution)

Some did, however, also welcome a stronger forward-looking component in the REF. Proponents of the idea included some HEIs and a few individuals and other organisations. It was noted that the environment statement – potentially merged with the impact template – would be the natural place for an enhanced forward-looking element. It was seen as a positive opportunity for the institutions to reflect on and be held accountable for future planning. A forward-looking element would also make it easier for new institutions without an established track record to gain recognition and develop new capabilities.

9.2 The use of QR funding by institutions

Some respondents additionally discussed the implications of an increased focus on future plans for the autonomy of HEIs and their control over the internal allocation of QR funding. Most respondents commenting on this issue, including HEIs and other organisations, expressed support for retaining un-hypothecated QR funding. It was argued that institutions

needed to be able to respond flexibly to unforeseen opportunities and emerging areas of interest without the constraints of a pre-approved strategy. Concerns were also raised that revealing future plans could make HEIs less competitive.

A few responses from individual researchers and other organisations, suggested that QR funding should be more directly attributed to the departments and individuals whose research had been assessed successfully.

Dynamism – views from stakeholder interviews

The interviewees largely considered that the RAE and the REF had helped to ensure that the UK research base was broad-based, inclusive and dynamic.

- The majority of contributors with knowledge of the UK research base in the 1980s, were adamant these national assessment exercises had caused a step change in university attitudes towards research and that the whole system was stronger, more serious and more competitive globally.
- HEIs saw no reason for REF to be more forward-looking, and that, since the research councils were playing this role, there was no reason for the REF to go there.
- HEIs and intermediaries urged government not to relinquish REF's commitment to institutional autonomy and thematic flexibility. Attempts to 'micro-manage' the system should be avoided as they could easily go wrong.
- Research funders and research users took a slightly different position, and expressed interest in seeing REF tackle the question of dynamism (of individual HEIs) more explicitly. There was no suggestion that REF should start to assess future potential in the way one might do in considering the case for a large capital investment (an ex ante impact assessment) but rather require HEIs to explain and exemplify their strategic capacity.

10 Other comments

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

The last part of the consultation provided an opportunity to comment more generally about the REF.

10.1 The purpose of the REF

A large number of respondents commented on the purpose of the REF, most of them insisting that it serves a range of purposes beyond the allocation of QR funding. In addition to accountability and benchmarking, many respondents cite the aim to ‘fund excellence wherever it is found’ in relation to many of the issues brought up across the consultation questions.

10.2 Burden of the REF

Many respondents, including all groups of respondents, commented on the cost or ‘burden’ of the REF. Among other things, the REF was felt to be time-consuming for individual researchers and increasingly complex and costly for institutions. While most respondents agreed that the REF was costly and could be improved, many also argued that it is relatively good value for money compared to perceived alternatives, such as the grant funding through the research councils. Among suggestions made to reduce the burden (discussed throughout the text above) were ideas to extent the gap between assessments, align data and processes with other research funders to avoid duplication and reduce game-playing.

Simplification and streamlining – views from stakeholder interviews

The stakeholder interviews revealed many views on cost savings:

- Continuity was seen as one of the best ways by which to contain costs so that the main savings for REF2021 would come through minimising changes. This includes impact case studies, which were arguable ‘gold-plated’ due to the novelty of the process in 2014 (HEIs and intermediaries).
- Some felt that a system designed to work for all HEIs – including different sizes, specialisms and research intensity – added cost and complexity. Other HEIs argued that this supported the diversity of the sector and allowed less research-intensive institutions to retain a research presence. There were concerns that changes to the current system would favour large universities at the expense of less research-intensive institutions.
- Inclusion of all staff would reduce costs associated with selecting staff, special circumstances and early career researchers (HEIs).
- The number of impact case studies could be reduced and the Impact Template could be eliminated (HEIs).
- Producing the main Environment Statement at the institutional level, with lighter information for UOAs, would save effort (HEIs and intermediaries).
- REF should return to basics, refocusing on outputs and research quality. Impact could be omitted and the environment statements simplified (HEIs and individuals).
- There is no need to pay panel assessors, they all have jobs and will do the work happily as part of their extra-curricular activity (other countries only pay T&S) (HEIs).
- Further harmonisation of data collection could provide savings, including urgently getting to a common understanding between RCs and FCs on new issues such as Open Access.

