Learning from the UK Research Excellence Framework: ends and means in research quality assessment, and the reliability of results in Education

Ian McNay

This article first reviews the objectives/ends of research quality assessment in several countries to draw lessons for the UK Research Excellence Framework and similar exercises. It then reviews work on performance management as a framework for reviewing the views of participants on the means to the ends – the management of their experience in submitting to the 2014 REF. Finally, it examines the outcomes and considers how true a picture they paint of the quality of research, particularly in Education. It concludes with recommendations for change.

Keywords: research quality assessment; performance management; education

Introduction

Results of the UK Research Excellence Framework (REF) submissions were announced in December 2014. In comparable systems, the Performance Based Research Fund purpose and process is being evaluated in New Zealand before the next exercise in 2016; there has been a national overview in Australia of the Excellence in Research for Australia system in time for its next cycle in 2015; the census date for the 2014 Research Assessment Exercise in Hong Kong was September 2013, very close to the UK timetable. David Willetts, former Minister

for Higher Education, suggested that an international exercise might develop across these four systems. This paper seeks lessons from the UK 2014 experience. It first examines approaches within different HE systems to defining objectives/ends of research quality assessment. Within the context of recent research on performance management, and drawing on a survey of active researchers, it then assesses how well the process criteria set by the UK funding councils for the conduct of successive exercises were met. Finally, it looks at REF outcomes and takes Education as a case study to consider validity and reliability of the results.

The ends: for what purpose/s?

The aim of the UK RAE has been kept simple. For several iterations it was 'to inform decisions on funding'. For 2008 it became 'to fund excellence wherever it is found'. That went with a change in scoring the units of assessment to introduce a profile, rather than a single average score. The rating criteria had also been criticised for a mix of vagueness and absolutism which did not help those preparing submissions (McNay, 2003, 2009). For 2014, the consultation on moving to the REF implied that it could be used as a means 'to change behaviour' (HEFCE, 2009) but with no indication at that stage of what the desired behaviour might be, except by implication through the introduction of discrete funding for 'impact'. The vagueness of the criteria remained. Documents from HEFCE after the results were announced made claims about achievements, implying a retrospective framing of them as objectives:

> Providing accountability for public investment in research
>
> Providing evidence of the benefit of such investment
>
> Providing benchmarking information
>
> Establishing reputational yardsticks for use within the HE sector and for public information
>
> Leading a cultural shift, with changed practices and new strategies linked to impact, though more among managers than researchers
>
> (HEFCE, 2014b)

The judgements had always been of past performance; funding was, therefore, a reward. There were no stated objectives related to increasing

research activity nor to enhancing quality or to developing capacity and capability, though a March 2015 press release on funding allocations, announcing no increase, even in cash terms, for research, did claim 'to provide selective support for the next generation of researchers'. For 2014, capacity declined: the number of people submitted as 'active researchers' has gone down between exercises, from 52,401 FTE in 2008 to 52,077, despite an increase in numbers of academics in the sector. So did productivity: the number of outputs declined even further as the number per head also declined, to 3.41 from 3.75 (REF, 2014a). The limited 'audit' approach may be because of concerns not to breach the autonomy of institutions and because HEFCE sees itself as a funding body not a strategic one.

That contrasts with elsewhere. In New Zealand, the Performance –Based Research Fund is allocated to institutions on the basis of Quality Evaluation (QE – 60 per cent), Research Degree Completions (RDC – 25 per cent) and External Research Income (ERI – 15 per cent) (Ministry of Education, 2013). All academic staff submit portfolios and receive an individual rating, whether or not they are research active. The government's aims in introducing the PBRF were to:

> *Increase the average quality of research*
>
> Ensure that research continues to *support degree and postgraduate teaching*
>
> Ensure that funding is available for postgraduate students and new researchers
>
> Improve the quality of public information on research outputs
>
> *Prevent undue concentration of funding* that would undermine research support for all degrees or prevent access to the system by new researchers
>
> Underpin the research strength in the tertiary sector
>
> (Ministry of Education, 2013, italics added)

