

# The Challenges of Research Assessment:

## A qualitative analysis of REF2014

Daniel Neyland

Sveta Milyaeva



European Research Council  
Established by the European Commission

**Goldsmiths**  
UNIVERSITY OF LONDON

January, 2017 ERC Project: MISTS 313173 [www.marketproblems.com](http://www.marketproblems.com)

## Executive Summary

This report provides an in-depth qualitative analysis of the experiences of REF 2014 participants. It draws on long, semi-structured interviews with 33 participants including REF managers, Main Panel members, sub-panel members from a cross-section of distinct disciplines, and impact assessors. The corpus of over 1000 pages of interview transcripts was arranged into five main sections, summarised here and reported in more detail in the following pages.

### A. Panellists' motivation:

Interviewees discussed their motivations for taking part in the REF. These included:

- a sense of civic responsibility
- a need to represent their academic field
- to oversee their department's interests and secure income for their department
- a desire to change the REF for the better
- to carry their misgivings into the process in order to make changes

Impact assessors also reflected similar motivations:

- a form of civic duty
- a hope to make change by bringing impact to public attention
- a sense of disquiet about the REF and its ability to assess such matters as impact

### B. Purpose of the REF:

Interviewees articulated a variety of purposes for the REF. These included:

- the selective allocation of research funding
- stimulating competition
- prestige and reputation
- valuing UK research as a whole
- enhancing accountability
- enabling community building
- increasing stress and pressure
- streamlining funding
- marginalising certain institutions

### C. The work of making the panels:

Interviewees reflected on the composition of sub-panels, reflecting on:

- how sub-panels might incorporate relevant forms of expertise
- the range of work that a sub-panel would need to assess
- how a sub-panel might look to its Main Panel and to audiences outside the REF

- how a discipline could have adequate representation
- who might fight on a discipline's behalf to defend its interests
- the adequacy of representation in terms of gender, geography, age, departments and university types (old, new, small and large)

#### D. Assessing academia:

Interviewees discussed:

- the challenges of peer review, the workload and delivering scores
- disputed scores
- spurious accuracy
- structural limitations of the scoring system
- calibration, normalisation, the relative ranking of outputs and their effects
- the strengths and limitations of metrics
- the production, assessment and inequities of impact

#### E. Effects of the REF:

The following effects of the REF were discussed by Interviewees:

- gaming through recruitment strategies, fractional contracts, research stars, the self-promotion of individual academics
- marketisation and the competition and constraints it might impose on Higher Education
- increased professionalism of institutions and accountability
- disillusionment with research assessment
- shaping of career choices
- the ways the REF has brought value to UK academic research

## Introduction

In this report we explore, in-depth, the challenges of imposing and carrying out a national system of academic research assessment. We focus on the UK as one of the most developed and long-standing forms of research assessment, from the Research Assessment Exercise (RAE) that began in the 1980s to the Research Excellence Framework (REF) that took place in 2014.

Research assessment in the UK over this period of time has attracted some attention and critique. For example, it has been argued that research assessment measures become targets that problematically shape the trajectory of academic work (Strathern, 2002), that research assessment enables the marketization of UK Higher Education (Brown, 2014) or that research assessment is costly and unreliable (Sayer, 2014).<sup>1</sup>

Rather than take these critiques for granted, we carried out research into the practices of the 2014 REF. Although it would be possible to analyse the metric outputs of the RAE and REF (as many universities do in planning their REF submissions and reflecting on their results), we felt that a greater insight into the practices of the REF would help bring to light in more detail how research assessment works. That is, we became particularly interested in how, for example, the REF structure of Main Panels and sub-panels works, how the mechanisms of peer review – so central to the REF – operate in practice, and how those who took part in the REF reflected on their participation. An overview of the practices of the REF, how these differ, a detailed comparison of the issues that arise in research assessment, and how these are resolved, has until now not been the subject of research. We decided to address this gap by carrying out qualitative interviews with REF participants, including Main Panel and sub-panel members, REF managers, and impact assessors. In the following pages we report on the results in five main sections. First we offer a background to the RAE and REF and further detail on our methodology.

This research was generously funded by the European Research Council (grant no: 313173; [www.marketproblems.com](http://www.marketproblems.com)) as part of a broader enquiry into the effects of interventions into public sector problems that are said to contain a market component.

## **Background to the Research Assessment Exercise (RAE) and Research Excellence Framework (REF)**

The UK Research Assessment Exercise (RAE) took place in 1986, 1989, 1992, 1996, 2001 and 2008. Subsequently it was replaced by the Research Excellence Framework (REF), with the first REF assessment taking place in 2014. University departments in the UK submit a REF return, traditionally composed of the most suitable research active members of the department based on internal assessments carried out by each University (although this may change for the next REF to universal submission). Each UK academic selected to take part should submit their four 'best' research outputs produced in the REF period (January 2008 to December 2013), although books can be double-weighted and part-time and early career staff can submit fewer outputs. When Universities carry out their internal assessments of what, and who, should be submitted, practices vary by subject, with Business and Management Schools mostly using their own ABS journal ranking to assess the strength of publications for potential submission and disciplines such as Sociology avoiding the compilation of a journal ranking. Whichever approach is taken, all articles are peer reviewed and scored on a four/five part scoring system (from 0 to 4, with 4 being highest) and the peer review should constitute the basis for scoring, not the form of publication (in principle, journals should be considered equal for example. This might also change for the next REF, as in 2014 panels could experiment, for example, with citation indices and in future other metrics might be permitted).

Peer review takes place through 4 Main Panels (Panels A, B, C and D), each of which covers a broad academic area (natural sciences, social sciences, etc) and is divided into 36 discipline specific sub-panels, also known as Units of Assessment (such as Sociology or Clinical Medicine). The chairs of each sub-panel make up the membership of the Main Panels, also under the stewardship of a further chair. Chairs allocate outputs (articles, books) to members of their sub-panel in line with their expertise to carry out peer review. On most sub-panels, two reviewers examine each output, although on some panels this is not the case. Departments can also highlight staff circumstances that panels should take into account in considering an academic with fewer publications (including maternity leave, sickness and so on). Further direction can be given by departments on such matters as inter-disciplinary work that might need the attention of assessors from other sub-panels.

Some individual assessors on some sub-panels were given between 800 and 1000 articles to peer review. If there are identified gaps in the expertise of the members of the panel, then a further specialist can be nominated and appointed to the sub-panel. Each academic then enters a score for each article they review in the computer system designed for the REF by the Funding Councils (the bodies responsible for managing the assessment system). These scores are then combined to work out the average score for each department that has submitted to a sub-panel and a ranking is calculated.

Along with research outputs (articles, books), departments should also submit an environment statement. This should provide a narrative of research life in each department submitted, setting out how much research grant income has been received, and how many PhD students are in the department, the major research centres and their activities, and so on. With the inception of the REF, higher education institutions should also submit a template that discusses the institution's approach to facilitating the impact of their research, and

departments should also submit impact cases. Impact cases should demonstrate the non-academic impact of the work of the department. Impact cases had to fit a 4 page template and set out the impact achieved, evidence for that impact and a link to an academic output that would score at least 2 stars on the REF scoring system. Although for assessment of research outputs, items had to be published within a single assessment period (for the last REF this was 2008 to 2013), impact case studies could draw on a 20 year history. Under REF 2014, outputs amounted to 65% of a department's score, with 20% derived from impact and 15% from environment.

The number of impact cases required by a department, depended on the number of staff submitted to the assessment (2 cases for up to 14.99 FTE staff, then 1 more for every 10 extra staff). Yet, the distribution of up to £1.6billion of Quality-Related (QR) research funding annually to English Universities alone, also depended on a calculation of the percentage of staff submitted. Although, then it might be prudent to submit fewer staff to reduce the number of impact cases required, this might then impact on a department's research power ranking (their score in relation to the percentage of full-time equivalent staff submitted). Some departments made smaller submissions to the REF on the basis that they wanted to get a high ranking for their outputs and reduce their impact obligations, aware that this would reduce their allocation of government money. These departments calculated that the amount they would lose in government QR funding was less than they would lose from other external sources if their reputation was negatively influenced by a low assessment score. Other departments sought to submit as many staff as possible in a strategy to try and boost their research power ranking and their QR funding. Well known, well-funded and established departments at some Universities expected to both submit most of their staff and score highly on their outputs.

This strategic decision making offers an initial entry into the complex decision making that pervades the valuation of UK academia. It points toward a selective assessment of what kind of result might prove favourable, who (and what kind of work) might contribute to such a result, and how this might be used to later narrate the success and strength of an academic department or even a University. Over 30 years these valuations have become a commonplace feature of UK Higher Education.

## Methodology

In this report we look up-close at the decision making that characterises the REF and the complex features of the organisation of the assessment system. As a research task, this is by no means straightforward. Beyond anodyne minutes of sub-panel and Main Panel meetings that offer few insights into the details of the valuation system, the REF and the RAE are characterised by a curious absence of accountability and transparency (particularly in comparison with other publicly funded initiatives which depend on openness for their democratic legitimacy). REF sub-panel and Main Panel members must destroy all records, including notes relating to any assessments they have carried out, messages they have exchanged regarding valuations, concerns or questions they have raised or any accounts of disputes that were resolved. No-one is allowed to record the means by which a valuation took place. Only the score is recorded. The REF is more black hole than black box awaiting to be opened. However, with agreement from the Higher Education Funding Council for England (HEFCE) and with a commitment to complete anonymization<sup>ii</sup>, we have carried out lengthy interviews with Main Panel and sub-panel members, specially invited academic assessors and impact assessors. This has produced a corpus of over 1000 pages of interview transcript.

We have arranged the research data into five main areas, some with sub-fields. This resulted from a detailed analysis of the interview transcripts and the development and refinement of a coding that allowed us to gather together common themes in interviewee discussion without overly restricting the range of data to be presented. To be clear, we did not code the interview transcripts in a way that sought to homogenise the responses we received. Instead, we coded according to shared topics of discussion. Where interviewees expressed distinct views on a particular theme, we have set those out in the following analysis. Our five themes are listed below and form the structure for the following sections of this report:

- A. Panellists' motivation
- B. Purpose of the REF
- C. The work of making the panels
- D. Assessing academia:
  - peer review,
  - scoring/calibration,
  - metrics,
  - impact
- E. Effects of the REF:
  - gaming,
  - marketization,
  - careers,
  - transparency, accountability and justice,
  - funding,
  - valuing research

## A. PANELLISTS' MOTIVATION

In this first section, we look at what motivated participants to take part in the REF. Initially we designed this question as a gentle opening to interviews, encouraging interviewees to begin to talk about their engagement with the process. However, as we reviewed the corpus of the interviews as a whole we realised that many of the themes that emerged from interviewees in responding to this question, were revealing of participants' views of the process as a whole and what ought to be their role.

Firstly, several participants suggested they took part in the REF to either represent or even secure their academic department's interest:<sup>iii</sup> As Interviewee 3 notes, being represented on a sub-panel is key to ensuring fair representation:

Interviewee 3.

I've actually been a member of RAE 2008 and REF. [...] Why did I agree? At the time of that REF nomination I was in a senior admin position at the university. [...] And I felt it was very important to be involved in the process, because actually... slight conflict of interests - I was also steering [my University's] submission to the REF. And so I think in a way the REF is a little bit of a tension between senior people from universities who've done their own submissions and they're then in a way negotiating how the outcome should come, you know? So I felt it was important that [University X] had a senior member of staff on this panel. [...] I think it's important to be represented. I think your University... I see it as exactly as the tension between people submitting on behalf of University and people in a big circle judging on behalf of all Universities. And if you are not represented I'm not sure that you would be treated as fairly as if you are represented [laughing]. I think it's important to be at the table, basically. So that's why I did it.

This interviewee captures the professional tensions of academics participating in their own valuation. The interviewee notes this is a 'slight conflict of interests' and a 'tension' but at the same time participation was vital to ensure being treated 'fairly'. Other interviewees developed similar themes, discussing how participation was important for their institution, and also their career (Interviewee 9 and 20):

Interviewee 9.

I said yes, because I thought it would be interesting to see, and quite useful for my institution to see the process from the inside. It also, yes, seemed like it would be an interesting way of seeing the process. It was quite a good thing to do at that point in my career, as well.

Interviewee 20.

I really wanted to understand, what are the factors in the end, to help me but to help my institution as well? Understand the internal workings of the REF? There's always so much controversy about it and whether it's really transparent or really fair and so on. [...] I think it's advantageous to you, as an individual, and also advantageous to your institution, to have people who have been part of the process. I'm not saying that you're feeding back to them at the time. But afterwards, for example, my vice chancellor asked all of the people who were on panels to debrief him. So that the institution gained some insight from the experience and whether it had been a good experience. Actually, we had a group of us together doing that, as well as individually. We did find there were different experiences.

What we can take from this preceding excerpt is that the importance of being on a REF panel for an institution is not limited to that single assessment, it also helps institutions to plan for subsequent assessments.

Other interviewees expressed different motivations for participation. For some, it was important for academia. In the following excerpts Interviewees 26, 15 and 12 discuss the importance of representing their field and the responsibility this entails:

Interviewee 26.

I just think it's part of my social responsibility as an academic to contribute to these things.

Interviewee 15.

I could have said no. Partly I suppose I was flattered to be considered. The people who were encouraging me to apply felt that I would be able to bridge the gap between rather different approaches to the nature of the subject, the nature of the discipline, the forms of scholarship, and I was seen as covering both bases. [...] I kind of saw it partly as an obligation to my fellow academics, that it's a form of service that you do if called upon to do it. I also thought it would be interesting to do and maybe I was looking for a challenge. So that's what encouraged me to do it.

Interviewee 12.

I suppose, you know, it's partly that you feel it's important for the community of people you're representing. [...] I suppose, I felt it was appropriate for them to be represented, and if I said no then that might not occur.

For Interviewee 14, it was not just the academic community that required support, but also the REF of which they were a strong advocate. Being involved was a way to maintain the positive benefits of the REF:

Interviewee 14.

Why did I accept to do it? Because I think it's a great service to the profession. I think we're very fortunate in the UK to have this mechanism, I think it does miracles to improve research productivity really, so it can only work if people are willing to do it, and that's why I did it.

Other Interviewees (18 and 11) sought to take part in the REF as a means to represent the field, but also to make a difference to the REF and the assessment process either through ensuring diversity (Interviewee 18) or a pluralist assessment (Interviewee 11):

Interviewee 18.

I was really interested in it, and I was really interested in the processes. I think people complain a lot about REF, an awful lot, but I think it's always better to change things from within. I've always felt very strongly that if you don't agree with something, then work to make it better. I felt very strongly about women's positions in higher education, and the ways in which I had been subject to maternity leave and was still expected to produce four outputs, in the past. It worried me that women were being – not just women, but people with severe caring responsibilities, or other complex circumstances were effectively being penalised. We believe in equality and diversity, so how do we achieve equality and diversity for REF? How do we put mechanisms in place that facilitate equality and diversity? I felt that there were ways in which I could play a role in that through REF, and create a more equitable REF process.

[...] I have a strong belief that research in [Discipline X] is important and contributes to the [...] economy. I have a strong belief in the fact that we create impact in our areas, and that often this is not recognised, or is not visible. How can we make visible the fact that our research does make a difference? And how can better conditions be created to enable the widest number of researchers to go forward, and to not marginalise those who have had maternity leave, are caring, who have complex circumstances in their lives? So how can we make this a process that encourages diversity and equality on all levels. So I think it was a mixture of all of those things.

Interviewee 11.

I had not long finished being Editor of the Journal of [X]. So I suppose thinking about what I thought the panels might need, then I thought someone who is quite general might be useful. Because, just from being an Editor of a very general journal, I was

aware that I was quite reasonably placed, relative to those people who are very narrowly specialised, so just thinking about getting things reviewed.

[...] I remember some people at the time, when it was announced, said, “Won’t that be a huge amount of work?” And I remember saying, “Well I kind of take the view that somebody has to do it.” As well as being a generalist I also think I am quite a pluralist, and in our field I think that’s quite important [...]. So when people say, “Why are you doing that? That will be a huge amount of work.” I think I said, half jokingly, “Well if I don’t do it someone else will be recruited and maybe that other person would be a bit less open to judging things on their own terms.” [...]

For the following interviewees (22 and 30), making a contribution to the academic community went hand in hand with having misgivings about the REF, either in terms of the amount of work it would require (Interviewee 22), or the limitations of the assessment system (Interviewee 30):

#### Interviewee 22.

Well, partly, it’s a service to the academic community. So, you feel- so, when you’re rung up- so, this time, I wasn’t so keen. But I was rung up by a persuasive chair, and after I thought about it, I didn’t mind doing it. You know, it’s quite a lot of work, but it’s quite interesting, and you feel you’re doing some service. So, on balance, I was happy to do it. I definitely don’t do it for the money (Laughter). That’s clear. It’s more- there is some interest in it, and the work itself is fairly interesting. You’re reading papers that you wouldn’t otherwise read. Of course, you read them fairly quickly, but it’s interesting to get a view of other work outside your own narrow area, so that’s a plus. [...] The other plus is that you feel you’re doing a job for the community. I think those two things would be the main motivation.

#### Interviewee 30.

I could’ve said no. [...] Well, to be honest, I had some misgivings about getting involved, because I feel that even well conducted research assessment exercises, they can fail to pick up some qualities of universities and university environments, which really should be cherished. But given that the assessment was going to take place, I had a sense of duty to participate, and make it as effective as possible, and try to ensure that full credit was given to leading researchers.

I suspected that if there were a substantial proportion of panellists that were not experts in the fields of assessment, you know, they might give overdue assessment, or overdue attention to, maybe, unreliable impact factors, the way academics publicise their work in some cases, when their claims are not substantiated.

For Interviewee 21, these misgivings about the REF and in particular a scepticism about impact, provided their main motivation for participation. They sought to have impact-scepticism represented within their sub-panel:

Interviewee 21.

Why was I prepared to put myself forwards? The choice of [sub-]panel chairs coincided more or less with the announcement that impact was something that was going to be assessed [...] This caused a fair degree of anxiety to the [Discipline X] community. And as a sceptic about impact I thought it was tremendously important that, among those that were going to be operating this, were some people who were robustly sceptical about it. And I thought I could probably manage to maintain my robust scepticism, as indeed I have. And so that, I guess, was quite a determining factor. And across the [...] [main] panel [...] the vast majority of sub-panel chairs at least started out as sceptics. Some of them became converts in the same sort of process, but not all of us.

For the following specialist assessor invited to complete the range of disciplinary expertise covered by a sub-panel, motivation also derived from a form of scepticism about the assessment process and a desire to work on it from the 'inside':

Interviewee 28.

I understand that they felt they did not have enough capacity in this particular area in terms of the actual members who had been appointed to the REF panel. In some way unknown to me, they chose me as a potential expert in that area, and they wrote to me and they asked me. After some discussion with colleagues here, I agreed. [...] I felt that it was an important exercise and it was going to have significant consequences. Although I have had my criticisms of recent developments in the REF process, although I was a supporter of the early years of the RAE, I felt it was probably better to work on the inside than on the outside to understand the process a little bit better. I'm now a senior professor in the discipline, and I felt a little bit of a professional obligation to be involved.

For impact assessors, their motivations for taking part on panels was slightly different. Alongside wanting to support the academic community, they looked to maintain the links between academia and industry (Interviewee 16), to give a value to impactful work (Interviewee 17) or because of a sense of civic duty (Interviewee 19). Interviewees 17 and 19 also seemed to have some reservations about the qualities of the REF:

Interviewee 16.

[Impact assessor] I guess the main reason I was willing to take it on was because it seemed to fit with what I'm doing as part of the job. We certainly have an interest in proving the impact of [Discipline X] research. We don't want it to be irrelevant to what's happening in practice.