Overall view on efficiency – views from stakeholder interviews

Interviewees gave contrasting views about the cost of REF, with organisations close to the process judging the costs to be reasonable where those at a greater remove see the final cost as being too large and in need of reducing:

- HEIs and intermediaries saw the REF as good value for money in light of the size and diversity of the system it works within, and that its seriousness (and cost) had been central to the high level of trust it enjoyed across the higher education community.
- Several HEIs argued that the estimate costs (REF 2014 Accountability Review) were overstated and reflected the community’s desire to “charge as much of their research management effort to REF as possible.” They argued that many of the costs were ‘business as usual,’ and that a good deal of the additional costs were self-inflicted by institutions investing heavily to ensure the get the very best scores they possibly could.
- There were numerous other suggestions for efficiency savings, which have been discussed already above, ranging from news rules on selectivity to a lower number of impact case studies and a re-basing of the environment statement at an institutional level.

10.3 Teaching

The issue of teaching and the relationship between teaching and research was brought up repeatedly in different contexts. Some respondents argued that the REF had led to a prioritisation of research over teaching at many institutions, and many proposed that ‘impact on teaching’ should be rewarded in the REF.

Many respondents expressed concerns about the increasing separation of teaching from research and the role of the REF in this process. The expected introduction of a new Teaching Excellence Framework (TEF) was seen as a risk of further competition between the two missions, especially if the responsibility for the two funding streams falls to different funding bodies. Others saw it as an opportunity to rebalance a relationship seen to be skewed in favour of research.

10.4 Equality and diversity

A number of respondents, mostly HEIs and other organisations, commented on the implications for Equality and Diversity of various elements of the assessment system. Contributors complimented Hefce on the new measures introduced with REF 2014 designed to improve the overall assessment process’s ability to take into account issues to do with equality and diversity. There was a general sense that the HE community needs to do far more in this space, and that REF has made a positive contribution and can / should do more going forward. That said, while people welcomed the introduction of the ‘individual staff circumstances’ process in principle, it was widely considered to have been unnecessarily burdensome and intrusive.

Several contributors commented that while the process was onerous in its scope, the thoroughness had brought new insights about Equality and Diversity issues within research and had prompted wider reviews / reform of institution’s own policies and practice.

There was support for requiring a fuller exposition of Equality and Diversity policies and practice within (institutional) environment statements, with common metrics and information about certification.

There were also calls for panel members to be briefed / inducted in the potential for unwitting bias, and the need to carefully consider Equality and Diversity aspects in order to ensure a neutral assessment. There were risks flagged about gender bias in bibliometrics.

10.5 Continuity and transparency

Many HEIs and other organisations commented at different points across the nine questions, that continuity and transparency would be a significant benefit, both in term of efficiency and embedding of new practices. It was also felt by some that, while more fundamental design-changes could be justified in the long run, preparations for the next REF were already too far advanced to make major changes. For example:

We welcome this review of the REF as a unique opportunity to outline ambitious suggestions for ‘direction of travel’ improvements to the exercise, not necessarily for immediate full implementation. We would stress that the magnitude and timescale of the review’s recommendations

must reflect the fact that exercise's cost and burden scale i) with the degree of change from the previous exercise and ii) inversely to the amount of preparation time available.

(Higher Education Institution)

Some respondents also urged for guidelines for the upcoming REF to be released as soon as possible to allow ample time to prepare.

Alternatives to REF – views from stakeholder interviews

Interviewees prefer to fine tune the existing system rather than change it radically.

- **Using research council income as a basis for allocating QR funding:** While this looks okay at an overall level, analysis suggests it fails at the field and institutional levels with a heavy bias in favour of STEM subjects and the top quartile of HEIs.
- **Replacing REF with a negotiated block grant:** While it is initially attractive to think that HEIs might access QR funds without the work involved in REF, interviewees expressed uneasiness about the criteria and process one would use to determine the distribution and the capacity of government to run the process (rather than outsource it). Instead, it was felt that the UK needed a process like REF that is transparent and consistent, and which the community largely drives itself.
- **Reducing the share of QR funding contested in each cycle:** The larger HEIs found this idea interesting, but reducing the refresh rate for funding is potentially 'sclerotic', which is the opposite of what is needed. According to intermediary organisations, smaller HEIs would feel especially at risk from such an arrangement. The smaller HEIs value the accreditation and reputational benefits at least as much as the QR funding secured, and want the opportunity to invest and make major gains in the quality and volume of their research as they move from one REF to the next.