Differences from the UK ends are clear. There is a specific commitment to quality improvement. There is a clear linking of research to teaching, specifically excluded in the UK in terms, for instance, of impact case studies. Other countries, such as the Netherlands, review teaching and

research together, with visits to institutions, possibly seen as not possible in the UK because of expense in using public funds in a larger sector. The Netherlands approach is also less stressful because there is no direct link to funding. In Hong Kong the UGC (UGC 2012) adopts an holistic approach, using all four of the Carnegie scholarships as a basis for judgement – discovery, integration, application and teaching (Boyer, 1990). In Australia, after several unimplemented schemes, including one with an imitative impact rating element, the Excellence in Research in Australia (ERA) also omits teaching. It emphasises:

> assurance to promote confidence among stakeholders
>
> stocktaking strengths and *'areas where there is opportunity for development'*
>
> identifying excellence across the full spectrum of research performance
>
> *identifying emergent areas and opportunities for further development*
>
> allow for comparison of disciplines nationally and internationally.
>
> (ARC, 2014)

There is another contrast with New Zealand: in reviewing the process. In the UK, in early 2014, the funding councils called for inputs to a review of the 2014 REF. The funding councils, however, restricted replies to one per institution funded (HEFCE, 2014a). That risks getting a distorted picture, mediated, if not dominated, by senior management, and excluding any voice of the researchers most affected by the process, even through, for instance, scientific and learned societies.

By contrast, the recent consultation in New Zealand (Ministry of Education, 2014) had 32 replies from institutions, including 15 from departments; six from HE 'peak bodies'; 11 from science and industry peak bodies and 78 from individuals. The consultation covered PBRF objectives and core policy design. The core process design remained almost unaffected. The main debate was over ends, and revised objectives emerged, reflecting the more globalised, competitive world in which HE now operates, and a stronger link to wider governmental policy objectives.

The primary objectives of the PBRF are to:

increase the average quality of basic and applied research at New Zealand's degree-granting tertiary education organisations

support world-leading research-led teaching and learning at degree and postgraduate level

assist New Zealand's tertiary education organisations to maintain and lift their competitive rankings relative to their international peers

provide robust public information to stakeholders about research performance within and across tertiary education organisations.

In doing so, the PBRF will also assist to:

support research activities that provide economic, social, cultural and environmental benefits to New Zealand

support the development of postgraduate student researchers and new and emerging researchers

support commercialisation of research and technology transfer to New Zealand businesses and organisations

(Ministry of Education, 2014:7)

The full account of the consultation and views of all who contributed can be seen at www.minedu.govt.nz.

My two points here are, first, to contrast the objectives of the exercises, and particularly to underline the developmental nature of those elsewhere, and the link between research and teaching – the 'harmony' that Humboldt thought should exist between the two functions – and second, to note the wide range of stakeholders and interested parties involved in review, which may reflect system size and so operational load, or system values and so management and leadership culture.

Means to the end: managing process and performance

I do not want to repeat the debate about peer review versus bibliometrics. My focus here is on the principles set out by the funding councils to underpin the process of submitting bids, and the way in which performance has been managed within institutions. I start with theoretical frameworks and some previous research.

Performance management: research findings

In 2014, while panels were busy assessing submissions made to the UK REF, reports were published on two projects examining performance management and motivation in universities, in particular among researchers. Kallio and Kallio (2014) studied management by results (MBR) in Finland, where it has been explicitly adopted since 1994 and used for budget negotiations between the ministry and each university since 1997. They summarise the approach:

> The essence of the MBR doctrine is to develop evaluation and feedback systems that encourage staff to focus their work output according to the organization's strategy … a person who performs well in line with the designated indicators is rewarded in the form of higher monetary compensation or other benefits.
>
> (Kallio and Kallio:574)

Their survey led them to conclude that

> The motivation to engage in creative, knowledge-intensive work, such as the work carried out at universities, is typically intrinsic. In the light of the empirical findings of the study, it seems that management by results is in conflict with intrinsic motivation and the very essence of the expert work undertaken in universities.
>
> (Kallio and Kallio:574)

Franco-Santos *et al* (2014), in a UK project funded by the Leadership Foundation for HE, noted that although

> traditionally, universities have seen themselves as stewards of knowledge and education, focusing on long-term scholarly goals, comprising the development of knowledge and the greater good for society at large

that is changing: 'they are currently becoming more short-term and results/output driven due to the increased pressures to perform' (p7). Among those pressures they include research assessment frameworks and rankings. They classified performance management systems as *stewardship-based* or *agency-based*:

> Stewardship approaches focus on long-term outcomes through people's knowledge and values, autonomy and shared leadership within a high trust environment.