I guess a lesser consideration was I thought it would be of some interest personally, but to be honest I view it as something that fits in with part of the job. My employer kindly allowed me time to do this work, so far as it was in working hours, but of course it involves quite a lot of work outside working hours as well, as it does even more so for the academics involved.

Interviewee 17.

[Impact assessor] Oh, because I'm an academic in my first life and I probably... I care about academics so, one. [...] I'm , sort of, agnostic, because I'm no longer an academic. I think, if you're going to try to scale outcomes, you need some assessment framework to do so. And this one is incredibly burdensome but I don't know if there's a better way of doing it. The Impact was added this year. I thought it was a very interesting exercise. I did feel that it was a worthy exercise and, because its trigger is that you could get credit even if you had a level two academic product, and Impact, as opposed to only getting judged by scholarship. And it's UK taxpayers money.

Interviewee 19.

Why did I agree? Well, I agreed because... Well, various reasons really. Because I'm not in academia, but I've had a career which is on the fringes of academia, so partly I like to keep up contact with the academic world, to be honest, I thought it would be quite good to do it. Also, I suppose, some sense of civic duty, if you like. I know that, for all its imperfections, the REF is the mechanism we have for assessing and allocating funding, and no one really likes it, but yes, it seemed that... Yes, I was asked, so if I could do it, I should do it, and the compensation seemed sufficient for that, because obviously it's a time commitment, but I felt the compensation was fair.

For other academic Interviewees, participation was motivated by more personal reasons. Several interviews identified that taking part in the REF was a means to enhance their personal enjoyment of their working life (Interviewee 13) or as a means to contribute to career development (Interviewees 23 and 24):

Interviewee 13.

I was curious. That was a big factor. I knew that every bit of reading that was necessary for it would be delivered to my door. I like reading and I have been an editor of a number of journals, and when I stopped being a journal editor, I missed reading at volume in a very directed and focused way. The fact that there was a

workload and that most of it was reading was something that was incredibly attractive to me. These [points] aren't in a rank order at all, but, thirdly, I thought it would be my civic duty to accept. I thought that I knew enough about it but not too much to be able to do a good job. I think I'm a reasonably fair-minded person.

Interviewee 23.

I was really delighted. [...] I made new friends, and contacts, and got to write a special edition, you know, professionally, people know who you are and ask you to write things, and do things. It was great fun as well, I enjoyed it.

Interviewee 24.

I thought it's an acknowledgement of your position, and so on, within the university and within the community, so that's why I accepted.

Other academic participants (Interviewees 27, 4 and 33) found that they came under pressure to represent either their University's interests and/or those of the field within which they worked:

Interviewee 27.

it was a way of expressing my subject's interests in some way or another. There was some, let's say, encouragement from [University X] to be part of that process, I guess partly so we could – why would that be? – so we could influence it, so for prestige, and I suppose in some way a feeling that [University X] wanted to punch above its weight in some way.

[...] I think there's a strong feeling of disciplinary 'proprietaryness', if that's the right word, about it – a kind of pride in the discipline but also a sort of feeling of... Of course, no individual or no group can direct it, in a way, but nevertheless you're kind of at the heart of this sort of engine of disciplinary direction – even some sense of excitement about it, in a way. That, of course, when you actually come to do it, is overtaken by all kinds of pressures, and irritations, and disillusionment at some point.

Interviewee 4.

Because a senior manager at the university persuaded me that it was essential for the institution to have people on the panel and they thought that I would be appointed to the panel. I was reluctant, having heard from so many colleagues about how frustrating and unrepresentative they had found the experience.

I agreed partly because [University X] said it was so important and partly because I'm involved with quite an innovative and not a mainstream part of the field and it was a

question of whether that would be represented and there would be people who could really speak to it. [...] one of the issues with the REF is it standardises everything and operates from a consensus middle, so the edges of fields and the more innovative and more experimental work have little presence. I thought I would be able to speak to innovative work that would be under-represented, and that's why I agreed. But in the process I found out that this was not the case.

#### Interviewee 33.

When the offer is made to become the chair, the university would be very upset if I declined the offer, because this is an important thing to do for the community and so forth. But also, it's tremendously important for university. At one level, if you're offered to become a panel chair or a panel member, no university will be happy if you say, no, I think. As I say, my employer would be upset if I said no. I'll not be doing it again, that's for sure.

For other academic interviewees (6, 7) along with some of these aforementioned motivations, a central reason for participating in the REF was maximising income for their institution:

#### Interviewee 6.

The decision, I suspect, was taken by [a senior manager in my University] - who I think had quite an influence on who sat on the [sub-panel]. [...] I'd written the 2008 RAE for [University X] [...] and I wrote the REF '14 for [University X] I thought, "If I'm going to do it for [University X], I might as well do it nationally as well." It is only a financial exercise. I never had illusions. So, when you do the REF for your university, it is the money that you're looking at. You're trying to maximise income. It's nothing to do with academic quality. It's entirely about the tactics to maximise income.

#### Interviewee 7.

I did RAE as well. And I think, you know, modesty aside, I'd done a pretty good job on that, and I reasonably had good standing in the community, both in the professional societies and in all the various constituent bodies that were there. I was also the leader of a large, you know, I'd been the head of department of a large department of [Discipline X]. [...] In an institution which was making a large number of submissions to the REF, and so on. So I suppose that was the basis of me being selected. [...].

Oh, I mean, there's no question. There's no question. It's the most important, you know, all the constituent- every single reason spoke to it, you know? You're able to understand and shape your field. You gain knowledge in shaping, and in influencing, shaping your field. You are performing a public service. It's like refereeing, and

everything like that. This is a critically important thing for the health of the field as a whole, and it's also a vital thing for gaining intelligence for my own institution.

So for all of those reasons, I was concerned with performing. You're effectively creating the basis on which this unit of assessment will receive a certain amount of money. [...] I mean, that was why we were there. We were there to distribute QR. Yes, absolutely. [...] I mean, I was a senior university manager, so I knew exactly, you know. Absolutely.

## **Summary**

In this section we have loosely organised the Interviewees' motivations for taking part in the REF. Although a sense of civic responsibility and need to represent their academic field has perhaps been the dominant motivation (most frequently mentioned and most strongly argued), other motivations have also been put forward. Interviewees talked of a desire to change the REF for the better, to carry their misgivings into the process in order to make changes, to develop their own interests or career, to represent their institution and secure income for their department. Impact assessors also reflected similar motivations: a form of civic duty, a hope to make change by bringing impact to public attention and a sense of disquiet about the REF and its ability to assess such matters as impact.

## **B. PURPOSE OF THE REF**

A further, seemingly uncontroversial, topic for discussion among interviewees was the purpose of the REF. As they had all been involved in either chairing, assessing or managing research assessment, we had anticipated a fairly narrow range of responses among interviewees. However, as this section of the report will demonstrate, interviewees spoke at length about a broad range of intended, partially intended and unintended purposes of the REF.

We begin this section with those who have had a role in managing research assessment. Their approach to the purpose of the REF – as a means of selectively allocating funding according to excellence – was closest to an official or formal definition of research assessment. As Interviewees 1 and 2 both remark, research assessment in its various forms has been deemed a necessity given limited supplies of government research funding. However, as Interviewee 2 goes on to describe, research assessment is not a neutral tool of measurement:

Interviewee 1.

in 1991 [...] when the polytechnics became universities, and the money for research didn't increase, [it was necessary] to find a way of allocating the money selectively, and the only rational way of doing that was on the basis of the quality of the research outputs. So the Research Assessment Exercise was the mechanism of helping us to decide where we should put the money. [...] it's a way of identifying the good research where it really happens. That's why it's important that it should be discipline based.

Interviewee 2.

On one level our funding policy is to fund excellent research wherever is that. Selectively find excellent research wherever is that. The need to have a policy like that derives from a scarcity of resource. If we had a hundred times more money to disperse than we currently have, then we might come up with all sorts of ways of distributing that won't be based on excellence, we might just give an allocation to each institution, or whatever, based on their size. So that's the policy, and in order to implement that policy you need a way of measuring excellence, and that's where something like the REF or RAE comes in. Those are ways of measuring excellence in order to implement that policy. [...] it's a bit more complicated than that, because clearly the way in which this assessment is constructed influences the behaviour of researchers. And I don't think it's right to pretend that an assessment is a neutral instrument.

[...] So, the assessment instrument is a policy as well, and it needs to be thought of in terms of incentivising behaviour that you want to see and de-incentivising behaviour that you don't want to see. And then you are into all sorts of complicated things around unintended consequences and you mitigate against those.

In contrast to the preceding technical definition of the REF (albeit a technical measurement device that also steers activity), for other interviewees, the purposes of the REF varied broadly. The most frequently articulated purpose for the REF was the distribution of funding. However, rather than a technical matter that simply followed on from the results of measurement, selectively distributing money was viewed as a policy choice (Interviewee 19), a focus for competition to get a 'slice of the pie' (Interviewee 10), prestige outweighing money (Interviewee 14) or producing focused league tables that become the monetary focus (Interviewee 22):

Interviewee 19.

my understanding is that it's essentially the idea that there is a finite pool of resource to be allocated to academic departments in this area, [...] and there has to be some way of determining- Well, no, there doesn't have to be. You could either just distribute the money equally, amongst all institutions, or you have to decide that some of them merit getting more resource than others. If you go for the second, which is what we do, then you need to somehow have a, however imperfect it might be, some kind of ranking of the quality of the department. So it's to determine- I see it as a way of trying to determine the quality of the department so that resources can be spent in ways that are following quality

Interviewee 10.

On the positive side, it's a snapshot of research activity. It identifies areas of strength and weakness. [...] It's now perceived much more as being to do with getting a slice of the pie. I don't think it was originally. I think it was seen as something else. I think was quite a shock for people to think there was this national evaluation. Although it was tied, in the beginning, to funding, I think people were not as focused on that part of it. It was more focused on identifying people who really weren't doing their job and identifying people who were doing a fantastic job and so on. Now, I think, everybody sees it as much more to do with allocation of resources.

Interviewee 14.

I think in top departments it doesn't have much of an effect because top departments compete globally. [...] But I think where it makes an enormous difference is below the top departments, there's very few departments in the UK, at least for [Discipline X], that operate in a global market. Most of the departments don't, they operate in a more local market. I think it really puts research productivity to the fore, and it prevents an equilibrium in which universities just become a teaching place which is, there's nothing wrong with teaching but if you're a research university you should produce research and the fact that there is this exercise every four years... even if the money attached to it is not that much, it's the prestige attached to it which is

important I think. So I think it gears resources towards research in universities, it makes research a priority and of course, being in a leading research institution, I think that that's a good aim.

Interviewee 22.

Well, obviously, its function is to aid the distribution of money [...]. But also, everybody knows it creates a league table, which- now, in some sense, I'm not particularly fond of the idea of creating league tables. [...] Of course, that's massively important.

Other interviewees expanded on this discussion of the financial purpose of the REF by talking about the pressure that such a focus on money brings. Interviewee 23, for example, suggests the REF has become a means to streamline funding to the top universities and to side-line teaching; a situation that creates pressure for newer universities:

Interviewee 23.

Actually, on some levels I really disliked the REF. [...] And so in terms of using it as a way to distribute money, because really, the REF panellists are sort of impotent, because they don't have anything to do with the delivery of the outcomes. I mean, it's a completely detached process. [...] Coming from a lesser university, there is this anxiety [...] that it's a mechanism for streamlining funding to the very top percentile of the universities, and I did worry about that, and I did worry about that being the agenda, but I tried to detach myself from that and just tried to get my head down and do a good job for this first stage, thinking a bit powerless about the next stage. So, that's a bit of a confession, really.

[...] I feel passionately that there should be links between teaching and research, and of course that's what I try to achieve here, and it's one of my passions, but that was another painful thing about the REF, is that teaching is completely side-lined [...] But you just get on and do it. Nothing's a perfect process.

Other interviewees suggested that the pressure that emerges through the financial focus of the REF is reputational. Interviewee 25 suggests that the importance of reputation varies between different types of institution and Interviewee 28 suggests that the REF has become a means to allocate prestige, not just funding (but perhaps through an assessment system whose results are misleading):

Interviewee 25.

Well, in a very narrow sense, of course, it's there to distribute the QA money. It's a funny exercise. Of course, it's quite interesting how different HEIs look at REF. Some

treat it much more as a reputational exercise. This is a broad generalisation, but, generally speaking, the newer, younger, less research-intensive universities treat it more as a reputational exercise. Whereas, the well-established Russell Group-type universities like [University Y] treat it as purely a funding thing. [...] We did quite badly in the previous RAE, but we got a lot of money and that was all that mattered to us.

At [University X], we don't rely on what position we are in RAE or REF for our reputation. It's got nothing to do with it. For other places, [...], it was important for them to come top. I think, actually, the one that sticks most in my mind was probably [University Z]. Actually, that was for the REF. They put in a relatively small fraction of their people. They, therefore, of course, lost a lot of money, but, nevertheless, came very high up the list in terms of all these ranking tables that people produce.

#### Interviewee 28.

My own view was that the original strategy for the RAE, which was a peer-reviewed strategy which appointed peers to the panel and gave them a great deal, as we understood it, a great deal of latitude and discretion in how they evaluated the departments that were coming to them, that that was a very successful exercise. [...] My view of the process as it developed more recently was that it had become very, very top-heavy, very, very onerous, requiring spurious accuracy in the judgements and generating results which ranked institutions on the basis of tiny little variations in their profile.

The profile seemed to me to be constructed through a rather arbitrary algorithm; the allocations seemed to me to be misspecified, in that if one took them very seriously – four-star research – if any individual in their career produced one piece of four-star research, then they would be very fortunate. The way in which these star levels had to be interpreted seemed to me to leave a lot to be desired. I generally thought that the whole exercise had grown to a size, and a complexity, and a cost that far outweighed its usefulness as an allocation device or as a ranking device.

[...] everybody who's on the panel knows that it's both a means of allocation – allocating funds – and it's a means of allocating prestige. Although nobody quite knew at the time what the formula was going to be which would translate rankings/star numbers into money, it was clear that that was the case. Nobody really knew what the consequences would be for institutions of where they fell out in the ranking by the grade-point average, because in fact we were not doing that grade-point average ranking until right at the end of the exercise.

The final point made by Interviewee 28 in the preceding excerpt – that the ranking was not apparent until the end of the exercise – links to a common theme in Interviewees' discussion of the purpose of the REF. As a form of research assessment, money and time were intricately linked. For Interviewee 11, the REF has to function to capture consistency or change over time in its allocations of funding. For Interviewee 29, time enters into

consideration through the amount of time allocated to try and succeed in the REF's financial allocations:

Interviewee 11.

It is about the allocation of money [...] the history of scientific endeavour is that there are innovations and there are new perspectives and disciplines and knowledge moved forward by innovation. And unless there is some mechanism to capture and monitor that, then you could find yourselves missing out on the potential to put resources into new and promising areas.

I mean if someone has always been good and continues to be excellent then of course that should be funded. [...] But you could look at other projects where really important innovations happen, but then after a while there are diminishing returns that set in. So I suppose the REF panel is amongst other things, making judgements about where the high quality and where the best value is in terms of investing these resources.

[...] So the recognition that, although it is a lot of work, it is probably like the Churchill statement about democracy. It's the worst system except for all the others. It's not ideal the REF and the RAE, but in the absence of it can you think of a better system for allocating resources to promote vibrant and dynamic disciplines and knowledge and things? That kind of argument.

Interviewee 29.

The question, of course, is all about the amount of work, and the amount of preparation, and the extent to which these full-blown exercises devour huge amounts of very qualified researchers' time. As we all got into the system, what we realised was just how much time it was going to take.

Other interviewees, linking back to the reflection on changes in research assessment mentioned by Interviewee 28, discussed at length the changing purposes of the RAE and REF over time. The focus on being an active researcher that research assessment has encouraged (Interviewee 31), the benefits of research assessment for women (Interviewee 13) and an emerging focus on the good of academia for society (Interviewee 26) were all identified as positive changes in the developing purpose of the REF. However, for Interviewee 13, these positive changes were also caught up with a move from a qualitative reading project to a focus on un-evidenced indicators:

Interviewee 31.

I think it starts out with a kind of meritocratic ethos for some, and so it's not just simply a neoliberal story, or it's a neoliberal story tied maybe to the meritocratic

version of neoliberalism if you like, that strand within the emergence of this entrepreneurial, individualised way of operating as an educator or as a scholar. [...] I don't think the old order was perfect by any means, it's not about a romantic harking back to how great things were in the 1980s, because they weren't great. They weren't great at all actually. There was a lot, I think, of wasted effort and wasted talent on the part of intellectualism, and that's what I care about. I care about ideas and I care about writing, and teaching. [...] So I don't have romantic feelings about that time because not all of it was great, it certainly wasn't. So the idea then that, you know, teachers would be active scholars, active researchers and writers, I don't think that's anything to question actually, I think that's something to be happy about.

#### Interviewee 13.

Back in 2004, it was primarily a qualitative reading project which focused on published outputs, in which people were expected to research and publish at a modest level but in a continuous flow. I'd seen, from the previous RAE, that the outcomes had been used to reward people who had produced published work at a minimum threshold level. I thought that it massively - and I still think this - benefited women, because you no longer had to have a penis to do well.

It wasn't a matter of, "He's a chap. He goes out drinking with us. We will promote him." It would be, "He has no publications. She has four."

[...] What's changed is that as this creature... It's like evolution, isn't it? It's like watching a dinosaur. As evolution has impacted, other measures, not just the qualitative reading of the published outputs, have become a major part of the overall outcome. Into it goes, "How much research income?" "How many research students?" impact, environment and all of these other things which are not a matter of colleagues in an academic discipline reading in a collegial way. [...] I'm not eligible to be on another one, but I wouldn't agree to be on another sub-panel because ... the emphasis now put on impact is largely unevidenced and is simply a case document written by the unit of assessment or somebody acting for the unit of assessment. It's unevidenced and highly problematic. But it has a multiplier in terms of its percentage impact on the overall outcome.

#### Interviewee 26.

So, I think in the RAE the emphasis was much more around individual scientists and high quality scientific outputs. In most people's minds it was about people, it was about individual academics. I think by the time we'd finished the 2014 REF it felt like a lot more of it was about what is the impact of good science for society and the economy. [...] Of course the down side of that is that within universities REF scores and outcomes are still used in terms of individual scientists. Who are we going to hire and fire. So there is a little bit of that loop that's missing at the moment I think.

In contrast, Interviewee 33 suggested that the purpose and function of research assessment has not changed much at all:

Interviewee 33.

So they introduce impact so that they could claim that serving on this panel would be an impactful thing [...] You know that the money follows. Because it's been doing it since the very beginning of the Research Assessment Exercise, it's always the money for research quality has followed the assessment. We didn't know how it would follow. At one level, there was no way we could know how it would follow because for HEFCE, who determine the formula, they're not going to reveal in advance how it will follow because they don't want panels to be game playing.

Beyond a focus on the allocation of money, other Interviewees also pointed to the importance of other purposes served by the REF. For example, Interviewee 18 and 7 both emphasised the importance of gaining recognition for UK academic research. For Interviewee 7 an emphasis on the purpose of the REF as a means to develop recognition for UK research would require a change in public dialogue regarding the purpose of research assessment:

Interviewee 18.

I think it was about identifying and locating excellent research. So for me, I think we do amazing research in the UK, and it's about recognising that and rewarding that on every single level. [...] It's about bringing a wider recognition of the excellence of that work to as broad a public as possible, both within HE in this country, HE in other countries, but beyond HE also, nationally and internationally.

Interviewee 7.