11 Summary of key issues

The call for evidence issued to support the review of the Research Excellence Framework (REF), chaired by Lord Nicholas Stern was issued to explore a range of issues arising from discussions in the early stages of the review.

This report contains a synthesis of 301 responses to the consultation received by the Department for Business Innovation and Skills (BIS). As shown in the discussion above, the responses raised a wide range of issues, only some of which have been presented here. The questions were generally open and invited a very rich debate about the research assessment system. On some issues, the responses coalesced around concrete options concerning the design of the future assessment framework. On a number of key policy options and issues where a significant number of respondents took a clear position, indications of the weight of opinion have collected.

Table 2 below provides an indication of the level of support of different groups of respondents to these issues as found in the responses to the Call for Evidence.

Table 2: Level of support for policy options and positions, by respondent group

Section	Option	Higher Education Institutions	Individuals	Other respondents*
(2.1.1) Number of outputs	Reduce number	<i>Moderate support</i>	<i>Limited support</i>	<i>Limited support</i>
	Keep 4 outputs per person	<i>Limited support</i>	<i>No support</i>	<i>Limited support</i>
(o) Staff selection:	Submit all staff	<i>Moderate support</i>	<i>Moderate support</i>	<i>Moderate support</i>
	Retain staff selection	<i>Moderate support</i>	<i>No support</i>	<i>Limited support</i>
(2.2) Impact	Broaden definition of impact	<i>Moderate support</i>	<i>Moderate support</i>	<i>Moderate support</i>
(2.3) Metrics	Extensive use of metrics	<i>Limited support</i>	<i>Limited support</i>	<i>Limited support</i>
	Metrics as support for peer review	<i>Very strong support</i>	<i>Limited support</i>	<i>Strong support</i>
	No role for metrics	<i>Limited support</i>	<i>Limited support</i>	<i>Limited support</i>
(3.1) Units of Assessment (UoAs)	Keep current UoAs	<i>Very strong support</i>	<i>Moderate support</i>	<i>Strong support</i>
(3.2) Link between outputs and individuals	Keep current link	<i>Moderate support</i>	<i>Limited support</i>	<i>Limited support</i>
	De-couple individuals from outputs	<i>Strong support</i>	<i>Limited support</i>	<i>Moderate support</i>

Section	Option	Higher Education Institutions	Individuals	Other respondents*
(3.3) Aggregation	Aggregate parts of environment and/or impact elements	<i>Strong support</i>	<i>Limited support</i>	<i>Moderate support</i>
(7.1) Influence of the REF	REF has negative influences	<i>Strong support</i>	<i>Very Strong support</i>	<i>Strong support</i>
	REF has positive influences	<i>Strong support</i>	<i>Limited support</i>	<i>Moderate support</i>
(7.3) Solutions to gaming	Restrict portability of outputs	<i>Moderate support</i>	<i>No support</i>	<i>Limited support</i>
	Change eligibility of staff	<i>Strong support</i>	<i>Limited support</i>	<i>Moderate support</i>
	Selectivity (<i>see above</i>)			
	Reduce use of unweighted scores	<i>Limited support</i>	<i>Limited support</i>	<i>Limited support</i>
(9.1) Forward planning as assessment criterion in REF	In favour of enhanced role of future planning in REF	<i>Moderate support</i>	<i>Limited support</i>	<i>Limited support</i>
	Against enhanced role of future planning	<i>Moderate support</i>	<i>Limited support</i>	<i>Moderate support</i>
(9.2) Allocation of QR funding	Keep un-hypothecated	<i>Strong support</i>	<i>Limited support</i>	<i>Moderate support</i>
	Earmark QR funding	<i>No support</i>	<i>Limited support</i>	<i>Limited support</i>

(*) The group “Other organisations” includes all categories of respondents other than HEIs and individuals



© Crown copyright 2016

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk. Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication available from www.gov.uk/beis

Contact us if you have any enquiries about this publication, including requests for alternative formats, at:

Department for Business, Energy and Industrial Strategy
1 Victoria Street
London SW1H 0ET
Tel: 020 7215 5000

Email: enquiries@beis.gsi.gov.uk

IND/16/1a