> Agency approaches focus on short-term results or outputs through greater monitoring and control. (p7)

Most academics find agency mechanisms 'unhelpful and dysfunctional'; professional staff find them helpful 'as they provide greater clarity and focus' (p7). At institutional level, stewardship approaches are 'associated with higher levels of staff wellbeing, (which, in turn) is associated with higher research excellence, students' satisfaction, students' employability and financial results' (pp7-8).

In New Zealand, Edgar and Geare (2013) found a strong link at departmental level between cultural make up and operational practices and performance. Their high performance group was collegial: greater autonomy, more egalitarianism and less emphasis on formality than the more managerialist culture of the low performing group, though there were some features of managerialism, such as competition (p789), but with a strong support for achievement and individualism (p774). They cite Yokoyama (2006) to explain why the differences emerged:

> paradoxically, it was claimed that the introduction of research assessment led to the managerialist drift. Weak cultures had weak resistance to the external pressures – 'more susceptible to having their values, and hence their managerial practices, influenced by other stakeholders' including those running the national assessment systems
>
> Edgar and Geare (2013:789)

Rebora and Turri (2013), comparing RQA in the UK and Italy, use neo-institutional sociology and operational control theory to give a fuller explanation of the introduction and development of New Public Management in universities. Organisational behaviour adapts to achieve acceptance and legitimacy by a process of imitation of a 'best model' that leads to isomorphism, based on a desire to refer to socially shared metrics using common criteria that then affects the behaviour being measured. Institutional leaders use the external measurement mechanisms as tools for orienting behaviour and changing culture in a way desired by government, eg over impact.

Leathwood and Read (2013) record the responses of academics to such pressures. They allocate them to contestation, compliance and complicity. Policy was contested almost universally, especially when

> critical, innovative, small-scale, qualitative and/or feminist research was all seen to be under threat. Particular concerns were raised about

the impact on early career researchers and hence about research capacity building for the future.

The power of others and protection of their own involvement in research, with conditions, led to calculated, strategic behaviour:

> Despite high levels of contestation, however, almost all the academics in this study were complying with the demands of research audit and performativity, often at significant personal cost. Most did not see any option but to comply, with compliance seen as being the only way to contribute to do the research they loved, and to remain in employment.
>
> Leathwood and Read (2013:1172)

One respondent to a 2014 survey, reported below, confirmed this:

> The approach was to move to be 'selective' this time around, having been 'inclusive' in the last RAE. This was achieved through a centralisation of decision-making where decisions were not effectively open to challenge. No senior member of the School questioned that the games had to be played according to the rules (though some whinged about the rules) so of course we became agents of the state!
>
> McNay (2014:e8)

The authors acknowledge their own complicity, which many of us also must do, in encouraging colleagues to operate optimally within the culture.

In 2014, I conducted a small pilot project, exploring academics' experience of being involved in the REF submission process. The Funding Councils had asked for feedback on this, but allowed only one response per institution; my concern was that the voice of those on the front line would go unheard. My research on leadership and management has always incorporated views from the led and the managed, where the perspective is different and so is the evaluation. I wanted to record the experiences of being managed as a professional in a process crucial to institutional reputation as well as individual motivation. Respondents were asked to rate the principles underlying the process laid down by the councils. They represent the end to be achieved *within* the process: means towards a higher end.

clarity of documentation

consistency across academic areas

continuity between exercises as far as possible

credibility to those being assessed

efficiency in the use of public funds

neutrality over research approaches

parity across different forms of output

transparency

Performance against these was rated on a 1-10 scale at both system and institutional level, with collated scores converted to an overall percentage. Average scores are given in Table 1, with the number of responses to each item in brackets: not a positive set of ratings. There was an open comment section, where views could be expressed, particularly on the experience of impact case studies, a new element for 2014.