If the purpose of the REF is just to distribute QR, that's one thing. But it's not just to distribute QR. Actually, it's about other mechanisms, other policy-related goals. It is about developing a community dialogue around quality. It is about drawing attention of researchers to the primacy which policy attaches to certain classes of impact. It is about driving up the global position of the UK, with respect to its competitors, against certain key measures. It's about transparency and auditability, and consensus, within the community. [...] So the point here is that we have a dishonest dialogue about the REF. The dishonesty being that we know its purposes are one thing, and yet we sometimes say it isn't.

So this is for government to decide. If government wants to decide the purpose is simply to distribute QR, then this is a monumentally inefficient way of allocating QR. If we believe that a different sort of dialogue is needed round research in the community as a whole, for the UK as a whole, then maybe what we're doing is the right thing, with some adjustments.

This theme of developing recognition for UK academic research was developed further by Interviewee 20 who suggested that the importance placed on recognition, led to the development of different REF strategies by different universities:

Interviewee 20.

I think, for me, it gives some assurance, it can never be 100%, but it gives some assurance about the quality of UK research with respect to, if you like, the international research agenda. [...] Obviously, money is important, but actually, in many ways, the ranking became the dominant factor. [...] But I think they're both important, actually, really, because it is part of the market. The brand. It was particularly important to most research intensive universities. My institution is one that's in the 20 to 30 bracket. It's particularly important for a university such as ours to make sure we retained that research intensive branding.

I was aware of that from my own institution's point of view. I was very interested by the fact that not everybody played the same game. For some, the ranking was by far the most important thing. For others it was really important, I think, for the bigger universities especially, that they got significant amounts of funding because they have got many more mouths to feed. [...] In the end, that strategy and the tactics that were used by different institutions to make the decision about the quality of the research versus the quantity of researchers submitted, that interplay between them is two things. It was an individual decision, I think. I think a number of institutions in that middle bracket made different types of decisions that affected their league table in one direction and in another direction...

Following on from the suggestion that distinct REF strategies enabled different universities to try and prefigure where their results might position them and how they might most benefit (in terms of reputation or finance), Interviewee 21 suggested that figuring out the purpose of the REF in this way could lead to quite elaborate game playing:

Interviewee 21.

interested in finding ways of assessing where the high quality research is being done, where there is a high quality research environment and recognising that fact. And one would worry if one thought that the results that the REF produced did not fairly represent, in fact, the places where high quality research was being done.

In fact, having been involved in the last two exercises I'm pretty sure that they have – at least in my subject area – fairly recognised areas where a high quality of research is being done. My sub-panel was not a sub-panel that suffered from any serious degree of game-playing in the Ref 2014. I think there are some sub-panel chairs who would say, actually, the results were distorted because the rules... We rigorously followed the rules – but rules allowed certain forms of behaviour that meant that departments

could represent themselves as a lot stronger than they were. And one was only allowed to go on the representation, as it were.

These were not the only views articulated on the purpose of the REF. However, there were only small numbers of other interviewees who placed emphasis on a particular purpose of the REF away from finance and reputation. For example, Interviewee 5 suggested the main purpose of the REF was about accountability (see also section E.5 on the effects of accountability), Interviewee 32 placed an emphasis on community creation and Interviewee 9 suggested that a purpose of the REF was marginalisation:

Interviewee 5.

It's the idea of institutions being held to account for the way that they have invested the public money and what has come out of it. Sometimes, if the first of those two reasons was questioned, then HEFCE officials would fall back on the second as being at least as important.

Interviewee 32.

So that whole process over the assessment exercise in the whole academic community across [X] main panel changed behaviour of individuals enormously to the benefit now of their own reflections on process. Because we broke down quite a few of those barriers and created nice communities that were working together but had very different perspectives. It was great.

Interviewee 9.

Ostensibly, its purpose is to help divide the available funds in an appropriate way. From the very start, it's been perniciously competition-inducing, stress-inducing. It's been [...] a [...] way of excluding and marginalising some institutions, and indeed some people. So, whether you count those as purposes or not, I don't know. I mean, I perceive them to be not unwanted aspects of its purpose.

## Summary

In this section we have noted a variety of purposes for the REF and its predecessor the RAE that were articulated by interviewees. Although a formal definition of the REF as a means of selectively allocating research funding was set out by the two interviewees most involved in the management of research assessment, other interviewees broadened and diversified what they perceived to be the central purpose of this exercise. Most prominent were discussions of financial allocations being linked to competition, prestige and reputation and the effects of universities identifying and pursuing these purposes. Others suggested that the purpose of the REF changed over time, had addressed issues of gender inequality, and had a positive benefit

for the reputation of certain institutions or for UK research as a whole, enhancing accountability and even enabling community building. However, other interviewees pointed to the ever increasing stress and pressure of the REF, its effects on streamlining funding and marginalising certain institutions.

### C. THE WORK OF MAKING THE PANELS

A key topic of discussion among interviewees that emerged from their responses rather than any particular prompting from our questions was the work done to put together the REF panels, both Main Panels and sub-panels. Here a delicate art seemed to characterise negotiating who would be nominated, how efforts would be made to get certain participants to accept, what the balance of the panel ought to be and what should be balanced (expertise, gender, geography, University type and so on).

In the following excerpts, interviewees discuss this work of panel building. As Interviewee 32 suggests, Main Panels were made up of sub-panel chairs, based on recommendations from subject associations and others. Getting the right kind of people seems crucial. But as Interviewee 3, 34 and 19 also note, consideration had to be given to representing different regions of the UK, age, departments, a gender balance and the types of University being represented:

Interviewee 32.

So the first thing was to get the Main Panel right, which was to pick the chairs of the subpanels, the sorts of human beings who would be good at doing that and we had a list of different ones that were put forward by various organisations. So because the human qualities of these individuals was going to dictate their performance in this new world of interdisciplinarity, we couldn't have narrow thinking, and it would be unacceptable to have somebody trading their own particular biased interest in such a setting. So the key thing was to get those chairs sorted first, and that was great fun, picking the right qualities of the individuals. [...] Once we got them, then of course working with them to pick the people who would populate the subpanels.

Interviewee 3.

we had gaps - somebody who died, somebody who left, that sort of thing. [...] at least you [need to] have experts who can defend their opinion on all the papers. [...] And obviously we've had a reasonable breadth across new universities, old universities, south, north, whatever. [...] The final [sub]panel was huge. It would have been really hard to get more people to interact even, I think. We were about 30 people maybe.

Interviewee 34.

Initially, the members of the subpanel, as distinct from the assessors who would be appointed later, the members were all nominated by the subject associations. One of the problems that I faced, because I, as Chair, had to look through this list of nominations and try and find a balance in terms of subject expertise, gender, kind of university, research approach, etc., etc. Sometimes we had the same people being nominated three or four times. That was why we had to go back, I think, twice, certainly once, maybe twice, to the subject associations and say to them, for example, "We need more women. We need people in these subject areas," and so on. [...] I

think we ended up with about a third of women on the panel. It should have been more.

Interviewee 19.

It seemed to have representatives from different kinds and sizes of departments. It had a decent range, I think, of level of expertise. I think it had younger and older people, but almost everyone was pretty experienced. Yes, I thought it was a pretty good spread. The only sense in which it perhaps wasn't representative was [University X's] mega department [...] didn't have like several people.

However, these were not the only considerations. As several Interviewees suggested, panel members were also drawn in on the basis of their expertise in carrying out research assessment. As Interviewee 25 notes, it was often considered that previous experience of the RAE was a necessity for a sub-panel as a whole:

Interviewee 25.

[the subject association] asked widely for suggestions for panel members' names, so we essentially ended up with a large number of names that had been suggested by all sorts of different stakeholders, as we were supposed to call them. Then, there was a selection panel [...] You wanted a mixture of people who'd done the RAE. I think we tried to get about half and half. Half had done it before, and half were new. Maybe it was a third and two thirds. Practically speaking, in the end, it was a third and two thirds because quite a few people had had enough after RAE and didn't want to do REF. [...] The other thing was both the male-female balance, and also the geographical balance. Of course, all those things are highly over-constrained, so, in the end, you can't do an awful lot about that. It's, of course, particularly difficult in [this subject] to get the female representation well because there are so few women.

Once composed, sub-panels often proved difficult to hold together. Interviewees spoke of small conflicts and internecine warfare between different sub-disciplines drawn together onto the same sub-panel and old wounds being re-opened in relation to problems that were said to have characterised the discussions that took place in previous RAE panels. As Interviewee 33 suggests, preventing a sub-panel of many different academics from fractionalising was an on-going task:

Interviewee 33.

HEFCE continually wanted to limit the number of people. I ended up appointing more and more and more in order to make it work. They did eventually agree, because every person has to be paid, they have to have a hotel room. It costs them money they

thought that they could reduce. They didn't understand it properly. So, yes, there are limits. But equally, keeping a panel together, I ended up having [N number of] people including the Secretariat sitting around the table, that's a very big panel to manage. To chair, to keep everybody together without it fractionalising and so forth. [...] If you were to try to do this with somebody who was a manager who was not fully aware of the potential political and professional and other influences, there's a real risk that the members would take advantage of them. Would sew everything up. Academics are very good at this. That's very intelligent to do. I think you need an academic to be chairing it.

This difficulty in preventing one sub-discipline from taking over a sub-panel ('sew everything up'), was only one part of the problem of disciplinarity in a sub-panel. Various Interviewees (7, 18, 25, 40) spoke of the need for a sub-panel to have the right kind of expertise to both carry out the assessment, but also to adequately be seen to represent the discipline both on the Main Panels but also to the broader public. Having to fight for your discipline's interests (Interviewee 25) appeared to contrast with indecisive chairs (Interviewee 40) who might generate the opposite effect:

Interviewee 7.

I think that our panel chair was outstanding. I think that he did an excellent job in representing the subjects' interests to the panel, and vice versa, in making clear the compromises that were going to be necessary in certain regards, in order to- he succeeded in making quite sure that there were no objections to our chosen methodology, quite rightly so. [...]

Interviewee 18.

I mean, how long is a piece of string? You could make a case that you would need 50 people, but you can't have 50 people. So you have to have people of a certain seniority. You have to have people who have a breadth, and who have particular expertise, and whose work is also of a particularly high standard. And you need to have geographical breadth. You need to have breadth across the kinds of institutions that are going to be submitting to your subpanel, and you need to pay attention to issues of equality and diversity. [...] not all of them say yes. [...the role of chair] it's managing the business, then managing the process, overseeing the governance of the process, but it is also about representing [...] Now that's really important. So, where do we have agency? And how do we make sure that our concerns, and things that are very particular to us and our ways of working are represented?

Interviewee 25.

It was quite a strenuous job, and it's very important that you have a clear view as to what ought to happen. Of course, for everybody who takes on this job, their primary

concern is to make sure that their subject is not disadvantaged. You have to fight very hard for your own subject. Not in an unfair way, but if you don't fight for it, nobody else will, so you have to make sure that you're not somehow or other being judged on an unlevel playing field. That was what we were trying to do. There's no doubt that for whoever does that job next time, it's a very responsible job. They've got to make sure that they don't get themselves disadvantaged.

Interviewee 40.

in [RAE] 2008 we had an extremely powerful [person] who was extremely bossy, and put the fear of God in you, really, and our panel chair this year [...] I can remember actually chatting to some people over dinner and thinking, "Gosh, I wish [this person]'d be a bit more decisive." Somebody says something, he didn't tell you what the resolution was, so everybody said something and nothing happened.

Alongside this effort to be representative and to adequately represent a discipline, Interviewees also discussed the problems that arose in cases of what they deemed to be inadequate representation. Interviewee 9 suggests that non-traditional universities were mostly absent, Interviewees 28 and 11 discuss a bias toward older, more senior figures, Interviewee 27 suggests decisions effect how a discipline is treated and Interviewee 4 suggests that more innovative figures from their discipline went unrepresented:

Interviewee 9.

I think the one flaw was that it was very under-representative of institutions at the lower end of the hierarchy. So, it was only one person from a non-traditional university [...] but really, it's such an unfair exercise that there was very, very, very little we could do.

Interviewee 28.

As far as the composition of the panel was concerned, I think it is difficult to find the best ways of composing a panel like this. Inescapably, the panel consisted of rather senior people. Being senior, that meant people who were of a certain generation, people who'd been in the academic world a long time. [...]

I don't know the age distribution of the panel, but very few of them would have been under 50 years old, I think. They were making their decisions on the basis of their understandings, which were inescapably shaped – not determined but shaped – by their view of how the discipline was, their view of what the important issues were, and so on and so forth. But that's what you get if you get peer review.

Interviewee 11.

what you have got is people who are further on in their career. There was no one under 30 around that table. (Laughter) At one level you might say, “Well of course there wasn’t.” [...] Let’s say the average age was probably 50s, which would include me as well, then I suppose it is just possible that there may be things that have just taken off, that don’t get sufficient attention from the people on the REF panels who are established and arguably at one remove from some of the new developments.

Interviewee 27.

while one would never, ever say, “You’re privileging certain things over other things,” nevertheless there were decisions to be made. Of course, that’s being done at the level of who you appoint as the assessors and all that sort of thing, which has an inevitable effect on how the whole process is regarded and how people think the discipline is being treated. [...] So, you are giving off – you’re undoubtedly giving off – messages about how you view the discipline, in every decision that is taken, to be honest (Laughter).

Interviewee 4.

I didn’t think the panel was adequate. . In part because its members were not always able to understand work that did not conform to scholarly conventions or experimented with new methodologies. These panels seem to look far more to the past than to the future. The academics and practitioners who would have had a better chance of understanding, would probably not take part in an evaluative exercise such as REF as it does the opposite of forwarding and rewarding new work.

I don’t believe in national auditory exercises. [...] So REF is hardly isolated, it’s part of a much larger picture dominated by surveys and constant evaluation that takes up time and contributes little to pedagogy, student well-being or lecturers sense of making a significant contribution So I don’t think that there’s anything that can be tinkered with in REF that’s going to make it better. The basic concept is flawed because it’s a vehicle, not for improving things but for mechanisms of generalized evaluation that justify austerity and lead to mainstreaming.

Finally, Interviewee 20 suggested that issues of representation also had an effect on the work of the sub-panel. As one of the few females involved in the assessment of her field, she felt ‘mildly intimidated’:

Interviewee 20.

I came in late as chair and I also think I was the only female. [...] But I think there weren’t any other females in the group. Most of them had done REF before, as well. I think you could say I was mildly intimidated by it because I’d not done it before. I

didn't know whether what I was hearing was what they'd heard in RAE 2008 or whatever. [...] But our main panel Chair was an extremely affable and helpful person. He looked after me. I think the Secretariat did as well.

## **Summary**

In this section we have noted a common concern for Interviewees who participated in our research: the representativeness of the REF. For many of the Interviewees this was a concern regarding the relevant forms of expertise that a sub-panel would require. This was a concern regarding the range of work that a sub-panel would need to assess, but also a concern that the sub-panel looked right to the Main Panel and to audiences outside the REF. It was also a concern for ensuring that the discipline itself had adequate representation, perhaps even someone to fight on its behalf to defend its interests. Yet Interviewees also raised concerns regarding the adequacy of representation on sub-panels of gender, geography, age, previous experience of the RAE, departments and university types (old, new, small and large). This appeared to have an impact for these Interviewees on the outcomes of the REF.

## **D. ASSESSING ACADEMIA**

As a central activity of the REF is assessing academic work in various ways (outputs, impact, environment), unsurprisingly Interviewees devoted much of their time to discussing the various forms, practices and issues that arise in carrying out assessment work. In order to present these discussions in a coherent way, we have divided this section of the report into four sub-sections covering peer review, calibration, metrics and impact.

### ***D.1 PEER REVIEW***

Peer review remains central to the assessment work of the REF and has been central to the preceding RAE over several decades. Although suggestions have been made for moving toward a more metric REF, it seems that peer review will retain its centrality, even in association with metrics. Interviewees spoke at length of the challenges and benefits of carrying out peer review. A frequently recurring feature of discussions with Interviewees was how many outputs they had to review and how easy or difficult they were to assess. For the following Interviewees, despite high numbers of outputs to review, it appears they thought it was ‘fine’ to review nearly 900 outputs in a week (Interviewee 6), found it reasonable because they were given leave (Interviewee 9), found it straightforward as they only had to assess a few outputs (Interviewee 10) or found it relatively straightforward as fewer departments were submitting to their panel (Interviewee 22):

Interviewee 6.

Eight hundred and ninety-two. [...] It was fine. [...] Assessing outputs is what I do; it’s my job and I’m doing it constantly. I’m reading CVs on a daily basis and assessing the quality of publications, so to do a big tranche... It took a week to do them.

Interviewee 9.

I think it was somewhere about 190 [...] Probably about third of those were books and the rest were articles. [...] my department gave me leave [...] I probably had about 20 things that I talked about with other people in my peer review group, so to that extent, it was always possible to do that. [...] Through the secure email, and sometimes... It was more helpful to do it at the meetings, actually. It was more helpful to send them the stuff, or to make sure they got the stuff, and then talk about it when you actually met them in person.

Interviewee 10.

The average on the subpanel I was on was about 250, which we all thought was terribly high. But in fact, we found out that other panels had as many as 400 and 450. They were a really high number. I, personally, had many fewer. They were also, of course, rated by other people on the subpanel, who had many more as well as the ones I rated to do. But I really had a much lower load than anyone else.

Interviewee 22.

I think it's a bit over 200, probably [...] I'm forgetting all the dates, now. But we started quite early on in 2014 with stuff, I think. We tried to do the outputs, I think first, or at least get a long way with them. So, we started reading fairly early on, I think. So, we had a few months. It was over a few months, the actual- and then bits and pieces would come through right through until summer, I think. You were reading on and off probably for, I'm guessing- I'm already forgetting the details, but five or six months maybe over that period. [...] I think it was fine. Well, I think it was quite a bit less than the previous one, partly because the number of, you know, submitting to us had gone down quite a bit (Laughter).

Other Interviewees discussed the difficult challenges of peer review. Interviewee 14 talks about taking the reviews on holiday to have sufficient time and Interviewee 7 found it quite challenging:

Interviewee 14.

I remember, we went skiing and my family was all skiing and I was reading papers (Laughter) [...] Maybe two hundred, two hundred and eighty or something like that, and we had to destroy the files so I can't even really check it anymore [...] I was happy with 99% of them but I had some which were more on the margin of the profession that I wasn't sure how to evaluate ...

Interviewee 7.

Eight hundred. [...] So given that the panel size is limited, you know, and given the nature of the discipline, there are fewer double-weighted submissions, the number of submissions is likely to be high. [...] I did it quite comfortably within that period of time. No, not comfortably. [...] It was a hard and difficult process. And now comes to a controversial thing, and I might as well get it out there for you, which is, the quality of the rating does not necessarily depend upon the length of time spent on the rating. On the whole, the more time I spent, I believe, the lower the quality of my ratings were.

Following on from the comments of Interviewee 7, this issue of time appears to have been central to considerations of the type of review it was possible to carry out within the REF. As we can see in the following excerpts, for Interviewee 31 the REF is not peer review, while for Interviewee 2 it is a type of peer review that is distinct from, for example, reviewing academic journal articles:

Interviewee 31.

How can it be, you know, that a prominent journal which uses three reviewers can publish a piece in their journal, that rigorous process, and then for that journal article to be read, within fifteen minutes or twenty minutes, there's another intervention here that's important to remember, and then be judged as two star, or three star. How can that be? [...] It's not peer review. The REF isn't peer review, the RAE isn't peer review. [...] They're judgments that are skimming the surface of work that's taken years to produce. That's not peer review. It might be being done by peers but it's not a review, I don't think, a serious intellectual engagement, it can't be. It's just impossible.

Interviewee 2.

There is a tendency when people talk about peer review to lump it together as if there is one thing called peer review. The process of reviewing a manuscript that is yet to be published and deciding whether it should be published or not is different from the process of reading an already published and peer-reviewed paper and making a judgment about where it sits on a quality spectrum. The amount of effort and in-depth thought that's required to do those two things is rather different. And so if you're to give a REF panel member the stack of outputs to review and say, 'These are all manuscripts and you need to decide whether they are going to be published or not', that would indeed be an impossible task in the time available. To give that same panel member a pile of outputs that says 'They've all been reviewed in one form or another – they've either been peer-reviewed, or the books have been through editorial processes (or whatever)', then that's actually a much more straightforward task.

The modified notion of peer review discussed by Interviewee 2 offers an interesting distinction between different conventions for reviewing. This alternative version of peer review certainly seems to be absent from the publicly available documents and discussions of REF. Other Interviewees picked out alternative issues they had with peer review. Interviewee 11 notes that peer review is a time consuming activity, but one for which there are few alternatives, Interviewee 12 suggests that there was not much time for each review, but the alternative of sampling or bibliometrics were not promising (also see section D.3 for more on metrics), and Interviewee 13 suggests that the way the REF now structures peer review makes it impossible to get a clear overview of the trajectory of people's work. Each of these views adds to a sense that the participants in the REF felt some disquiet with the review techniques involved, but also failed to see much value in alternatives:

Interviewee 11.

I am not in favour of peer review just to have lots and lots of meetings, because clearly REF is an incredibly expensive and time consuming process and all the concerns that there are about it. But I think back to the Winston Churchill point, that actually it's still less bad than the alternatives.

Interviewee 12.

I think, you know, we evaluated the amount of time that we had available, and we divided it by number of tasks that we had and worked out how much time we had, and realised that it was going to be less than the amount of time you would spend on refereeing if you were sent a paper from a journal to referee it. [...] I mean, you know, [...] I'm aware that some people have studied how well bibliometrics works, and I think my general impression is it's pretty bad. [...] I was involved, to some extent, in the idea of whether one might employ some sort of sampling methodology, combining bibliometrics and peer review, but that doesn't seem terribly promising, from what I've seen.

Interviewee 13.

It wasn't possible to do it this time, but the first time around, it was possible to read everything. [...] This time, it was delivered in a very different way. You could only access, on the members' website, the outputs that you had been nominated to read. [...] You get no sense of holism. [...] You can't look at the trajectory of people's work, think, "Well, they started out a bit rocky, but they're doing better," and give them the benefit of the doubt [...] When the assessment is driven by these more technical considerations of how you fill in the software, that kind of more qualitative assessment, of not only what you're doing but also what they're doing, is not possible.

Interviewees also spent some time discussing their efforts to agree on disputed scores. Interviewee 4 felt their scores were over-ridden, Interviewee 3 spent as long on discussing scores as carrying out reviews and Interviewee 43 was a member of a panel with a hierarchical scoring mechanism (1<sup>st</sup> assessors were given more weight than 2<sup>nd</sup>, with 3<sup>rd</sup> assessors 'often just arbitrating'):

Interviewee 4.

One marks with another member of the panel and in consultation with them so there is a discussion. On occasion that discussion is put to the panel which is not always useful as expertise in research details don't always indicate originality of thought or stature of research. The panel Chair can override the decisions and our chair did so

frequently. While various calibration exercises were done in advance, it is only when you experience a panel at work that you understand how little you can actually influence a judgement, as protocols override knowledge or opinions.

Interviewee 3.

I had about 700 papers to read. [...] we were allocated papers that were pretty well in my own specialty. [...] So I found the assessment relatively easy. I mean it probably took me a working week [...] When I'd gone through that, I found that a lot of time was rationalising scores with my co-panellists [...] a lot of it was around 2 or 3 boundary, basically. Where the very best papers were obvious, we didn't have that many very bad papers, but this 2-3 boundary took a lot of time. And we would have to have a discussion then about every discrepant score, and so I spent another working week, pretty much, discussing.

Interviewee 43.

Sort of 750ish that kind of thing. [...] I knew what was coming and the way we did it was every output was read by two people, with a third in case... [...] So the ones who were the first assessors had more responsibility than the second. The third quite often just arbitrating. [...] So for the outputs, I think we had about three or four months to do them. So I was sort of setting aside Mondays and Fridays to work on REF.

Alongside some disquiet regarding the peer review process, several Interviewees also raised concerns regarding the scores and the ways in which results were tabulated. As Interviewee 28 and 29 both suggest, this led to a kind of spurious accuracy:

Interviewee 28.

how this all figures through to the way in which the GPAs work out when everybody tries to rank all the institutions, and there's 0.01 of a... This is a kind of spurious accuracy, I think, in some of the ways in which this was done.

Interviewee 29.

What didn't work? For me, I felt there was a spurious false precision in the whole process. I'd have been much more comfortable had we been clustering institutions into bands. I don't think that we were... The precision that separated grade-point averages to two or three decimal points is spurious precision.

Finally, Interviewee 43 also noted that it was a useful idea to continually change assessors as they had got ‘too blasé about doing it’:

Interviewee 43.

I’ve probably got too blasé about doing it. I think it needs, once you’ve done it I certainly felt near the end that my little mental formula for scoring a paper was starting to become too formulaic. Early on I really struggled with how to score a paper, somewhere in the middle, at the end of the first one and the beginning of the second one I was doing it, I was reading the paper properly and thinking about it. Quite near the end I was feeling, “I know how to do it. Read the abstract. Look at the results, figures, and tables and give it a score”. I don’t think that’s good.

## ***D.2 CALIBRATION***

What became clear in carrying out interviews on REF panellists’ work was that peer review was not a standalone task. As we have already noted in the preceding section, scores were often agreed with other panel members. In this section we will see that the scoring itself was also underpinned by calibration exercises (that tried to establish a comparable basis for scoring) and by forms of normalisation (that sought to render different assessors and sub-panels commensurate).

As the following Interviewees articulate, calibration was about directing people in their assessments (Interviewee 24) and identifying assessors that were outside the range of scores of the rest of the sub-panel (Interviewee 6):

Interviewee 24.

It’s a calibration exercise, so we’re given then a limited number of papers to look at. [...] and there was a huge scatter of marks, that was the whole point of the exercise, to try and direct people into what they’re looking for, and so on.

Interviewee 6.

We did a calibration exercise and realised that most people were pretty consistent, and the outliers were identified quite early on. That was quite a useful exercise.

As Interviewee 10 notes, however, this was not a one-off process. Calibration had to be continually re-explored as assessors entered their marks into the REF system:

Interviewee 10.

I actually thought it [calibration] was very useful. Again, more to the point of helping individuals understand whether they were looking to identify strengths or weaknesses. People really differ in what they think they're doing when they rank stuff. There wasn't a big enough sample to do much other than give yourself some pointers. But it did, I think, it suggested to some people that they were out of line with others.

Although, interestingly, that kept having to be re-explored. We were each sent our own individual distribution of scores regularly, as we uploaded our rankings. Then I could see where I was relative to other people, for example. I think calibration is something that was ongoing. There was a calibration exercise early on. [...] But there was [also] constant calibration.

This Interviewee goes on to talk about the delicate politics of calibration and how it can enable sub-panel members to see whether they are over or under marking

Interviewee 10.

Do you go out cheerleading and give everything in your field really good ratings because you think it's really important to support it? Or do you sit there and then pick holes in things on the grounds so that work could be better and so on? Without a comparison with other, assuming every sub-panel is going to have a similar distribution, just statistically speaking, it's really important to see that. Because you can then say, "Oh my goodness, we're ranking everybody far too low because our curve is way down here and everybody else's curve is much higher."

Similar points were made by Interviewee 35 on consistency, Interviewee 38 on avoiding systematic difference, Interviewee 41 on the sub-panel being aligned with its Main Panel and Interviewee 25 on the importance of building bridges through shared scoring:

Interviewee 35.

it was useful to begin to get a sense of how different people judged. We did have a reasonable amount of recalibrating to do, to make sure that we broadly agreed. Inevitably there were the tough guys and the softies, but we were able to see that and begin to move people closer together. [...] So whereas in pre-2008 the individual subject panels had a fair degree of autonomy to shape their own understandings of the criteria the idea, in order to get greater consistency across panels, was that more of this would be a top down process. [...]

Interviewee 38.

What we did in our subpanel, later on, was a lot of statistical analyses of how people were grading papers. [...] we controlled for a number of factors, controlling for

journals, controlling for where the outputs- everything you could control for, we put into the thing. And pretty much everybody was- you couldn't distinguish, looking at people. Nobody looked more- hardly any difference in terms of generosity or being harder. So, I think the calibration- and this was done very, very thoroughly. It was quite extensive analyses. It wasn't obvious there was any systematic difference between people.

Interviewee 41.

I can remember in our meetings, we would get feedback from our chair who'd been to the main panel chair, and he would say, I don't know if he was allowed to say it, but he said it, he'd say, "We're doing okay, chaps. We're on the right track." So, that was going on the whole time, and we had a lot of pressure on us to get done by these deadlines. There was a lot of time-pressure for the deadlines, and specifically so they could do the calibration. And it's something we spoke about a lot, because it was such a disaster [...] before.

Interviewee 25.

I'm not sure if anything would have changed if we hadn't have done it, but it was very good in establishing the appropriate community interaction inside the Main Panel. It's very easy, if you're in charge of [discipline X], to be at war with [discipline Y], and this exercise, I think, helped to build bridges so that we all felt part of the same exercise and that we weren't going to dissolve the whole thing into civil war. That isn't to say that civil war didn't occasionally almost happen.

Continuing these themes of the importance of calibration and the need to build bridges, in the following excerpts Interviewee 29 discusses the importance of sub-panel members educating each other and Interviewee 32 talks about the importance of avoiding problems that emerged in the previous RAE:

Interviewee 29.

That initial calibration exercise was very educational, because there was such huge disparity. [...] You know, it was just, sort of, very stark that we needed to educate each other, to be open-minded about certain kinds of research, which otherwise might not have been given due credit to or countenanced in that initial calibration exercise, yes.

Interviewee 32.

[calibration] was a huge success. Had that failed, it would have been a disaster, it really would, because last time in 2008, there were some issues between subpanels

not being able to agree a level playing field in different disciplines. Whereas this time, every time I went around the subpanels as I did, the first thing you say is, “We’ve got to get the benchmarking process right. We’ve got to make sure that if somebody puts a four in a [discipline X] panel and a four in [discipline Y], that they have the same sort of currency. You can’t have one very much higher or lower than the other.” That was a bit of a challenge, but an enjoyable one

Although calibration was identified as hugely important to Main Panels and sub-panels, various Interviewees also talked of the normalisation of scores. At various moments, the scores of a whole sub-panel could be balanced (Interviewee 4) or an individual could be confronted with their low scores (Interviewee 20):

Interviewee 4.

There was internal dissent in the panel with disagreement on the actual criteria for marking. There was also the presence from the REF administration that wanted panels to balance their conclusions, rather than having some be very generous and others be very harsh in their judgement, and in their opinion to ‘sabotage the field’. Eventually an algorithm was applied that brought grades into balance both within the panel and between panels. Such an application of a computational model does not allow for the breadth and multiplicity of positions in a field to be recognized.

Interviewee 20.

There was an individual in our panel who scored very low, right the way across the board. The other panel members were uncomfortable because they were consistently getting different marks to him. I had to deal with that. I had to say to him, "You don't appear to be in line, why is that? Is it that everybody else is wrong?" We put quite a few papers against him with lots of different scores to see whether it was just a particular area. He didn't completely capitulate, probably rightly so, because you'll always get differences [...] If everybody agrees, if everybody says, "Look, if we all agree, we can go home and we don't have to worry about this," that's the easy way out. Those people who are more polarised in their marking, actually brought some credibility to the process, in my view.

Building on this theme of what ought to count as the correct score for an output, other Interviewees discussed a variety of problems with ensuring consistent scoring. Interviewee 14 found that they had to review their own scores, Interviewee 33 found that they had to ensure that new assessors on a sub-panel merged from distinct disciplines could score together, and Interviewee 30 found that papers that were difficult to reconcile, eventually had to be scored through an agreed formula:

Interviewee 14.

We had a million meetings, they lasted forever. [...] I remember reading papers and then going back to them after having read more papers, because the opinion you get, you're really happy at the beginning and then you see better and say "wait, that wasn't so good" (Laughter) and then you go back and forth so that you get a consistency across many papers.

Interviewee 33.

The calibration exercise I chose to do [...] is you calibrate the instrument. So it's the panel members are the people who are being calibrated. [...] So, actually, what this is about is that... There were certain people who, maybe a quarter of the panel, had been on the previous RAE. The other thing that we had [...] was a merger [...] That meant I had some people who had previous experience of the other panel and then other people with previous experience of this panel and then a whole lot of new people who had no experience of doing the exercise at all before.

Interviewee 30.

Eventually, it worked that as the deadline for completing the assessment approached, you know, November 2014 or whatever it was, and we realised there was so much assessment still to complete, we introduced a formula for reconciliation. You know, if there was agreement within a certain band, you just took an average. If it was outside that band, then you had a discussion.

As Interviewee 7 made clear, the important aspect of scoring was not so much that a paper scored a certain number (particularly given the possibilities of normalisation). Instead what was important was that the score enabled the outputs to be positioned in a relative ranking to each other from good to bad:

Interviewee 7.

My individual ratings for papers, or whatever outputs, doesn't matter. It's irrelevant. What is relevant is, how accurate my overall assessment of the set of papers was. [...] I mean, how well it represents the quality of the outputs against the criteria identified. [...] In short, actually, this is about the law of large numbers. And actually, the law of large numbers works better when there are large numbers. [...] So you ask me, any given paper, you know, was it a three or a four? Well, you know, I did my best. But overall, over the full set of papers, do I think I performed a reasonable assessment? Or do I think we, collectively, performed a reasonable assessment? [...] The point is that individual rankings for papers does not matter. It's not what the methodology is about, it serves no purpose whatsoever.

For Interviewees who were new to the REF (Interviewee 9) or joining a merged sub-panel (Interviewee 12), working out the criteria on which scores would be given and then calibrated or carrying out calibration on unfamiliar papers was a challenge:

Interviewee 9.

[pre-meetings] I think it wasn't entirely clear, probably because I wasn't completely immersed in the process by that point. It wasn't entirely clear that what we were doing was working out the criteria, and it wasn't clear how much was up for grabs, I think, or what kinds of comments we could usefully make. It felt like we had some input, but not very much, really.

Interviewee 12.

[calibration] was critical, really. I think the question is more to do with 'how do you do it most effectively?' I think there was a difficulty in my sub-panel, in that it was formed from the merger of [several] previous sub-panels, and so you've got very different academic subjects. So, you were being asked to calibrate papers from completely different areas that you didn't know anything about, but, you know, this was the grand idea of REF, that, of course, "If you have fewer sub-panels, it must mean that we will get much more consistent scoring." So, the calibration exercise is meant to somehow promote that.

### ***D.3 METRICS***

Within these discussions of the assessment practices characteristic of the REF, one area that seemed to produce contrasting viewpoints was the possibility of using metrics. As this has been an issue under discussion for some time (through the Metric Tide report and also to an extent through the Stern Review of the REF), perhaps the level and detail of discussion was unsurprising. What was clear, however, was that viewpoints on the value of metrics were divided. For example, Interviewee 23 was clear in their support of a metric REF, stating simply:

Interviewee 23.

I think you could replace it with a metric system.

Other Interviewees (1 and 2) suggested that metrics may assist the REF (and, for example, citations are used by some sub-panels) but could not provide everything that the assessment system required:

Interviewee 1.

There was pretty good research done at the time, showing the metrics wouldn't do the job. Certainly metrics alone wouldn't do the job. And that's why I think they came up with a policy that was no different from how it was done before, but it was more formalized. And that was that metrics could inform the peer review but the peer review would remain at the core of the process.

Interviewee 2.

some of these measures, some of these bibliometrics tell you things that are useful, you just have to be careful how you interpret them. And they generally only tell you something about one aspect of the research in a way. Whereas excellence we know is a multi-faceted thing.

For Interviewees who sat on sub-panels that used metrics, again views on their value were mixed. For example, Interviewee 3 found the citation index mostly unhelpful, except as a guide for bringing discrepancies to attention:

Interviewee 3.

we were supplied with the citation index. So, I didn't find that that helpful. [...] And would read the paper quite carefully then if it was high-citation low-impact journal. Okay? The other possibility is something in a very high impact journal that is hardly cited. And it's been published for 5 years. [...] But I'm only looking at those very specific cases, really. In general, what I know is that if you published something in a decent international journal it will be cited 20 times in it's whole lifetime. And I think there was very... in that range there's very little discrimination.

[...] I didn't use the journal impact factor, but I know which journals have a strong reviewing process. So how can you avoid that [laughing]? Because we're all submitting to these journals. So my opinion is if someone's got through the reviewing process of Journal X, then I know that paper is a good quality paper. And there is nothing I can do to wipe that information from my mind [laughing].

This view was somewhat contrasted by Interviewee 6 who used metrics and found them useful, but would also like to see future changes in the REF toward more, and more sophisticated usage of, metrics:

Interviewee 6.

I think the next REF will be quite different, certainly in [Discipline X]. I'm sure that it will be metrically driven for outputs, and that will be a very good thing. [...] Most of environment can be metrically driven, and I think [they] should concentrate on reviewing the impacts, and that needed more work. [...] I think that if we had scored the publications, the outputs, metrically, it would have made no difference to the ranking, so a very large amount of work could have been avoided.

[...] In the STEM subjects, where a more metrically driven approach would be, I think, both feasible and acceptable, you could run a cheaper operation. The key thing is, I believe, that all contracted staff should have to enter the REF. So, I think that this business of doing limited submissions in order to boost where you sit is misleading. So, I would make it a level playing field and you'd just have to submit all employees.

Impact would still require peer review- [...] because you can't do it metrically. The metrics are subjective. [...]

Although a broadly pro-metric stance, Interviewee 6 clearly felt that more metrics could have been useful. In the following excerpt, Interviewee 7 explains how one sub-panel had already developed and used its own metrics:

Interviewee 7.

we wrote all our own software. [...] I didn't write it, but we wrote it. The community wrote it. [...] It's much better than other people did. It was a much better method, and everybody should use it, because- and it's based upon this understanding that what we've got is individual scorers, each producing unaligned distributions, and actually you need to normalise. This is all about generating large numbers. [...] could we achieve the ostensible purpose of the REF by using metrics? Yes. [...] In other words, we can reproduce, within reasonable bounds of accuracy, the results of the REF using a relatively small basket of metrics

[...] Okay. If the purpose of the REF is just to distribute QR, that's one thing. But it's not just to distribute QR. Actually, it's about other mechanisms, other policy-related goals. It is about developing a community dialogue around quality. It is about drawing attention of researchers to the primacy which policy attaches to certain classes of impact. It is about driving up the global position of the UK, with respect to its competitors, against certain key measures. It's about transparency and auditability, and consensus, within the community. And for those purposes, metrics will not suffice. Then there is the much more complicated question about, could we do a better job of metric usage in the REF? Of course we could. It was lamentable, it has been a lamentable, the use of data has been lamentable for some time.

For Interviewee 7, then, metrics could be used to more or less accurately produce the results on which QR money could be distributed. However, if the broader purposes of the REF are to stimulate certain kinds of activity (for example, discussions of quality, impact, excellence),

then metrics alone might not achieve these goals. This measured caution toward the REF, its goals and the role of metrics was also reflected upon by other Interviewees. Interviewee 33, for example, is sceptical about bibliometrics (particularly journal impact factor), Interviewee 30 is cautious about metrics being too rule-based and Interviewee 29 suggests that metrics can offer a good first indicator of the value of an output, but should not determine its result:

Interviewee 33.