TABLE 1
Perceived performance ratings for operational principles, REF 2014 percentages and number of respondents

	Rating of processes at system level	Rating of processes for own institution
Clarity of documentation	88 (25)	63 (26)
Consistency between academic areas	44 (21)	58 (19)
Continuity between exercises as far as possible	40 (21)	46 (22)
Credibility to those being assessed	46 (25)	54 (25)
Efficiency in the use of public funds	36 (26)	39 (26)
Neutrality over research approaches	50 (24)	55 (23)
Parity across different forms of output	42 (23)	51 (24)
Transparency	55 (25)	59 (24)

35 requests were sent by email, from which I got 31 replies. 11 of those did not provide detailed ratings, but 7 of those provided usable comments and 2 gave an overall rating of both system and institutional processes. 12 respondents were at professorial level. 4 respondents had led the preparation of the submission for their UoA. 9 respondents were from my own institution; the others came from 15 institutions, 5 being modern, 'post-92' universities, 10 being more 'research intensive' pre-92 universities. 18 had been part of submissions for assessment in both 2008 and 2014; 5 in 2014 only; 3 in 2008 only and 5 in neither. These responses were supplemented by ten made as part of workshop activity on research strategy with research staff from two modern universities in the north of England, of whom 8 had been submitted in 2014. So, 38 returns had useable content.

Overall, internal procedures are more highly rated than those at system level, perhaps moderating the effects of external expectations. This is most true of the northern group and not true for my own institution. A picture emerges of an approach based more on agency than stewardship:

> It proved most divisive. The sense of them and us became very pervasive. The small amount of money for research projects available through our research centre was limited to those projects deemed REF-able by the person controlling funds. This meant, for example, that a new researcher applying for a small amount of money was refused. This does not nurture new researchers, nor develop a research culture.
>
> (n8, a team leader)

There was other evidence, too, of a serialist approach, with emphasis more on managing the exercise than leading research, with little connection to a longer term strategy (if any) or continuity between submissions – one of the principles above, but not helped by changes by the funding councils:

> While starting early was important, we started (here) before the panel criteria had been published. Poor feedback based on the *previous* RAE had a very deleterious effect.
>
> (e3, a professor, leading the submission).

One final recurrent theme was on poor staff management. One of the northern group (n3) had spent much early morning time while at a London recruitment fair getting urgent documentation back to the SMT for an impact case study, only for the SMT to realise a week before

submission that his recent arrival made most of his work ineligible for consideration. Others felt exploited:

> I felt very much like a resource in this exercise – not someone with interesting and valuable research to share. I found the process very impersonal. The institution in general has never shown much interest in my research and I've had limited support to do it. But all of a sudden my research was needed for the REF (and I was happy to participate, valuing the credibility it might give my researcher identity). In terms of the process leading up to the return, I just remember that I was usually asked for things quickly (copies of papers, rationales for research, completed special circumstances forms), but was never really sure what were the best things to submit – there wasn't a great deal of guidance. I found the process quite stressful, and I remember trying to calculate whether I had scored enough points to get in or not. I had to remind myself that this was for the university as well and that actually it was unlikely to impact on me that greatly (except perhaps my ego) if I did not get returned. I was interviewed for the impact case study – but I never saw what was written or whether anything of mine got in at all.
>
> (g4)

One professor elaborated on the drift from good intentions:

> Locally, while early attempts were made at the documentary and central support levels to try to organise the process as effectively as possible, regrettably those early aims were quickly forgotten. The process was subject to significant, unreasonable and unexplained delays, impact case studies were pulled together hastily, external assessors were hugely inconsistent, unreliable and indeed frankly quirky in their judgements. It was disappointing as a process and required a huge amount of work to be done by busy researchers numerous times, as items that were submitted were deemed to be lost or needed for yet another parallel system.
>
> (g8)

The comments below attempt to reflect the balance of views related to the process principles (neutrality and parity drew few comments, though there are references to equity of treatment).