I'm sceptical about whether you can use bibliometrics. There are certain places which have bibliometrics and they're quite good. Most bibliometrics are not to do with the journal article itself, they're to do with the impact factor of the journal. [...] Specifically banned, but that's because academics don't like it. It's about a measurement of the publisher not a measure of the author. [...] There are certain disciplines which are subject to orthodoxies. In those areas, bibliometrics and even the journal in which you publish, are classed as being, if you're in the excellent journal of something or other, then this is going to be a four star paper.

I completely think this is wrong. Yes. I think those subject areas should be opened up and looked at in great detail, because I suspect that they've closed down academic discussion and debate because of orthodoxy. I'm very suspicious of bibliometrics and metrics of that sort around those kinds of things.

Interviewee 30.

We were not allowed to bring up bibliometric data in connection with making the case for a grading. So, there was not discussion of bibliometrics at the panels, in relation to specific papers. [...] I think many of us actually did look at that number, but then tried to put it in context. [...]

You know, the problem with a rule, you just give a paper which has been accepted in a major journal, a top grading, and low grading to another one, is accepted for publication in a journal is pass/fail, isn't it? There would be huge variability within articles in a journal. Then, there's a phenomenon of... I know some of the most influential journals or articles in my field have been published in, frankly, poor journals, because the author has wanted to publish, as a whole, for example, an article which exceeds the length limits of major journals, or wants to get it out immediately or something.

[...] You know, sometimes you will have a paper which is so good, it just closes down discussion because there's nothing more to say. So there won't be so much follow-up, but the author has achieved something fantastic, and this is credited by his colleagues.

Interviewee 29.

We do have fairly comprehensive data available. We didn't formally use much of it, because that wasn't in the rules, procedures, but as a community we have large amounts of this information. To imagine that there is no correlation between those bibliometrics and the quality and impact of a paper is odd, it seems to me. There is, but of course it varies from community to community.

Within [Discipline X] there are high-quality areas, within which there are a small number of practitioners and so the actual magnitude of citation is very different than those who work in areas where there's currently large amounts of activity and interest. I think it can't be a complete substitute, but it can very often give you a first approximation. [...] we have to be careful, as I say, that the tail isn't wagging the dog.

This cautionary approach to metrics continued with other Interviewees expressing a variety of viewpoints that either gently promoted or rebuked the future use of metrics in the REF. Interviewees 32 and 28 suggest that metrics could have value but citation metrics, for example, could be misleading:

Interviewee 32.

when we talked about metrics and using some of these other instruments, almost unanimously the subpanels felt that that was underestimating the quality assessment. In fact, if you look at journals, the way journals generate their impact factors and the way papers generate their citations is a populist approach, really, to research. You can have paper very, very highly cited because it's awful, you know, and that's why it has been very, very highly cited, not because it's really excellent. I had several examples of that in the exercise.

Interviewee 28.

I am less opposed to metrics, done properly, than many colleagues. [...] First of all, it is heart-breaking when you look at the citations and you realise that so much work that so many people have put so much time and effort into is never, ever, ever cited again, that nobody makes any use of it. That's kind of heart-breaking, really. [...] I find that sad. Of course, we all know that what people cite is partly a matter of fashion; one cite can lead to another cite that can lead to another cite. We also know that there is corrupt practice, or quasi-corrupt practice. There are citation rings, formal citation rings, or informal: "I'll cite you if you cite me" kind of things (Laughter). We know that those things happen.

On the other hand, if one wants an indication of whether or not a piece of writing has had implications for the community, the people it is trying to address, then citations are one way – one way – of thinking about it.

[...] I think citations, done sensitively, being aware of timelines, being aware of fashion, not just treating the number, the bald number as the only thing that you think about, they are helpful; they are helpful guides. I don't think it was very sensible for

panels' members to be advised not to look at citation indices, as they were in some panels. How they interpreted them, that's another question.

[...] I think metrics – some careful metrics as a guide – are maybe not the best, but they're the least worst. I still think there should be a peer-review process, but I think the peer-review process can be guided by metrics.

Alongside this cautionary approach to metrics, other Interviewees were keen to articulate a strongly negative stance. Interviewee 9 suggests metrics are 'completely unreliable' and Interviewee 25 suggests they are 'utter nonsense':

Interviewee 9.

I think metrics are completely unreliable, and the times they've tried to look at it for [Discipline X], it's very distorting. It would have a terrible effect on publications, and just no. Citation is not a reliable method of calculation of anything in [Discipline X]. It just isn't, and it shouldn't be. I can't begin to think how horrible the effects would be on the field. You know, if you're doing somebody a favour every time you cite them, it's bad enough as it is.

Interviewee 25.

the metrics, which a lot of my colleagues are very keen on, which I convinced myself very early on was utter nonsense. [...] Well before the panel was appointed, I knew this was going to be an issue, so I did a little dry-run exercise myself. I took 20 outputs from the RAE and marked them myself. Then, I looked at the metrics. I was astonished. [...] They were astonishingly different. Not always, of course. Quite often, there was agreement, but every now and again, and quite often... Of those 15 to 20 outputs I looked at, I would guess that there were probably four or five where there were enormous discrepancies in the citation count between them. They were discrepancies which were not obvious.

Other Interviewees (10 and 18) were against metrics, but in relation to certain subjects or forms of publication, or because they would limit the chances of junior colleagues (Interviewee 11):

Interviewee 10.

What about things not published in journals? You don't have that measure at all. If somebody publishes a monograph? You could use citations, but it may only have importance to a small discipline... Something might be quoted by every single person working in [Discipline X], but that may not be many people. So it may be a really fantastically important piece of work, but nobody's going to cite it. There may be just

100 people in the field, so the most anything would ever get is 100 compared to something else.

Interviewee 18.

I think it's really hard for bibliometrics, and especially metrics, to be accurate when you have less than [X]% of the outputs [that are] articles. Less than [X]% of what we assessed was articles. So it's really hard for that to be an accurate way of measuring the research quality.

[...] I think it is very particular to a certain number of disciplines, of which we are not one. [...] Peer review should stay, I think it's really important. Informed peer review. Because it allows for discussion; it allows for... I think REF is very strict about conflicts of interest, and articulating and making clear that it's a conflict of interest, and so the peer review is- Every system has flaws, but I think it is as good as we're going to get, to have a select number of individuals who have a track record of undertaking informed work in this area, coming together to offer expert peer review on the materials submitted.

Interviewee 11.

so immediately you have got an issue about how something that is published in the last couple of those years, you know it hasn't had as long to develop a profile. If you look at journals and they do various metrics and so on, yes there are some articles that do have an impact straight away, make their presence felt straightaway. But I would be surprised, I haven't done any analysis, but I would be surprised if that wasn't biased towards established scholars.

[...] So I think that would be one worry about citations, is that that might disadvantage newer entrants to a discipline [...]. And if institutions are making those sorts of calculations, then I think the last thing we need is institutions being even more restrictive about who to include at the start of their careers.

[...] I just think if we encourage institutions into a risk avoidance strategy and a restriction of who the earlier career people are, then we risk going right back to your very first question about what's the point of the method. You know we risk discouraging innovation and novelty and fresh thinking.

#### ***D.4 IMPACT***

As a new feature of research assessment in the 2014 REF, impact was a topic that many Interviewees commented on. For several of the Interviewees, impact had been a broadly positive development. For Interviewee 24 it is about demonstrating results beyond academia

and for Interviewee 32 it was an opportunity to assess and make available ‘amazing’ impact cases:

Interviewee 24.

I think it’s important to know that your research, after a few years, has produced some results on the market place. That’s why they bring in industrialists to judge that, but what they’re judging is impact of research that was done outside the REF period. [...] so they’re looking at long-term impact, so really you did some research a long time ago, has it been used? That’s what they were expecting to see, so yes, I think it’s a good measure. Whether you’ve influenced the market, whether you’ve influenced the policy. You’ve persuaded the minister to change his mind about something, that’s the kind of information that they were after.

Interviewee 32.

I suppose that was one of the most wonderful things that everybody thought about the REF. I mean, the panellists loved it, the people on the panels. Of course, we had external assessors who were outside of academia and come in and join the panels to help them in assessing the process. They loved it. In fact, it was almost the highlight of the whole exercise, the impact. People, again, thought this was going to be difficult and a bit of drudgery, but the opposite turned out to be the case. Suddenly, people realised how amazing some of these case studies were that they were looking at across the world and in this country, or whatever. I think that particular set of processes, getting the impact assessment into the REF again transformed the mood music of the whole exercise, really.

One of the most frequently articulated positive aspects of impact, was the drawing together of a set of impact cases that could become a resource (also see E.6 on Valuing Research) as Interviewee 2 and 29 make clear:

Interviewee 2.

there is now a source of evidence that allows universities or individual researcher to make decisions about how they carry out their research and how they take that research to deliver benefit that wasn’t available before. And so I think that’s a strategic value. [...] I describe this as a national asset; clearly anyone in the world can look at it and draw those inferences. [...] And there is a certain amount of rhetoric in that, you know, I wouldn’t pretend that there wasn’t rhetoric there to actually send a message to government that says ‘This thing, there is a future value’. So I would hope one might end up using that rhetorical device to achieve that.

Interviewee 29.

we had a quite interesting formulation of impact. I think it's been a very helpful feature for the community. Although it's had to adapt, we now have a lot more information about where innovation from research is happening; we have a lot more information about the engagement of industrial and business organisations in the research, the impact it's had.

I think those impact statements have become a really valuable resource for the community, in lots of ways. I know that impact case studies are still used as reference models for departments to talk about what they're doing. Yes, so I think that turned out to be a really important feature.

I think there is an issue around the rules about what's admissible. You weren't able to submit impact case studies that applied to the higher education system itself, which is perverse because it's an economic system; you have impact in it. Why shouldn't you be able to claim credit for systems which transform the way in which the sector itself works?

Following on from the final aspect of the preceding excerpt, Interviewees often combined positive views on impact with notes of caution. Interviewee 25, for example, suggests he is not 'an unqualified fan' of impact, but also recognises that impact was not a disaster:

Interviewee 25.

It was 'Impact' that was the new thing that people were getting very upset about. There were lots of people who were writing letters to 'The Times' and all sorts of things about how this was completely going to ruin blue skies research. It was all complete nonsense, of course.

Not that I'm an unqualified fan of impact assessment, but, even at the time, it was obvious that it wasn't going to be the disaster that people were saying it was. Some people have no sense of proportion. There was a big discussion and debate going on about 'Impact'. That is the thing that's mostly in my mind about that period - how we were going to do 'Impact' and what it was all going to be about.

These mixed views on impact were also reflected in approaches to assessing the impact cases. The following Interviewees (3, 44, and 25) suggest that impact cases were relatively easy to assess, and yet problems also arose in the assessment. For Interviewee 3 the level of evidence was 'pretty rubbish', for Interviewee 44 the cases were 'very varied' and Interviewee 25 suggests there were problems with how cases had been written:

Interviewee 3.

partly because of the subject matter, I think the impact was a little easier for us. [Experts in this discipline] are used to assembling that panel of evidence. [...] actually, one of the weird things about impact was that the form where people capture their impact, the evidence level on there was much lower than [is typical in Discipline X]. So we found the evidence level pretty rubbish. [...] at the panel we weren't allowed to look up any additional information. We're supposed to look at this impact statement. [...] So I think there is some discipline-specific ease with which you can judge impact.

Interviewee 44.

[impact cases] They were very varied, very varied. So some of them read like they'd hired a PR manager to come and write a brilliant advertising slogan for their [...] output. Some of them were really scraping the barrel of trying to turn something that wasn't really an impact story into an impact story. A few of them were just stunning, really great examples of how [Discipline X], over a long period of time, through well managed and integrated relationships between academics and industry or academics and government makes a difference in the real world. There were a few really great ones. [...] If you could only see it [case study] approaching it [impact] then actually you were supposed to give a zero mark. We did end up with a few that got that because they actually hadn't had the impact they were just doing applied research.

Interviewee 25.

None of us had really thought about how we were going to do the measurement of impact. When you were presented with the case studies, it was quite interesting that several people assessed them all and there was a broad range of agreement.

We were worried that everybody would have different ideas and that there would be differences of assessment, but there weren't. That isn't to say that the whole thing was straightforward. There were lots of problems.

People didn't know how to write case studies and people didn't put the right information in, but that was what you expected, of course.

The relative ease with which the preceding Interviewees assessed impact cases, stands in contrast to Interviewees 14, 20 and 8 who each suggest that impact cases were difficult to score:

Interviewee 14.

we graded them all of course. That was the hardest part because we know what makes a good academic article, we don't know how to prove impact. The guidelines I found - this is anonymous, right? - I found them to be a bit rigid to tell you the truth, because

the standard of proof that was required is not really a standard of proof that normally applies to the influence the research has on the public debate. [...] The other thing is that, sometimes, impact, it would have been better if it didn't exist because poor-quality research can have more impact (Laughter) how can you judge that? It had a lot of impact but it might have been better if it didn't.

Interviewee 20.

We did have one big problem. [...] We had, I think, six submissions where the impact scores were 3.5 or above the mean impact scores. I was sent back from the main panel to say, "How have you got six submissions that have got these really high scoring scores? We want you to relook at this." Unfortunately one of them was mine as well, my own institution.

What I did was, I asked four panel members, they were panel members who weren't involved in submissions. They were institutions that aren't returnable to the REF. I took four or five of those people, and the deputy chair whose university wasn't involved in it, and I asked them to score all of the impact case studies independently: I was not involved at all. They came back with, pretty much, the same score.

[...] It was difficult because my subpanel were not happy about being challenged on their scores, having previously been challenged that they were too hard. They were then being challenged that they were too soft at the top end. They weren't happy about it but carried out a re-review as requested.

Interviewee 8.

I think people did struggle with [scores] despite the guidance and what have you. So there were issues, you know, the difference between dissemination and impact, there were lots. I think some of the rules as well, because impact stays with the organisation and doesn't travel, so what I tended to find was you could tell when a case study had been written by somebody else. So somebody had obviously died or they'd moved on and the responsibility for demonstrating impact fell on somebody who actually didn't really know very much about it.

I can see there's a bit of unfairness there because I think there are some things that you think, "That must have had really quite good impact," but they haven't been able to demonstrate that in the narrative. [...]

Some things were dubious and there was a lot of over claiming I would have thought. Well, I say a lot, but you could tell there was over claiming. Another thing I thought was difficult was that thing where if you're the only one doing research in an area and something happens, that's easy isn't it? If there are lots of people operating in an area, how you demonstrate your unique contribution, so I think that is a very difficult thing, I'm not sure that was ever properly resolved.

So somebody who has an impact because they've been pushing on an open door, is that one, two, three or four because it wasn't that difficult. So lots of people already

did the work or whatever, or it's an area where you were basically invited to come in and do something as opposed to somebody making impact in an area where the door has been shut for a long time. Is one better impact than the other? I'm not really sure, I don't know if we answered that.

These more negative views of impact were expanded upon by other Interviewees. Interviewee 13 suggests that impact 'stinks', Interviewee 9 was among several Interviewees who felt that the impact template was a problem and Interviewees 10 and 42 suggest that both the production and assessment of impact cases was an issue:

Interviewee 13.

[impact is] a problem for two reasons. One is that case documents are unevidenced, but you can't take that into account. You can't say, "I'm going to go poking around," you have to take what comes.

It's unevidenced, in that sense, which is not the same with a publication because you've got the publication.

That's one reason. Looking at the other reason, it's a single measure of what counts as impact. Whereas, on the ground, different units of assessment, different research groups or individuals within that, and also different kinds of universities, have adopted very different stances to how they want to do impact in their local situation.

[...] It cannot be measured adequately, appropriately or acceptably to me, unless it includes education itself. A measure of the impact of the work that we do that rules out of consideration academic impact is a problem. [...] What do academics do? We teach people. That is the greatest impact that we will ever have.

Interviewee 9.

The [impact] template was diabolical to assess. It was really formulaic.

Interviewee 10.

But one size fits all... It doesn't work [...] There are some financial implications to it. I think it's very hard, and I think people outside the [Discipline X's] particularly, feel that they're being asked to do something that doesn't fit.

Interviewee 42.

I thought the panel was very cautious; there was a lot of questioning about how we were going to do this, and people thought it was ridiculous, and the conclusion we came to is, "Well, we'll just do it. Do it as it comes along." We don't know how we're

going to do it, we can't write to you how you're going to do it, but we'll do it, and anyway. So, I sort of thought the same thing, really. I thought, "Well, it'll sort itself out. I'm not going to get anxious about it."

And then, to my surprise, I think the whole panel thought it was much more interesting [...] there was a lot of enthusiasm for the content, which turned into more enthusiasm for the process, I think

As Interviewee 42 suggests, this sense that the assessment of impact cases would be a problem, dissipated over time. This was a view shared also by Interviewee 27:

Interviewee 27.

the impact work [...] assumed a kind of a looming presence which at first was very alarming [...] So, it felt quite stressful, I would say, for about a year, not least because we didn't have any actual stuff; we didn't have the things to look at, which, as you know, when something is so both abstract and difficult, it sort of assumes a kind of looming presence. Whereas, when you're actually looking at what some poor soul like ourselves has actually written and you read about their work, and about their relationships, and all the absolutely fascinating things that they've done, and changes that they've made and so on, then suddenly it becomes a personal experience.

You feel a kind of empathy with that person, whether you happen to... Often you know them, of course, and you know their work, you know their department, so you know what they're doing and what they're putting over. So, you then get, as I say, that sort of empathy. The problems, though they still existed, they started to be diverted into: "What are we going to think about assessing this particular piece of work?" rather than setting up endless abstract criteria.

Many of the Interviewees noted that the impact criteria were a crucial matter of concern. For Interviewee 5, discussion centred around the general criteria of reach and significance, whereas for Interviewee 39 it was the specific hurdles required to avoid a zero and for Interviewee 21 it was how to fulfil and assess the criteria of changing others' behaviour:

Interviewee 5.

We were quite open to the idea of impact through affecting public life, through public discourse and through affecting politics. [...] Of course, we had these two criteria of reach and significance that we had to apply. There was some discussion about how we should really think of the reach one. It wasn't restricted to geographical reach, but there might be importance in terms of the domain.

Interviewee 39.

I think the REF people had done a very good job of trying to make it clear what all the rules were. You had to go over various hurdles to avoid getting a zero. That was all laid out very clearly. [...] the criteria. It's quite hard to be very precise. There was a lot of judgement involved. I think, with the outputs, it was clearer, in some sense, what we were looking for. [...] I think it was okay. I didn't feel it was a bad- I think we did a reasonable job, and the rules were reasonably clear.

#### Interviewee 21.

The further thing which I know some of my colleagues felt very strongly about was the requirement that, in order to demonstrate impact you had to demonstrate that in some sense people had changed their behaviour as a result [...] And one colleague in particular said, "Look, public engagement is what we do." Demonstrating that people who come to your lecture change their behaviour is much too high a bar. What they do, is see the world differently after that. Could they describe where their world view came out of? No, because it's made up of a little bit here, a little bit there, and a little bit somewhere else. Hence, how do you produce evidence for this? You can't produce evidence for this. All you can do is to say, "Look, actually I kept 100 people riveted for an hour."

It's not that one doesn't appreciate there's a problem here. It's impossible in terms of producing that evidence, to show that they were riveted for an hour rather than asleep for an hour (Laughter). All you can do is list that you gave the lecture. But nevertheless that means that, again, a lot of impact is being missed because it doesn't exist in this way where someone can actually get someone to produce a testimony to say that, "As a result of this I went and did whatever."