The corporate attitude to inclusion and involvement is reflected in comments about ***clarity and transparency***, with implications for ***credibility:***

> The way it was done in our school and department was such that I never saw any of the documentation, and was not given any info about the process. I literally know nothing about it! I suppose that says a lot about lack of transparency. It was almost as if we were not allowed to know, nor were we allowed to know who else was being entered. A culture of opaqueness?
>
> (g1)

> All I did was fill in boxes that the institution sent to me; I was not given much indication of how this fitted in with the institution's overall submission.
>
> (e18)

> The process was extremely opaque… and there were never any official meetings, documents or discussions about the REF process…Overall the internal process seemed very rushed, based on personal relationships more than merit of work, and discriminatory against certain fields, research methods and positions within the institution.
>
> (e13)

> I found the transparency of judgement in our mock REF was low. I was a reviewer, but I have no idea if my judgements were in line with other reviewers or completely out of kilter. There was no moderation process.
>
> (e14)

By contrast

> Our institution was very open and honest about decisions and what they did.
>
> (e1)

and one experienced professor said

> 'I was only really aware of communications at the level of my school – all of which were excellent from my perspective'
>
> (e4)

When I suggested this was a contrast to many returns (including another professor in the same UoA), the response was

'It was just that they left me completely alone'

(e4)

Same experience, differently rated, because of expectations about documentation and process.

Only a few were able to comment on ***consistency across academic areas:***

> Overall the REF process was pretty excruciating although going in with Health has led to new research contacts, which is positive. There was a revealing experience of having my published articles reviewed by a reviewer for Business who graded them 2* and the reviewer for Health who graded them 3* with an additional report graded as 4*
>
> (g2)

A case study might help. One Business School sent all the REF submissions out to an external referee. This ONE person – so much for reliability – who remains anonymous, then scored the submissions, often reducing articles graded 3* by the Association of Business Schools list to a 2*. He also added scathing comments which had a very negative detrimental effect on the recipients (g7).

As someone who has been an external assessor several times, an internal one for 2014 and a subpanel member for the 2001 process, I share those concerns. At times I was expected to grade work at or beyond the fringes of my expertise and saw panel judgements influenced by knowledge beyond the documentation provided. There were other concerns about ***credibility***:

> This being my first experience of research assessment, I was satisfied with the way it was managed internally. At a system level I have less faith, especially around claims that panels will not use journal ranking/reputation in making their judgement. Having been through many Ofsted inspections, I'm used to subjectivity being presented as objectivity, anyway, so this caused no great alarm
>
> (g3)

> I think a review process that is not anonymised is of great concern. I refer to this in the context of my institution and personal experience. It would seem that in the human sciences where ideological stances can be obviously reflected in research, it is important for fair play that anonymity is given serious consideration. This of course, calls

into question the credibility of assessors

(g6)

The principle of ***efficient use of public funds*** got the lowest score of the eight. Respondents were concerned about opportunity costs:

Relentless, exhausting and distracting from the actual business it is meant to improve

(e3)

The REF is a lot of work, uses scarce resources and could be much simplified. I wonder if a random sampling system would be simpler with some recognition of the different outputs in different disciplines

(e14)

It seems that a lot of time and effort has gone into 'preparing' for the REF – in the sense of doing trial runs, trying to work out how to play the game. In that sense it has not been an efficient use of public funds. That said, I suspect that it has pushed some academics (including me) to produce more and better quality outputs.

(e20)

One person (g4) said simply: 'it seemed a lot less complicated in 2008'.

That duality – of cost and benefit – was replicated in comments on ***impact case studies***, which also raised other points about this major ***discontinuity***:

For those of us having to write an impact statement, it was amazingly time-consuming.

(e9)

Case studies were quite difficult to put together and to verify claims; movement of researchers between institutions meant that previous impact did not count and undermined a key case study proposed

(g5)

The process of writing the case study was fairly longwinded but the research office provided a journalist, who was knowledgeable about the field, and who turned the case study into an acceptable lively account for a non-specialist reader. A different consultant reviewed each impact case study and was not particularly useful

(e2)

> The university brought in external consultants to offer support but they knew little about the arts or social sciences and drew on the sciences as a model for assessing impact, which was unhelpful.
>
> (e6)

> 'Impact' was… in danger of becoming another example of what was countable leading the process rather than any real thought to impact – eg how can we measure how our students have taken our practices into their own professional areas.
>
> (e21)

> I was pleased to have the opportunity to work on an impact case study. This was duly included in our submission, and I also developed it as a book chapter.
>
> (e10)

Two respondents in northern institutions acknowledged benefit, though not recognition through submission:

> I can honestly say the unbelievable (sic) amount of work helped me articulate my practice more clearly and I'm certain contributed to my successful Readership application
>
> (n3)

> It was a good experience – made me think much more carefully about the work I was doing
>
> (n5)

One colleague was less enthusiastic:

> I was given the impression my work was 'second rate' research, but for the purposes of the REF case study, it 'ticked a box'
>
> (n6)

The other major discontinuity was the greater attention to equity and special circumstances. Early career researchers were allowed to submit fewer than the normal four outputs, though, as a group, there quality profile was on a par with more established staff. On those with other circumstances affecting their output, there were

> very complex arrangements regarding reduction of outputs in special circumstances and some definitions of who was to be included;

> definitions of early career researchers were not clear and not consistent with the institution.
>
> (g5)

> There was greater attention to issues of equality and diversity in the REF than in previous exercises, and this was very important. However, it also shed light on some darker corners of current practice, and these will need to be addressed …. Through my involvement in the equality and diversity panel, I learned much about the trials and tribulations that many colleagues have to deal with while still producing good work
>
> (e10)

Returning to the issue of cost-benefit and 'value for money', which triggered the first exercise in the 1980s, two professors approaching retirement were realistic and positive, though relieved it was their last time:

> I personally don't object to having to put forward outputs, to describe them and say why I consider them to have quality, value, some impact etc, etc. I know that those not included were not going to feel good about it, but there has to be a cut-off point somewhere and the need for four outputs of reasonable quality is not an unreasonable requirement … The whole thing in terms of the costs of time makes it poor use of public funds, but how else can you do this unless you just take it out of the hands of the institutions altogether and do it mechanistically – and nobody would like that either.
>
> (e12)

> I felt that the REF provided a unique opportunity to evaluate the quality of research across the entire sector. I think it's important that this should be demonstrable both within and outside the UK
>
> (e10)

Finally, in this section, I offer extracts from a thoughtful internal overview by one leader of a submission. His positives included 'it focused minds and led to discussions about 'quality', and led to a clearer set of goals (ie ends) and clarification of role diffuseness'. The lessons learned were about the need to develop teams to work on submissions because of snow blindness and workload; to emphasise enabling and enhancement as well as managing; to get structured, coherent and consistent feedback: 'we got contradictory expert advice … feedback

from the strategy group was contradictory and unhelpful: demotivating for some staff. We need critical friends who engage with the department over a long period'. This might allow more recognition of learning from cognate departments, but also nuances across the disciplines.

Outcomes

When results came out, there had been a big leap in the quality profiles across all subjects (Jump, 2014). *Units of Assessment* were given an overall profile from 4* (world-leading), through 3* (internationally excellent), both of which then lead to funding, to 2* (internationally recognised, whatever that means), 1* (nationally recognised) and U, unclassified. The proportions in the top two grades across the UK's four systems were 30 per cent and 46 per cent; in the 2008 RAE, they had been 17 per cent and 37 per cent (REF, 2014c). In 2014, 'impact' replaced 'esteem' as a contributing criterion and the balance across the three elements – output and environment remained from 2008 – was set as a constant across all panels, rather than varying by panel. Between the two exercises, the Grade Point Average (GPA) rose from 2.56 to 3.01 in the *Times Higher Education (THE)* calculations (Jump, 2014). So, *prima facie,* there had been a significant leap in overall quality. Up to a point. Given that the ratings are a *profile of distribution* of submissions, some of the 'improvement' came from greater selectivity. The number of units submitting went down by 15 per cent. Despite an increase in the numbers of academic staff employed across the sector, the numbers submitted fell, except in physical sciences. These reduced numbers submitted fewer outputs per head – 3.41 on average, from 3.75 (REF, 2014a) because of greater allowance for extenuating circumstances. The average proportion of *outputs* rated 4* went from 14 to 22 per cent, an eight point increase, rather than the 13 points (from 17 per cent to 30 per cent) shown by the overall grades. Much of that change can be attributed to a reduction, of over 12.5 per cent, in the number of outputs submitted (REF, 2014a), mainly by cutting off the tail of the quality distribution. The difference between the output profile and the overall profile relates to the other two factors. Whereas 22 per cent of *outputs* were rated 4*, for *impact* the figure was 44 per cent and for *environment* was 45 per cent (REF 2014b). Since these two factors contributed 35 per cent weighting to the final score, the only direction was up. The effect was greatest in medical and life sciences, where units had an average 60 per cent rating in the two factors (Else, 2015).