Other interviewees discussed the problematic criteria involving number of impact cases required (Interviewee 42), the need to come to agreement on the criteria (Interviewee 20), how to balance being open and overly prescriptive with criteria (Interviewee 36) and issues with the scoring system (Interviewee 28):

#### Interviewee 42.

I spoke up quite a lot about this and made myself unpopular, because these big research institutes who might have 200 people, they were going on and on. They were saying, "It's not fair that they have to produce, every 10% of people have to do it. That means that if there's 200 I have to have 20 case studies. How can you expect us to have 20 case studies?" And I said to them, "Well, you've got 200 people." If I've got a submission of 20 people, I've got to do 1 or 2, 2 case studies. It's the same ratio, but they didn't get it, and even now they're lobbying, I'm sure, that it's not fair that if you're a big research institute you still have to give the same number of case studies. So, there was a bit of bad feeling about the number of case studies in the large institutions.

Interviewee 20.

because I'd been leading my institution and had been working with colleagues on impact case studies, and had done some work with editors as well, I was quite surprised at how raw the understanding of the panel members was about what constituted a good impact case study. We had to do quite a lot work to fit our scoring with the published criteria. People's interpretation of the published criteria was quite broad, quite wide. So we had to come to agreement about what it was we were looking for.

Interviewee 36.

Originally, in our main Panel [X], we had decided to be fairly vague you might say about how much guidance we would give, in terms of what we thought impact would look like. [...] Maybe because we found it difficult but I think genuinely, and this is what we said, we didn't want to be too prescriptive and close down potential examples of impact. But what came back from the consultation was that this was no good at all because we weren't really giving them enough guidance on what to do. At that point then, we realised we had to give them more by way of examples or illustrations. If you look at the criteria and the guidance documents, they still try to stress that we're not being prescriptive.

Interviewee 28.

I understand the desire to take account of the impact of research, but, as you know, there are lots of ongoing reviews of how one might do that impact work. I don't really think that the criteria were particularly helpful for doing that impact work [...] I think to try and assess impact and environment using the same one-, two-, three-, four-star system and trying to put a profile to it was not very satisfactory, in my view. The panel struggled; the panel I was on struggled to do its best, but I don't think it was a terribly satisfactory system.

An issue for these Interviewees in articulating concerns about impact assessment were the guidelines offered on how to produce an impact case. As Interviewee 33 suggests better methods are required to ensure that future assessments are fair:

Interviewee 33.

Next time, if I was doing it again, I would be trying to say some slightly different things in the guidance to institutions about how they should present [impact]. I would probably think of some slightly better methods for how to ensure that the assessment was as fair as possible. [...] There was this basic thing in the exercise as a whole that

was to do with the reach and the significance of the impact. What exactly do these mean, is something which there is a lot of discussion around. One of the key things that some institutions, I think, got wrong, including, probably, my own one here, was to try to claim impacts that very large, very wide in terms of reach and significance. When, actually, it may be easier to evidence smaller impacts that have less reach and less significance. But, because you can evidence them well, then it's easier for a panel to say, "This really was that impact of that research, this whole chain of where the impact comes from." Yes. They have to give it a good mark.

The problem of not providing clear guidance on impact was matched by the problem of having to assess unclear impact cases. As Interviewee 11 suggests, assessing an impact case was similar to marking an essay:

Interviewee 11.

I suppose the challenge was, it's a bit like when you are marking an essay. Sometimes you can see what someone is trying to say, but they haven't quite said it and do you say, "Well I'd better give this essay a very good mark because potentially it was very good." Or are you saying, "Well I have got to mark what is here, rather than the sorts of things that you haven't quite drawn together."

[...] the users were very open to the full range of ways in which Impact had. So it was a pleasant surprise. I was expecting that there would be disagreements and is it a three or is it a four or is it a two and that sort of thing. Whereas actually I think the users were very open.

When you have got a group of academics there is always a concern that non-academics will not necessarily see the value of blue skies thinking or the value of things being interesting in an academic kind of way.

As Interviewee 11 suggests at the end of the preceding excerpt, the impact assessors were a key group that academic assessors had to work with. Experiences of working with the impact assessors appeared mixed. Interviewee 6 suggests they 'didn't contribute much,' while Interviewee 21 thought they were slightly less integrated into the REF, Interviewee 39 felt they were good people who perhaps took a while to acclimatise to the REF, while Interviewee 25 suggests that one of the assessors did not turn up to meetings and was late with assessments:

Interviewee 6.

I worked with different end-user assessors. I worked altogether with three, because of the way that they were divided up. One was from industry. One was retired from [field-specific industry]. They had a particular view of the world, and they tended to be more generous scorers, actually, than the academics, and less discriminatory. It

was interesting, having them there. They didn't contribute much. [...] They tended to think it was all wonderful.

Interviewee 21.

I don't think there was a very significant difference in handling the users from handling the subject experts. The fact that the users were there only for the discussion of impact meant that they were less well integrated into the whole process than, obviously, the main sub-panel members. That wasn't, I think, in the case of our panel, a serious problem. They quickly realised what the nature of the task was. They picked up well on the way the discussion was done within the sub-panel, I think.

Interviewee 39.

Yes, so they were very good people. Very good people, but they were busy. As I recall, I don't think all of them came to the calibration [...] But I think coming from outside into this thing, just dipping in, as I said, it does take a while to get a feeling for the level. And calibration is very important [...] the users played a big role in coming up with the numbers. They were very important.

Interviewee 25.

Now, when we got down to the impact assessors who weren't members of the Main Panel, then there were one or two quite problematic ones.

One, because he was really much too busy to do the job properly. Although, in the end, he did quite a good job, he would miss meetings and he would be late with all of his assessments. It was a real pain. [...] The rest of them were all fine. When it came to the assessment, [Expert X] who, again, is an old friend of mine... You're getting the impression that all these people are all friends of mine. Not all of them were, but some of them were. I didn't choose these people. He tended to have a somewhat peculiar view, definitely in the wings of the distribution. Often, we had to renormalise him. He was maybe not quite so useful as he might have been because he tended to take a strange view of things.

This theme of very busy impact assessors was continued by Interviewee 20 who suggests that the scores of one impact assessor could not be used:

Interviewee 20.

There was one Impact assessor who, for one reason or another, just never got to any meetings. And I think that's one of the difficulties with industry people especially, that their business comes first. [...] And, in the end, we didn't use his scores, because

there were very few scores, so he was able to complete and because he hadn't done the calibration process

Other Interviewees (27 and 42) felt that the entry of the impact assessors was a meeting of different cultures, with negative and positive outcomes:

Interviewee 27.

the thing that was really different about the REF, as you know, was bringing in the so-called 'industry assessors'. [...] There was a sort of implication that users know what they're using and why they're using it, but it was never fully explained, I don't think. They certainly felt that and so that was, I think – in my view, anyway; I don't know what others have said – that was the most difficult part of the whole REF process, was this sort of meeting of quite distinct cultures which thought they were both doing the same thing in some way (Laughter).

Interviewee 42.

it was slightly different when you've got the users in the room, because it's not just a load of academics talking to each other, and so you had to be very respectful of them, and they sometimes had quite a lot to say.

And actually, it was valuable, because some of those users had actually turned themselves into impact case study experts, and been tramping around the universities giving guidance on how to write impact case... And there was one person on our panel, I think, got up some people's noses, because he proclaimed to be a real expert, more than any of us; which he probably was, because he'd been thinking about it for a lot longer than us. So, he was a bit prickly. So, there was a little bit of prickle, but I think they were necessary because they kept us grounded.

The contrast in positive and negative experiences with impact assessors can be noted by drawing together the views of Interviewee 7 ('a failure') and Interviewee 36 ('worked quite well'):

Interviewee 7.

So, we had insufficient real end-user assessors, and I thought they were pretty useless, frankly. You know, if you want people, you need... I wasn't impressed, let's put it that way. I really wasn't impressed with the external assessors, you know, on balance. Quite rightly, the people you really need to make those assessments, you know, industry was not prepared to let those people spend that amount of time, nor were they ever in the domain to do that. So I think that was a failure, certainly in our area.

#### Interviewee 36.

I think that, as a process, it worked quite well. We had great assessors. Our impact assessors that we had, again, all took it very, very seriously and were excellent in the discussions. They brought their experience to the table. I think, also, for them, the discussion that they had with us meant that they had a better idea by the end of what impact might mean to academics in the same way that we learnt what impact might mean to users.

[...] I think, as a process, it worked well. I think the case study format worked pretty well. We felt there was a degree of over-claiming, but maybe that's inevitable, really.

For the impact assessors, being part of the REF and assessing impact cases raised many similar issues to those already articulated by academic Interviewees. For example, Interviewees 16, 17 and 19 all express a concern with providing correct scores, while Interviewee 17 also suggests providing clearer advice to universities, and Interviewee 19 discusses relative scoring:

#### Interviewee 16.

There was a concern that we might be marking people more strictly than other panels. So I didn't want to be in a position where I was, purely on the basis of my own subjective views, helping to achieve a result which would mark down a sector of the academic world which I actually supported, relative to other academic sectors. So once I'd realised that my marks were consistently low I just increased them all, but without changing the relative judgments on particular institutions.

#### Interviewee 17.

if they [scores] weren't within half a point we'd email back and forth to see if we could find a way of resolving it before we were on site. If we couldn't resolve it before we were on site, we did it face to face in negotiation around the table. [...] to me, Impact is taking your scholarship to the people, which actually I do think is a public good as public employees at universities in this country are. That's a good thing. [...] if anything, I would suggest that there would be some more concrete feedback from the ground, so the universities were a bit clearer about how they'd construct their case-studies.

#### Interviewee 19.

it was all quite new and in particular because impact was the area I'd been called in for, it seemed quite obvious that no one really quite knew what to expect from that,

and that remained true to a certain extent, there is a way in which everyone was feeling their way through it from the beginning [...]

when we looked at the overall scores. I mean, I suppose – it's a good job this is an anonymous thing – there was a certain point where I had to make, at some point, what might seem as reasonably arbitrary adjustments. But I mean, in a sense, I don't think that's necessarily a bad thing. I mean, the fact is, the numbers are kind of meaningless, it is relative; what's important is the relative scoring, not the absolute numbers, so if you have to adjust your numbers, that's fine.

[...] there were issues certainly around what kind of impact you could expect, because I think in the end we took the view that you assess the impact on the basis of what kind of impact it is reasonable to expect, do you see what I mean? [...] It's got to be relative to what you're talking about. [...] if it's as big an impact as you can possibly imagine from that, that's top marks.

One means to draw together this variety of concerns with impact is to contrast the experiences of two academics who had their impact assessed. What we can note in the following are the very different experiences of the two Interviewees:

Interviewee 31.

[submitting impact cases] a rational process where you don't really know the rationality of your parameters or the decision-making. [...] It's a complete exercise in guessing. [...] It's complete guessing. I remember I was sat on so many reviews, and externals would come and of course we would all invest tremendous hope in the external having a superior capacity to guess than us.

Often the externals had been on REF panels. But you're - my abiding memory of that was how much, even the people who had sat in the room who had made the decisions in the RAE or the REF, they were still guessing too. They were still guessing in terms of "Okay this is a submission that hasn't been made yet, it hasn't gone through that process, which one should we choose and why?"

I think about my own submissions last time, and I deferred to the judgment of somebody else who I, at the time, thought would know better than I do, and on reflection now I think that was wrong actually. That was - their guess was wrong. It was very flattering in a way, because they valued things that I wasn't sure about actually, that were probably a bit more risky and stretching the parameters of what a [Discipline X] submission to a REF could and should be, but if I could have my time again I would have taken back some of the things that I submitted last time and put other things in their place

Interviewee 6.

So, obviously, I was responsible for doing the impact for [University X], so I had worked extremely hard on the impact side. We'd done the pilot impact in 2012. We were one of the universities that took part in the pilot [...] which was absolutely critical to understand the process, and really helpful. [...] So, when I did the pilot, we put, I think, 20 impacts into the pilot, and I deliberately chose a range in order to test the system, so it included all sorts of different impacts. The feedback was extremely helpful, because it was clear that HEFCE panels were going to take a very conservative view of what impact was.

[...] So, the most important thing that I did in the REF 2014 was, when HEFCE asked for universities to identify possible staff to be seconded to HEFCE, I seconded my research administrator. She worked with HEFCE on drafting impact, the text. So, I was able to see that at an early stage, and indeed input into it. So, I was quite involved in the wording of the impact definitions and the text, so I was very happy with it.

## **Summary**

The length of this section of the report reflects the detailed discussions and number of issues aired by Interviewees around the assessment activities involved in the REF. This is not surprising given that the central activity of the REF and the members of Main Panels and sub-panels was to carry out and agree on assessment scores. Drawing some of these issues together, various Interviewees raised problems with peer review. Interviewees discussed how easy or difficult it was to manage the workload of outputs they had to assess, how challenging it was to provide scores and even to resolve a disputed score. Several Interviewees felt the REF created a kind of spurious accuracy by creating fine-grained differences between output scores and others thought the REF imposed various structural limitations (for example, eliminating the possibility to have an oversight of an academic's work). However, it was also clear that alternatives to peer review were considered equally or more problematic.

In moving on to calibration, Interviewees discussed the detailed methods by which distinct Main Panels and sub-panels established what ought to be a score and how a score ought to be achieved. Interviewees suggested that this was not a one-off process, but a continual activity to ensure that scoring was consistent and commensurate. Normalisation offered a further means to impose and reinforce appropriate scoring. What was clear for Interviewees was that the score they gave to an output was less important than the relative positioning they provided of one output in relation to others. Metrics were then discussed by Interviewees, not as an alternative to peer review or calibration, but as a potential aide in the scoring process. Views were divided as to whether metrics had a great deal, some or no value at all to the process of the REF. Several Interviewees pointed out that while citations might offer a limited guide, impact factors were limited. Furthermore, metrics might work against disciplines that were small or mostly book based, might work against junior colleagues just starting out and might make the assessment process even more rigid and rule-based.

Finally in this sub-section, assessing impact was discussed by the Interviewees. Impact was noted positively as contributing to the provision of a resource of case-studies that could promote UK research. However, there was variation in the degree to which academics found

the cases easy or difficult to produce and assess. It was suggested that guidance to universities could be made clear, that criteria could be clarified, and that the definition of impact broadened and the impact template removed. The experiences of non-academic impact assessors also reflected the same sets of issues. We finished this sub-section by contrasting the experiences of two institutions and their impact cases.

## **E. EFFECTS OF THE REF**

In this final section of the report we move on to consider the effects of the REF highlighted by Interviewees. There was no single consequence of the REF that seemed to dominate Interviewee discussion. Instead we have divided this section of the report into six sub-sections that capture the main issues that were introduced. These comprise discussion of: gaming; marketization; shaping effects; careers; transparency, accountability and justice; funding; and valuing research.

### ***E.1 GAMING***

Several Interviews expressed the view that an effect of the REF was to encourage institutions to try and ‘game’ the system. This comprised recruitment strategies, fractional contracts, with people presenting themselves as research stars or 4x4 academics, and the hiring of external assessors. As Interviewee 5 suggests, game playing appears both convenient and disturbing:

Interviewee 5.

It does seem to encourage some strange game playing. A number of people who have a one day a week contract at five institutions, so they can be scored in all of them. People who are brought in as visiting academics, very conveniently just over the census period. That's all a bit disturbing. Some people think that's inevitable, it's just an unfortunate by-product of any systematic way to measure things. It does suggest to me that, to prevent too much game playing, it's necessary to have a look at structure each time and to try and adjust it.

This theme was taken further by Interviewee 25 who suggests institutions with academics on fractional appointments received lower assessments:

Interviewee 25.

The game-playing aspect of it [...]. We saw people in not so much this exercise but the previous exercise who recruited various Nobel Laureates on 20% contracts for, presumably, six months and paid them a lot of money. That sort of thing, as I said, didn't seem to happen this time so much. I don't know whether HEFCE managed to find a rule that stopped it. I'm not sure they did, but, anyway, it didn't seem to happen as much. We kicked back at the groups that did that and more or less told them that they were playing the system and that we weren't stupid. Nevertheless, it didn't seem to happen so much this time.

There's no doubt in my mind that that whole game-playing business, and the fact that people started to need to hire people who were seen to be research stars, somehow broke up the whole rigid rules that used to be here and in many other universities -

particularly here - about (a) what you could pay people, and (b) what their duties ought to be.

The changing practices alluded to by Interviewee 25, also involved the allocation of a great deal of resources by institutions. As Interviewee 25 and 11 suggest, the process for the next REF is well under way and as a sub-panellist, they are in demand and resources are being mobilised to try and attract their attention:

Interviewee 25.

I'll get asked to go and talk to 15 or 16 different universities when we get nearer to the next REF, advising them on what outputs look like. Certainly, people I know, who were on the previous panel but not on this one, were asked to go and look at universities' submissions and see what they thought. The people knew what to expect. They ruthlessly got rid of the things that looked like two-star.

Interviewee 11.

it leads to the allocation of a huge amount of resources on the part of institutions and the preparation and that's starting again. I had an email recently saying would I be a reader of an institution's outputs, as they prepare and rate them one, two, three, four. And I declined because I have got so much else to do. But institutions are prepared to put loads of money into paying external assessors.

[...] You could look at the REF process, or any of these competitive processes, as encouraging fairly short term strategies. [...] But I think the sense must be that, certainly with at least some parts of the REF there was quite a bit of short term game playing. People say that with each iteration of the RAE and then the REF, the rules change to try to discourage game playing. But it just takes a new form, because in some cases you are dealing with subjects that actually have expertise in game theory.

So as long as you have got an economics department you will have somebody who will then be able to advise you on game theory and strategy and all that sort of thing.

[...] You do get some individuals who play that system, so it's not just institutions, it is also people playing institutions. [...] Some people use this expression, "I am a 4 x 4, so I have four, four star publications already in the bag, here on my CV, we will see it here, and therefore you should obviously offer me a huge salary and a very large office and no teaching and no administration and lots of research assistants, because I am a guaranteed 4 x 4." The term is used tongue in cheek. So I think that is something that is fairly short term. The risk is that people strategise for the next REF rather than for the longer term.

As Interviewee 19 suggests, these kinds of strategies and the actions of individual academics can shape the recruitment practices of the Higher Education sector:

Interviewee 19.

I think there is a lot of concern it distorts research priorities, because people are thinking, when they're appointing people, "Will this get us a good impact or not?" Whereas what they should be saying is, "We'll appoint the best people and get the most impact out of their work as we can." I think it would be naïve to think that's actually what's happening.

This combination of game playing by individuals and institutions was reflected on by Interviewee 31 who suggests that the system itself has done nothing to stop, and so by implication has enabled, a variety of 'toxic behaviour':

Interviewee 31.

There are different strategies, yes. They are very different strategy. One strategy is not to submit at all and to get a better ranking, maybe as a consequence you get less money. So there are different ways of gaming the system, which I think is profoundly wrong. [...] the system has allowed that kind of gaming, and as a consequence it's licensed all kinds of toxic behaviour as a result. [...]

After the last RAE where [University Y] did so badly, a friend of mine who works at [University Y], who carried that weight and also the consequences of that very, very personally, we were at a social occasion and I remember saying in the context of the social occasion that the victories in the RAE are as hollow as their defeats. The victories are as hollow as the defeats. I think that - I really, genuinely believe that.

## ***E.2 MARKETISATION***

A distinct effect of the REF that was discussed by Interviewees was the possibility that the assessment system was involved in the marketization of UK Higher Education. That is, the REF was identified as, possibly, contributing to the transformation of universities into competitive, market participants. However, the extent, form and effects of marketization was by no means agreed upon by different Interviewees.

For the following Interviewees (17 and 20) the REF was clearly instrumental in the marketization of Higher Education:

Interviewee 17.

I think we're in a market, whether I like it or not. So, does it adequately create tiers? I'm sure it created tiered universities with structures. [...] I'm probably getting too old

to say this but, you know, my sense is that that is the way it is. I'm not sure the mechanisms of marketisation can- I mean, this government is, and other previous governments, including Blair marketised and modified education. That's where we are. But in terms of having some calibration by peers of recognition that some scholarship is really good, I don't disagree with it, because it was still a peer review. [...] That's the difference. And the peers took it very seriously [...] You know, this was not an easy thing to do but we took it- We did the work, we couldn't do it without, particularly those who read thousands of pages. So, in some ways, the REF is a calibration of government policy.

#### Interviewee 20.

those things are amazing to me, and I'm sure we're moving in to new realms of different sorts of markets. And I know through all the strategies to try to enhance our position, we used maybe differently, but they were used by our competitors. So it's actually really difficult to move yourself within that market. Unless you do something really, really different.

Both Interviewees above seem to identify mixed aspects of marketization – that perhaps peer review was a positive benefit for Higher Education, while the REF also established some constraints. For Interviewee 33, the increasing presence of market-like activity among universities has been a positive change:

#### Interviewee 33.

This university is very dependent on its research reputation as a mechanism to attract good students from around the world, particularly postgraduate. There's a direct link. It is one of the reasons that people in the university so much value the Research Assessment Exercise. Actually, it's because it's valuable far beyond the QR income. Yes? It's valuable in exactly that, in the student marketplace, in PhD students and so forth.

I think it's a really good thing for the UK. The UK has a really good assessment system which is, I think, as good as or better than anywhere else in the world, and actually being copied by everywhere else in the world. That places universities in the UK at an advantage in international terms because people can read the ranking and they believe what they see. [...] The big unknown is actually the political whim. We're subject to a market which can do that [assessment], but we're much more subject to political whim. [...]

It's far more predictable is to work with an economic cycle than it is a political cycle. [...] I hope that's not unorthodox. That's absolutely neoliberal ordinary. For me, actually operating in a market, is actually [...] fairly straightforward. You know what you're doing. We need to please our students. We need to please our research sponsors.

Despite the clear views of the preceding Interviewees, there was not universal agreement on the extent or form of marketization. The following Interviewees (1, 22 and 16) recognised that there was an increase in competition among UK universities, but not necessarily a market:

Interviewee 1.

I don't think that we have markets in research, but competition. I think even when you've got state funding, universities will compete for it. One can envisage competition very easily without markets necessarily. [...] I don't call the REF marketisation, because that is just a mechanism for... it's a distribution mechanism. How otherwise do you decide who should have research money? And the Research Councils do it by assessing proposals. Is that a market mechanism assessing proposals, allocating money according to research proposals? I don't know how you define market, I don't know our definition of markets. [...]

Although I can see marketisation very clearly on the teaching side, I don't see on the research side. [...] The equivalent on the research side would be the government doesn't fund research, universities have to go to industry or wherever to get funds for research and the industry will give them the money if they think it's in their interests to do so. [...] and I don't see we're anywhere near that.

Interviewee 22.

Well, it's not really marketisation, is it? It's an external body. The REF is deciding whether you're any good or not. It's not the market deciding it. But I don't know how you'd do it, really. There isn't a clear way. If they want to give money out for research, then I suppose you're either going to do it on some historic basis, which doesn't seem fair, on grants, or you have to have something like the REF, I suppose. But you can have a better system than the REF, can't we?

Interviewee 16.

[impact assessor] [...] it's not obvious from the exercise that I went through that it's got anything to do with markets, it's quite the opposite. It's a bureaucratic process where people outside the market process sit down and decide on the merits of things so that government funded institutions can then distribute the money. It's the complete opposite of a market process.

### ***E.3 CAREERS***

Interviewees devoted some time to describing the effects of research assessment on academic careers. As we will see in this section of the report, the REF appears consequential for academics' choice of department or university, it can positively shape career chances for women, it has enabled diversity and equality to become a more central focus in recruitment and career development, and to some extent, the REF (and research assessment in general) has become a career in itself.

In the following excerpt, Interviewee 10 captures several of these points. Academics are forced to think about the suitability of their departments for their own research interests, departments have to move to recruit REF submit-able academics, polarising successful and unsuccessful departments, and those at the start of their careers may have little chance to develop their work:

Interviewee 10.

It certainly forces good people out of weak departments to move to other institutions. It's very hard to sustain academic activity if you're really the only top person in your department and most other people are not research active and so on. Universities find it very hard to retain people like that because they feel they get victimised by the, generally, the low performance.

Of course, universities then engage in these kinds of football transfer window activities, which encourage people to move. You see a lack of opportunity for someone who's not already in a strong department to develop. I think it polarises academic life between the haves and the have nots.

[...] I think the problem is it doesn't give an opportunity for people to develop research other than at the very beginning of their careers. [...] Basically, if you are not a star by the time you're 27, that's it, you're finished. I think that's a very poor use of resources. It may disproportionately affect women. I think that hasn't really been looked at because they're often delayed in building a research career. They're not in a position to change jobs while they're having a family and so on. I can see that it's not just the case that the strong people transfer, I can see a good argument for that. I think it's that the people don't get the opportunity to become strong, so we're losing potential research endeavour.

The point made by the preceding Interviewee on the possibility of a disproportionate effect on women, was by no means reflected by all Interviewees. Others suggested that equality and diversity of opportunity had increased as a result of the REF. For example, Interviewees 32 and 20 note a positive value in the changes made in taking into account special circumstances between 2008 and 2014:

Interviewee 32.

The first thing which I also thought was a great thing was the equality and diversity bit, which was much better in 2014 than it was in 2008, and that it benefitted those

who were in some way disadvantaged by having to take time out to have families, or look after their parents [...] I thought that was great, really, really good. Even the equality and diversity panel, they had very, very difficult cases, and I think what people, and what I realised – although I wasn't a member of the panel but we got reports – was people in the most extraordinarily difficult circumstances, terrible pressure, whether it was family or financial or whatever, were still delivering high quality work. It was wonderful. That said a lot about the endeavour here.

Interviewee 20.

'special circumstances' was a major development for REF 2014. [...] So, you know, I think that's a humanising element of it that I was pleased to see. Because, otherwise, it does become this great, big machine.

Furthermore, Interviewees 34 and 13 felt that the development of the RAE and REF had been positive for women in particular:

Interviewee 34.

whatever you tell people, they know that their next promotion will largely depend on their research. How closely your REF performance ties in with that will vary from school to school and university to university, but people are always aware of that. [...] I think research assessment has been very good for women academics, actually, over the years. If I look at the pattern of promotion for women over the course of my career- maybe it would have happened anyway. When I started out, it was almost unheard of for a woman to be a professor.

[...] I think, of course there would have been more promotions of women, because the world has changed, but I think research assessment has been a real driver in promotions in general, and therefore, promotion of women. So if you look now, still not perfect by any means at all, but I think in my discipline, there are far more women who are promoted all the way up the scale, and that's because, in large part, they have shown in research terms that they are valuable assets. So I think it's very important, and I think it still has a degree of importance in that regard.

Interviewee 13.

I still think, wholeheartedly, that one of the main factors which has had good gender benefits in the academic disciplines is the institution of the RAE, because it takes no notice of ... anything but output.

It merely takes cognisance of what people have produced and, in a positively evaluative frame of mind, what colleagues make of that.

However, as Interviewee 13 continues, this positive development for women does not mean that REF as a whole has been positive for career development, with those who are REF submit-able starting to make various demands in relation to their workload enabled by a REF obsessed university management structure:

Interviewee 13:

My feeling is that the previous experience - before 2004 - had been a good one. It was basically an evaluative exercise, where the emphasis was not entirely but predominantly on outputs - on published work. That is no longer the case. [...Now there are] managers [...] trying to do dreadful things. And one of those dreadful things is in some institutions already we're seeing that the people who have been returned and who have earned quite a lot of money out of that are saying they shouldn't have to do as much of X or Y as the people who weren't returned. So the implications of what looks like a technical decision - those people go in and those people don't - has implications for the day-to-day fabric of how people relate to each other, how people see themselves, what their working lives are like. And that wasn't where I came in the door, you know... Where I came in the door was a collective exercise to make the academic disciplines better. Not for the kinds of things that we're now seeing, where it's been used, whether advertently or inadvertently, it's been used in a way that creates massive divisions between colleagues, between former colleagues.

Developing this point in a slightly different direction, Interviewee 26 suggests that the REF has indeed become a basis for orienting the development of an academic career, but notes this as a positive development:

Interviewee 26.

It's in the back of everybody's mind. I've been doing other appraisals recently for some of the academic staff in this unit and it's amazing how much they are all thinking about what their papers and what their impact case study is going to be for 2020. This has become a very dominant bit of academic thinking and it's certainly part of the formal promotion process here. It's not written down that way but if you look at the criteria you can see how they map what REF reported. That's fine, I think having that focus for academic outputs has been really helpful. It's not how many papers you write, you've just got to have four really good papers over a period of seven years or something, that you're proud of and you can show they've made a big difference

Finally in this section, Interviewees also noted that the REF in some sense becomes a career with the University of Interviewee 6 employing people to work solely on the REF and Interviewee 33 noting that the REF takes over people's lives:

Interviewee 6.

People, I think, do have a pretty clear idea of what the university expects of them, and that's no bad thing. I did, and still do, resent the amount of work that the REF has been in the university. It does essentially take over the university for a year nearly. [...] It's a massive operation [...] I'm not sure it's the most productive way, and it costs quite a lot. We employ a fair number of people to do nothing but to manage the process. So, I'm not sure, in that sense, it's cost-effective, but having a research review process seems to me to be no bad thing.

Interviewee 33.

With the pressure on staff to be doing teaching, research, enterprise, helping administer the university, is a big load. There are individual pressures on individual staff members, of any discipline actually. The ones who get writer's block and just can't get writing. This drives people to nervous breakdowns. These are serious. This whole assessment, metric assessment of things, is a major pressure that people, individually, feel. I think that's the negative. That's the leading negative, I would say.

#### ***E.4 TRANSPARENCY, ACCOUNTABILITY AND JUSTICE***

A complex concern aired by several Interviewees without single agreement, was the issue of to what extent academia ought to be transparent and accountable, how the REF might be made more transparent (and the consequences of doing so) and the sense of justice or injustice that the REF promotes.

In the following excerpt, Interviewee 17 is clear that academics need to be held to account for their use of public money:

Interviewee 17.

And I'm afraid I don't have a lot of sympathy for academics who don't realise that, actually, they are part of the public good and that they have an obligation to that public good. I feel very strongly about that.

However, various Interviewees raised concerns regarding an absence of what we might term mutual transparency and accountability. As Interviewee 10 suggests, a lack of transparency within the REF is leading to frustration:

Interviewee 10.

I think there's a growing sense of frustration and disillusionment. That would be the sort of thing that would make me think this is a bad way to keep proceeding. The trouble with this whole thing is, in the end, there are no rewards. There are only sticks, no carrots. If there's never more money going in, and yet everybody is exhorted to do more and more research, then... [...] I think it probably can't be addressed within the REF itself, unless it was more transparent. The problem is the universities put in submissions without, as I've said, knowing really what's going to happen as a result of the outcomes. Perhaps, if it was more overt and clearer, and everybody knew that there would be this addition of funding relative to scores. And that, if the quality improved, more money would come into the system. [...] And how much more, then I think it would begin to, at least, function as something in which universities could make sensible decisions. I think, at the moment, because what seems to happen is, the rankings go up every time, but there's never any more money. [...]

Everybody knows there is lots of stress and dissatisfaction and really long hours worked in academia and much frustration at failure to obtain funding. That's only aggravated by the REF. I think the functions have shifted over time, but it's now not quite clear that this is going to benefit anyone. Once it's not beneficial to the academic enterprise then it's, so what if you got fantastic ratings if you can't get any money because there is no money in the system.

Other Interviewees pointed to distinct problems with what they felt was an absence of transparency and accountability. Interviewee 13, for example, suggests that the way the REF is micro-managed in comparison to the RAE, now prevents an overview of scores and systemically limits justice:

Interviewee 13.

as a panel member, the thing that has changed most dramatically is not, "How much reading do you have to do?" but, "Can you do justice?"

Now, you can no longer do justice because you're doing everything in such fragmented way, and you are being micromanaged as you do it. At various points, I would guess that all of the sub-panels - or certainly a high proportion of them - got told, "You should look at that again because we don't think you're doing it right. Knuckle under."

In the following excerpt, Interviewee 37 develops a similar point, suggesting the REF enables no more than a 'snapshot judgement':

Interviewee 37.

If I have my REF subpanel chair's hat on, of course I can say all this is doing is providing, kind of, a snapshot judgement which enables the funding councils to distribute funding, it's a form of accountability, it is no more than that.

But I know that, for my colleagues, in particular my junior colleagues who are starting out, it's enormously stressful to, more or less constantly, be told, "Are you on track for your next submission?" "Are you putting enough effort into this?" "Where are you going with your next piece?" "What's the timescale going to be?"

However, emphasising the complexity of transparency and accountability, Interviewee 37 also goes on to discuss the upside of academic accountability for the seriousness of research and the difficulties involved in making the REF itself accountable:

Interviewee 37.

I think that it has provided, over the years, quite an important mechanism for bringing academic research and scholarship; making it more accountable. If I think back to how research was organised and dealt with 25 years ago, we did live in a different world then, things were more relaxed, you pretty much did what you wanted. I think it was a fairly comfortable and complacent environment. My own view is that the sequence of assessments has made people take it more seriously.

[...] In an ideal world, one would wish to be completely open and above board and to say, "This got a three star because of the following reasons." I think there are a couple of things that mitigate against that. First of all, the sheer volume of material means that all you could record, and of course it was recorded, but as you say, not published, would be key words like, low originality or whatever.

[...] If I'm being more hardnosed about it, I would say that as soon as you say, "This is a three star. This is a two star because of X, Y and Z," then you lend yourself open to challenge. I certainly didn't want to find us being judicially reviewed on the basis that our judgements couldn't stand, and of course, it wouldn't succeed. But it allows people to begin to second guess.

In the following excerpt, Interviewee 18 suggests that a further downside to making REF decisions open and transparent would be the possibility of 'performance management' based on REF scores:

Interviewee 18.

I think that one of the reasons why feedback is very careful in REF, is that REF can be used for performance management [...] And by performance management I mean managing people out of institutions. So I think it is a process where you want to make sure that you are protecting the integrity of the exercise, but also protecting the

disciplines, and making sure that REF cannot be misused to push individuals out of positions, or to allow Vice Chancellors to make decisions that can be blamed on REF, or can be attributed to REF. So I think it's a very delicate exercise. [...]

A similar point is developed further by Interviewee 21, who suggests that academic colleagues noted value in keeping REF as a black box, limiting the possibilities of using research assessment as a management tool in the University. Interviewee 21 also notes that opening up decisions to potential dispute would make the role of panel members extremely difficult, a point also made by Interviewee 24:

Interviewee 21.

it became very quickly clear that, as far as most of my colleagues in other universities were concerned, the more REF was a black box, where you fit things in, you have no idea what happened, and then you got some results out, the better. Because their major fear was that the more that was said about how the process worked the more their university administrators would try and second guess how these rules worked and would impose inappropriate requirements on them. So they wanted it to be impossible for their university administration to guess what happened within the black box, so they couldn't interfere.

[...] I think that if it were known that the discussions were going to be open it would be very hard to get sub-panel members to volunteer to do the job. [...] The prospect that for a year-and-a-half to the end of it, one was endlessly dealing with queries and complaints and so on, is not an attractive prospect. Not because one is not willing to explain absolutely everything that one did, but simply that the burden of that explanation would become greater than the burden of the exercise in the first place.

Interviewee 24.

It's very difficult to preserve anonymity. [...] It's very difficult to express your opinion freely, because quite often you're talking about people that you know very well in your field. I think it would impede people from expressing their opinion, having a recording of everything that you say. [...] It would change the whole way that the exercise is working. [...] You might be more reluctant to criticise a piece of work, or you might give higher marks than you intended to, because, that guy's a friend of mine, he'd hear that.

In place of a more encompassing sense of transparency, other Interviewees (20, 25, 27) noted various REF practices that might be made more accountable:

Interviewee 20.

the issue of how far did you go to, if you like, calibrate your scores against others, and discuss them and defend your overall average scoring, I think those could well be made publicly available. I think, if you're talking about the individual scores of individual papers, you would then get Freedom of Information requests from every university in the country. It's a little bit like public examinations. I don't know. It would cost even more money than it costs at the moment to do it.

#### Interviewee 25.

The trouble is that HEFCE were paralysed with fear of being sued. With Freedom of Information being what it is, if they'd kept the records, then they would surely have been in danger of... I don't know whether there was a real danger, but, anyway, they were paranoid.

[...] Closed peer review is, in my view, the backbone of the whole peer review system. Therefore, since REF is peer review, it would be impossible to reveal the actual marks for individual papers because it would be clear who had done it. [...] A mark for every institution, of course, is revealed, but not in any detail, and the feedback is very anodyne. I think there's room for the feedback being more pointed, but, again, they're very worried about somehow revealing too much about what's gone on on the panel and then being sued. If you say, "Too many of your staff are not producing outputs," that may be true, but then they would be worried about somebody suing them for it.

I have no objection to there being more detail there for individual HEI departments, for sure, but not at the individual level. It could be very distressing for the individuals involved if they're putting four papers out that they think are marvellous and they all get two stars. It's going to be fairly depressing.

[...] Something like this will surely go on forever because governments like this sort of exercise, and you can't really argue against it because it's all about accountability [...]. We're forced to live with something like this, and so there are bound to be continuous exercises of some sort. The real danger is that they make it all metric-based, and then it will go bananas.

#### Interviewee 27.

The minutes do tell you what was discussed. [...] Of course, they're written in a very formalistic way, aren't they?

[...] After all, what sorts of things would we be talking about here, because obviously the minutes couldn't reflect things which could be either attributed to individuals or attached to particular institutions, I don't think? [...] There is a sort of a sense that the whole thing is deliberately obfuscated and surrounded in mystery, isn't there? [...] I know. I wish it were as exciting as that (Laughter), because most of it is pretty boring and it's just sort of dealing with awkward things around submission dates, and this, that, and the other, and how you interpret the special circumstances, and just all sorts

of things like that. I'm not sure that there are any wonderful scandals ready to emerge here, even if they were to be repeated more fully. [...] But I know that they must seem, when you dig up the minutes yourself, as if they were written deliberately to tell you nothing, and I don't know any more than that, really (Laughter).

In the following excerpt, Interviewee 31 suggests that the absence of transparency and accountability around REF scoring opens space for gossip:

Interviewee 31.

it's such a farce, the whole thing. Because there's all these high-minded statements about confidentiality and all this sort of stuff and the deliberations of the panel going no further than the room, it's absolute nonsense. Academics are, vocationally, the biggest gossips perhaps of any profession, of any workspace. It is the most leaky of professions in terms of the circulation of professional gossip.

So everybody knows exactly who the assessors were, who made the judgments, who's responsible. [...] That's where it becomes farcical, and that's why I feel so frustrated by the lack of - one the one hand, the absolute proliferation of gossip and leakiness, and on the other absolutely no recourse to question, no transparency officially, limited transparency. Weasel words of feedback, that's the thing that I feel is such an affront. On the one hand there's no transparency, at the same time it's like a sieve, the system. [...]

Finally in this section, Interviewee 33 was alone in suggesting an alternative basis for accountability and transparency – having anthropologists study the work of Main Panels and sub-panels:

Interviewee 33.