The rest of this section looks at Education to examine those factors more closely and to raise issues of general concern in any similar exercises.

Research in Education is contracting: productivity is declining, if REF statistics reflect reality in a balanced way. There was a net reduction of 6 in the number of submissions. In Wales, Cardiff is the sole representative. It will, therefore, get all the QR funding informed by REF ratings. None of the other units in Wales from 2008 submitted in 2014. I think such a monopoly is unhealthy and the Assembly government may need to intervene. FTE staff submitted totalled 1,442, down 15 per cent from 1696 in 2008 which was itself a 15 per cent reduction on 2001. Outputs submitted went down by 23 per cent. Modern universities were more likely to submit more staff, in England 15 per cent more, indicating an investment in developing research from a low base. Most pre-92 universities, with a longer research tradition, made reductions in numbers submitted, to add to those that had done so in 2008. Some cases were drastic in the extreme and illustrate the skewing of results by controlling the sample:

> Nottingham reduced numbers by 55 per cent from 51.2 to 24.6. That loss of 26.6 FTE increased the proportion of FTE with 4* activity by 3.3.
>
> Scotland saw the severest cuts: Edinburgh submitted 45 fewer (-54 per cent); Glasgow 25 (-40 per cent); Strathclyde nearly 30 (-55 per cent) – a round total of 100. The gain in FTE allocated to 4* overall activity rating was ... 11.

So, 'excellence' may be achieved, but the price is exclusion of much of the academic workforce.

The scale of research activity is still considerable – 3,625 doctoral degrees awarded in the review period, the highest per FTE staff within the social sciences. Many would be from EdD programmes (figures are not recorded separately). External research income totalled £292m, or over £200,000 per FTE staff.

The overall quality gap between units in pre- and post-92 universities widened, because of the different approaches to inclusion and the different ratings for impact and, in particular, environment. I have long campaigned for recognition of relevance as a criterion of quality, and have no complaints about the use of impact ratings. Three units – Durham, Nottingham, Sheffield – scored 4* on impact for 100 per cent of their work, and Nottingham had a double top with a matching score on environment, along with six others. All were in pre-92 institutions. Nottingham's 4* rating moved from 31.2 per cent for outputs to 55 per

cent for overall activity; the Open University moved even more – from 16.8 per cent to 38 per cent at 4*. Even at the bottom end, Sunderland had no 4* outputs, but had 5 per cent of its overall activity given that top grade, because of its impact score.

Many modern universities moved the other way. The most poignant contrast came from alphabetic juxtaposition. Staffordshire had a 25 per cent output rating at 4*; just above it on the list Southampton had 18 per cent. When impact and environment were factored in, Staffordshire went down to 16 per cent, while Southampton went up to 31 per cent. So, researchers produce good work, but suffer because of other factors. Staffordshire had had that before: I was on the subpanel awarding their grade in 2001, when their rating was held down to a 2 (on a different scoring system) because it was a first submission from a new unit set up as a directorate research think-tank on access. Its output, in my view, was of clear 3 quality, but I was outvoted on the final judgement. Later, a traditional department was given a higher grade than its output merited because it had a good history, so the evidence was seen as a blip. I believed then, and still do, that performance should be moderated by context, but in the opposite direction, as is done in considering contextual factors affecting A level results for university admissions. The burgeoning quality of the work at Staffordshire was shown in 2008 when over 35 per cent of activity was judged as 3* or 4*; in 2014, over 70 per cent of its output was at least 3*, with 25 per cent world leading. That has been achieved with little encouragement through the national evaluation exercise. Should not good work in a less supportive context be rewarded, or at least incentivised, more than similar work with lots of institutional resource and reputational capital?