I've already said that I think it would be a wonderful thing to have an anthropologist observing. You would handle the issues of personal information and anonymity, a stage beyond that. You would have an anthropologist who was able to talk about the process and to observe it and to write it up. But, obviously, they wouldn't reveal that we had had an argument about Professor X's paper. [...] I think that shouldn't be revealed. [...] The main reason is that I think HEFCE have a process which is about assessing how good research is. It's an assessment of an institution. It's not an assessment of an individual academic. The individual's academic research is put up, but what you don't want is an institution to have any good reason to then sack particular individuals on the basis of this process.

## ***E.5 FUNDING***

One of the key effects of the REF that stimulated varied responses from Interviewees was the distribution of funding. For some Interviewees this was not a concern at all while for others it was a deeply problematic feature of their participation in research assessment. In this section we group together the variety of responses.

In the following excerpt, Interviewee 18 suggests the REF was not primarily about the allocation of funding:

Interviewee 18.

So I feel that the primary purpose of the REF, for me, was not to inform the allocation of funding, that happens afterwards. For me it was about identifying the quality of research, in [Discipline X], in UK higher education institutions.

For other Interviewees, funding allocation and its consequences was something that they were at least partially aware of, as Interviewees 10 and 27 indicate:

Interviewee 10.

I think everybody is always aware, but because you don't have any idea what the actual factors for the money are, you're always basically doing things in the dark. I find that a bizarre way to do business, honestly. So there's not much point about talking about the impact of money because you have no idea whether departments that are four star rated are going to get five times as much money, twice as much money, ten times as much money.

You just have no idea. It's obvious that if departments are rated as very weak, it's going to be a financial problem. Everybody can see that and one talks about it. Particularly where there are one or two stars, outstanding people in an otherwise weak department. There was discussion about the implications for those people. [...] But, generally, there's nothing to talk about in terms of money because nobody has any idea. It's obviously better, the higher your ratings. The point is we're deliberately very vague.

Interviewee 27.

I'm not sure that it's so distinct from the hat that you change when you're reviewing a research grant, or even assessing a PhD or whatever. It's a very odd feeling in the REF to know that every time you meet somebody, you may not be directly, but you're assessing their work. Every single academic, [from Discipline X], that you meet during the process, you are conscious of the fact that you are somehow sitting in judgement at the same time, but I think that's only an extension of what we do a lot of the time, even in our own work, isn't it? We're constantly reviewing and assessing.

That's not quite the same thing as putting a mark on, a score on somebody and then giving them, effectively, government money, I know, but when you're doing it, it just feels like you're doing a normal part of academic reviewing. That's just what it feels like. Then you're discussing with another one or two people how you feel about the certain piece of work. The only difference is this wretched score that you have to attach to it; otherwise it's really interesting (Laughter).

For other Interviewees (24, 30), funding allocation and its consequences was central to the work they did in the REF and it was something that made them feel uncomfortable:

Interviewee 24.

Oh sure, that's [funding distribution] probably stuck in the back of your mind. It's not the first assessment we've been through, so we know the effects of that, and-, yes, but I mean it's the same for everybody. [...] I think you do that with a hand on your heart, and you say I'll do my best, and obviously you want other people to do their best as well in handling your papers.

[...] it certainly has an impact, a very serious impact on some institutions. I know whole departments closed down because they didn't do well in-, even in top-ranked universities, that happens. Certainly, it's important in terms of the money that comes in. It certainly shapes opinion in the university as to how research should progress, and for example now we are planning for the next exercise. We're having meetings, and we have a new [senior manager] for research, we are organising everything and trying to anticipate what's going to happen next. [...] You know where the marks are going to go, they're going to go to the people who are doing excellent research, so if anything, I think the REF pushes funding in that direction, to the detriment of new research

Interviewee 30.

To be honest, I was very uncomfortable about this [...] You know, to be frank, we are operating in a 'winner takes all' culture. The, philosophy is you identify the leaders and shower them with resources. In my view, that's not the right way to maintain the health of the university system as a whole.

I mean, even if you get the assessment right, there is wastage in really substantial rewards to a few supposed leading groups. So, I was very aware of the financial implications of this. It's not just the allocation of money, but the way universities are able to make use of a good assessment to promote themselves and attract students and future donors.

[...] I think the overall rankings were... I didn't have misgivings about them, I think it was broadly correct. It was heart-breaking that some very... In my opinion, departments who were doing good work tumbling down the rankings because, somehow, they didn't fit the template.

This sense of discomfort was continued by other Interviewees who noted ‘departments have been closed’ (Interviewee 12) or ‘people’s jobs ultimately depend on it’ (Interviewee 15):

Interviewee 12.

Of course, you see how well things come out, that the measures are used, and you know they’re quite important. So, in some ways, that’s partly why you feel some responsibility that you’d like to get them right, because we know that some things... In my subject area, quite a lot of departments have been closed down since the RAE began. Presumably, if they scored badly in the RAE, the university would say, “Oh well, we’ll close down that department,” or, “We won’t recruit any more staff in that.”

Interviewee 15.

people’s jobs ultimately depend upon it. So I think REF permeates every aspect of what academics do now.

As Interviewee 26 and 13 note, the emerging funding consequences of the REF arrived too late in the assessment process for them to offer any useful response:

Interviewee 26.

you see them [scores] all from a particular university coming in really high or really low. Then you go, “That’s going to be bad for them”, or “That’s going to be good for them”. By then it was too late to do anything about it.

Interviewee 13.

In a way, knowing the effect of the way that the overall assessment of a unit of assessment would turn out was not something that we knew until the end of the process. In terms of the timetable to keep to, we only saw the effects of the different parts of what we were doing in terms of output and, in particular, impact not long before we finished. [...]

I’d become very concerned about the effects on [Discipline X], of previous RAEs, and the fact that there were fewer and fewer units of assessment returning to [Discipline X].

I wanted various steps to be built into what we were doing, such that we could look at what the outcomes would be on the different measures and say, “Is that fair and reasonable?” There was a point when we did that, right at the end, when we looked at

it and ... said, "That is not reasonable. The technical exercise has produced full justice." ...

It wasn't that people were being horrid or anything, not even with the micromanagement. It was the technicalisation of the process. It breaks everything down into little fragments, and you don't get to put them all back together until right at the very end. Then, "Oh, God, what are we doing?"

[...] Collectively, if we think that there is something amiss with this technical exercise, we can do something about it. Of course, that's what the taskmasters don't want us to do. Partly for good reasons. It's about bias, and they think that we're going to pat our friends on the back. Also, it puts back into the hands of a discipline - which I think is where we started out way back - trying to produce good outcomes for the discipline as a whole.

Interviewees also devoted some time to discussing the broader problems with scoring and the incentives that funding provides. For example, Interviewee 1 worries that the funding allocation for research has diminished the importance of teaching and Interviewee 5 expresses their frustration with administrators continually attempting to guess the funding outcomes of particular scores:

Interviewee 1.

One worry that I'll always have about [...] the REF is the distorting effect it might have on universities and their decisions about research or what to do. And I think that is a worry. [...] The focus on the REF, the focus on the RAE, have I think probably got to the point where it's dysfunctional and damaging to other aspects of university life, particularly to teaching. So the TEF, you know, could provide an antidote if universities take it seriously enough, if the rewards are sufficient both in terms of finance and in terms of reputation

Interviewee 5.

there were several occasions on which I slightly told off the administrators for assuming that the old formulation would just continue through. In other words, there would not be a change in the way that the money was allocated. I had to remind them that that was a decision for the HEFCE Board that would not be made until after the whole exercise was completed.

Interviewee 5 then continued to offer one of the few alternatives offered by research participants to the current funding allocation model:

Interviewee 5.

It would be possible to set a geographical minimum by region. This is something I'm quite attracted to. [...] Yes, It's been said that the money should flow where the excellent research is, but could be subject to a minimum, so that each region has some protection. [...] Simply to make sure that all of the research money didn't end up getting sucked into a golden triangle or just a small number of large institutions. There's been a lot of worry about the so called cold spots. Areas of the country where there's poor higher education provision. Maybe, at some point, it would be sensible to use the REF mechanism to make sure that each region has some minimum amount of research funding to keep it sustainable.

## ***E.6 VALUING RESEARCH***

In this final sub-section of the report we look at the value that the REF has brought to UK academic research. Many of the Interviewees were clear that the REF does not just involve assessing or evaluating research, but is also focused on creating a sense of the value of UK academic research. As Interviewee 2 suggests, demonstrating value goes hand in hand with attempts to change behaviour:

Interviewee 2.

I think probably from a government perspective, central government perspective, part of the driver of introducing impact was to incentivize behaviours, and to encourage researchers to pay attention to delivering impact from their research. [...] I think it was also driven from a desire – and that's probably more or less a Hefce funding body's objective – to improve the evidence base around impact. I think whereas Treasury might come at this from a 'we need academics to do more things that deliver impact', I think [Organisation X] were more coming from a perspective of 'we need to be able to better demonstrate the impact that academic research is already having'. And so the extent to which there is a policy, it's a kind of mixture between a sort of behaviour change and advocacy type of activity. And I think, actually, you know my assessment based on the evidence, [Organisation X] probably achieved both of those objectives, and I think [it] caused behaviour change, certainly at the institutional level.

For Interviewee 27 this demonstration of research success and the impact of research through the REF is now a necessity to defend HEFCE and perhaps UK research from government cuts:

Interviewee 27.

in the previous RAE, HEFCE was regarded as kind of the enemy, actually. [...] I think that we, certainly on the subpanels and probably academics generally, up to a point at least, suddenly twigged [...] that we, as subpanels, needed to help HEFCE,

but, of course, already by that stage there [...] were threats in the air and somehow that they had to sort of pull the rabbit out of the hat with the impact stuff. [...] Whereas in previous exercises it was all competitive between the institutions all the time, the bigger show here was to show that these subject areas matter, and that we have an international and disproportionate reputation for the work that we do, and that we push the boundaries in practice research and all those kinds of things.

Somehow, at some point at least, even though, as we all know, the money – the pot itself – would stay the same or go down, nevertheless that we had a sort of a responsibility to our subject area and to the broader [disciplinary] sectors as a whole. I think that was a bit of a change. [...] almost became a kind of, “Let’s see if we can save HEFCE, or at least save QR and any future REFs, for fear of what worse may follow.”

For other Interviewees, the importance of the REF has been in getting universities to take research seriously. As Interviewee 25 articulates, the improvement in research quality over the time of the RAE and REF has been clear, even if the competition itself has led to multiple institutions each attempting to claim they are top of the rankings:

Interviewee 25.

It’s amazing how many people said they were ranked top in research according to the REF, because they’d picked a particular cubbyhole [...] space where they were top. It was remarkable. I think there were about seven or eight people who said they were first in the REF rankings. (Laughter) It was all there and they weren’t lying. It’s just that they were finding a particular narrow definition of what they were first in. [...]

From that point of view, it has been certainly a catalyst for competition, and I think that’s a good thing, on the whole. The other thing that you could argue - and, again, I’m not quite sure how far I’d want to push this - is that there’s no doubt that the standard of the research outputs, now forgetting about impact, has gone up since I was involved, and by most measures, from the very beginning of RAE. [...] This increase in quality is probably a real thing.

[...] It’s got many faults, so I’m not particularly pushing REF, but I think you have to say, in all fairness, that there must be some fraction of that which is due to REF. People have concentrated their minds. They’ve tried to make sure that they’re producing good results and that the stuff they publish is world class. I think it’s had some effect. I think it would be perverse to somehow discount REF entirely from the improvement that there clearly is. I think it’s had some effect, but I couldn’t say how much.

Finally in this section, Interviewees noted that the value of the REF is for Universities. According to Interviewee 7 in the following excerpt, the REF is successful at a macro level

demonstrating the strength of UK research and at a micro level, enabling departmental decisions for future investments:

Interviewee 7.

From macro to micro. I think the evidence suggests that at a macro level it is highly successful. The UK has gained, and if anything, improved its position, in terms of its global productivity. And if you look at whatever measure of global research productivity, the UK performs exceptionally. It has sustained that position in the presence of very strong global competition, producing a system which is highly beneficial to the UK. You could argue that it is almost the most successful thing that the UK does. [...] Research. We do it in the UK, and REF has contributed to that. [...] To that success. So macro level, it is a successful thing.

[...] Then we zoom down to the departmental level, you know, from a managerial point of view, you can make investments and see the return on those investments. You have an effective benchmark for your performance REF is a benchmark, not a signpost. [...] So as a manager, I would- I should say that the department I headed was number one on GPA in the REF 2014. (Laughter) So you can invest, and you can see the return on those investments. Is it a big ask that people put forward four outputs, and a department tells a small number of stories about how it's made a societal difference? It's not a big ask. And any well-managed institution would be doing that anyway.

## Summary

In this section of the report we have drawn together the varied discussions of Interviewees on the effects of the REF. For several Interviewees one of the central and more disturbing effects of the REF was gaming of the system. Here Interviewees suggested that recruitment strategies, such matters as fractional contracts and the hiring of research stars, could all be attuned to the timing, structure and scoring system of the REF. In this way, a department and its members might be strategically shaped to try and benefit from the REF. At the same time, individual academics were noted as trying to play the system, presenting themselves as research stars that ought to accorded certain benefits or as impact experts, for example, who might successfully steer a department's REF return. Marketisation was noted as a further potential effect of the REF, imposing forms of competition and constraint on the Higher Education sector. However, Interviewees also noted that competition had brought positive benefits in the organisational professionalism of institutions and in making universities accountable to an economic rather than a political cycle. Other Interviewees noted that competition was not quite the same as a market, with the REF operating as a mostly bureaucratic exercise. Perhaps the REF signals the evolution of a hybrid form of intervention amalgamating market and bureaucratic logics.

Interviewees were also clear that the REF has significant effects on academic careers, shaping individual's choice of which department to work in and even signalling the viability of departments as research organisations. At the same time, equality and diversity was improved

by the REF according to Interviewees by taking seriously the circumstances that shaped people's working lives. One factor that was indicative of the growing importance of research assessment for UK universities was that the REF itself has become a career.

The subject of transparency, accountability and justice became a focal point for articulating a variety of viewpoints. For several Interviewees the REF was clearly lacking in transparency and accountability, leading to a disillusionment with research assessment. For other Interviewees, greater accountability was impossible as this would lead to complaints, career management and even litigation. However, some Interviewees also thought greater efforts could be made to consider which parts of the REF could be made more transparent, how greater feedback might be given, perhaps even to give a greater sense of justice to research assessment. In a similar manner, Interviewees also expressed broadly different viewpoints on the effects of funding distribution. For some, research assessment could be carried out without regard to the funding consequences, while for others this remained a profound and on-going concern. For the latter Interviewees it seemed that being a sub-panellist was a stress inducing matter. Making consequential decisions that further into the future might mean job losses or departmental closures was a matter of concern.

Finally, we ended the report looking at the ways Interviewees suggested that the REF brought value to UK academic research. The impact cases were noted as resources that various audiences could use and that demonstrated the value of UK academic research. The REF also offered the chance, according to Interviewees, to demonstrate the value of academic research more broadly beyond impact. At the same time, the research assessment and its move beyond a narrow assessment of academic worth also enabled the HE sector as a whole to demonstrate its value to the UK, perhaps defending HEFCE and defending the field against cuts.

## Conclusion

### A. Panellists' motivation:

The dominant motivation for Main Panellists and sub-panellists to participate in the REF was a sense of civic responsibility and a need to represent their academic field. Other motivations included a desire to change the REF for the better, to carry their misgivings into the process in order to make changes, to develop their own interests or career, to represent their institution and secure income for their department. Impact assessors also reflected similar motivations: a form of civic duty, a hope to make change by bringing impact to public attention and a sense of disquiet about the REF and its ability to assess such matters as impact.

### B. Purpose of the REF:

Interviewees articulated a variety of purposes for the REF. These ranged from a formal definition of the REF (based on the selective allocation of research funding) to discussions of the complexity of financial allocations being linked to competition. Also prominent in the discussions were prestige and reputation. Others suggested that the purpose of the REF changed over time, benefitting UK research as a whole, enhancing accountability and even enabling community building. However, other interviewees pointed to the ever increasing stress and pressure of the REF, its effects on streamlining funding and marginalising certain institutions.

### C. The work of the panels:

Here Interviewees spoke at length of the extent to which the REF managed to be representative. This included a concern for relevant forms of expertise, the range of work that a sub-panel would need to assess, and how a sub-panel might look to its Main Panel and to audiences outside the REF. Ensuring a discipline itself had adequate representation was also important, along with having someone to fight on its behalf to defend its interests. Concerns regarding the adequacy of representation on sub-panels included questions of gender, geography, age, departments and university types (old, new, small and large).

### D. Assessing academia:

Interviewees discussed the challenges of peer review, both in terms of the workload and in giving scores. Disputes, spurious accuracy and structural limitations were all noted as problematic. Interviewees also discussed calibration and the methods by which a score ought to be achieved. Continual calibration, normalisation and the relative ranking of outputs were all subjects of discussion. Further discussion focused on metrics not as an alternative to peer review, but as a potential aide in the scoring process. Impact was then discussed as the final aspect of academic assessment, raising issues of the promotion of UK research, the ease or

difficulty of producing and assessing cases, the expansion of guidance to universities, broadening the definition of impact and the removal of the impact template.

#### E. Effects of the REF:

The effects of the REF were discussed by Interviewees in several ways. Gaming through recruitment strategies, fractional contracts, research stars, the self-promotion of individual academics were noted as disturbing aspects of the REF. Marketisation was considered a potential effect of the REF by some Interviewees, in terms of competition and the constraints it might impose on the Higher Education sector. However, other Interviewees also noted that competition had brought positive benefits in the organisational professionalism of institutions and in making universities accountable to an economic rather than a political cycle. The effects of the REF on academic careers were also discussed in terms of the shaping of individual's choice of department, improvements in equality and diversity, and the development of the REF itself as a career. Transparency, accountability and justice was a point of contention between those disillusioned by the absence of transparency, those calling for selective transparency and those suggesting that transparency was unachievable. A further point of contention was the effects of funding distribution and the extent to which it was a concern for REF participants. Finally, we ended the report looking at the ways the REF has brought value to UK academic research, through creating a resource of impact cases, by demonstrating the value of academic research more broadly beyond impact, and by enabling the HE sector as a whole to demonstrate its value to the UK, perhaps defending HEFCE and defending the field against cuts.

---

<sup>i</sup> Brown, R. (2014) *The Marketisation of Higher Education*. Available from: [http://www.uwl.ac.uk/sites/default/files/Departments/Research/new\\_vistas/vol1\\_iss1/vol1\\_iss1\\_art1\\_23April2015.pdf](http://www.uwl.ac.uk/sites/default/files/Departments/Research/new_vistas/vol1_iss1/vol1_iss1_art1_23April2015.pdf) (last accessed: 6th January, 2017); Sayer, D. (2014) *Rank Hypocrisies: The insult of the REF*. London: Sage; Strathern, M. (2002) *Abstraction and Decontextualisation – an Anthropological Comment*, in S. Woolgar (ed.) *Virtual Society?* Oxford: University Press.

<sup>ii</sup> In order to make the interviewee responses anonymous, we have simply randomly allocated numbers to each interviewee (Interviewee 1, 2, etc). We have replaced the names of all institutions with 'University X', unless a comparison is being made in the same excerpt with another institution, in which case the second institution is referred to as University Y. We have followed the same procedure with disciplines and sub-panel names (now referred to as Discipline X and Sub-Panel X) and the names of colleagues, sub-panellists etc whose names have been changed to 'Expert X.'

<sup>iii</sup> In the interview excerpts [...] signals text removed by us in writing the report, either to retain coherence or to protect anonymity. In the excerpts '...' signals a pause or change in thought by the interviewee.