Or was there a change, not continuity in standards, to skew the overall ratings? For environment, in 2008, the top score for 4* was 75 per cent. Only 5 institutions had a 4* score of 50 per cent or more. In 2014, 18 units scored 50 per cent or more and 8 scored 100 per cent at 4*, a perfect environment. Only two modern universities scored more than zero at 4*: that is where the main difference comes to expand the ratings gap.

For esteem, in 2008, the top 4* rating was 40 per cent, achieved by only 2 units, 11 more scored 30 per cent. In 2014, for impact, the replacement measure, 11 units scored 40 per cent, 13 between 50 per cent and 90 per cent and 3 scored 100 per cent. More perfection: it appears that the criteria for esteem by academic peers may have been more demanding than those set by impact assessors.

All those getting top ratings for environment were large departments, submitting more than 20 FTE staff for assessment. The link between size

of the research active group and a 'well-found' department is clear. That is an issue for the post-92s with a major commitment to teaching and teacher training, where, for many, research is squeezed into corners of the timetable or squeezed out into private time. That, in turn, affects the scope and scale of research that can be conducted. For the third exercise running, the panel feedback emphasised the need for more large-scale, longitudinal, cohort studies using quantitative methodologies and sophisticated data sets, and extolled the virtues of the growing work using such approaches (HEFCE, 2015). That characterisation of quality work is beyond the reach of many small units in modern universities (who had little representation on panels) since it needs a critical mass of people to form a strong team.

Conversely, work close to professional practice drew criticism. In teacher education 'the majority of studies were qualitative and often small scale…not strongly theorised'; in secondary education 'outputs on pedagogy, teaching and learning…spanned the spectrum of quality, …however, some of the work submitted was characterised by small-scale work of modest analytical or theoretical rigour', just as in primary education, 'some studies, close to practice, lacked originality, significance and rigour' (the three criteria for judgement). The feedback report confuses adult education and higher education – not good! – and issued a warning about work in further, adult and vocational education: 'the sub-panel noted the lower than expected volume of submitted outputs in these areas, and it was felt that this may reflect a longer-term decline in major research investments with a programmatic focus'.

So, mode 2 research, starting with real world problems and seeking to solve them is rated lower than strongly theoretical work. Yet many professionals in teaching do not come from a background strong in a social science discipline. They bring expertise in practice. In further and higher education they may have had high standing in another profession and seek to develop their successors. The REF does not encourage developing research from such professional bases.

We have noted the impact of impact scores in the profile calculations. Yet, there are only three paragraphs, out of 56, on impact in the feedback report. Theory seems more important: a strong discipline base seems essential; pragmatic but rigorous grounded theory does not feature. Nor action research, nor professional reflection, nor evaluation studies, and paradoxically, the criteria for assessing impact specifically exclude impact on teaching, though not *through* teaching, which may have to be re-labelled 'knowledge transfer'.

The best submissions 'specified the context of their work in terms of *potential* stakeholders in research outcomes…were able to tangibly

describe the pathways to impact *to which they aspire* and the infrastructural support available to staff in achieving impact … targeted staff development, promotion policies, publication of research summaries, creation of media strategies'. It reads like intent not actuality; more like selling a product, post-production, not sharing a project; an add-on, which was the concept underpinning consultations, not an integral part of design and delivery.

User members and impact assessors 'took away new knowledge, contacts and learning', but raised queries about the value for money of such a time consuming exercise'.

Conclusions and recommendations

The criteria for quality and the discipline based preferences of the panel stack the odds against those working with professionals in fields such as education, health, social work, even law and architecture. But those links and partnerships are strengths in a factor carrying a 20 per cent weighting. If an 'impact culture' can be made fundamental to individual and departmental strategy, that may also help in claiming a more positive environment. At the moment, weaknesses in these two areas drag down the overall profile of most small submissions, mainly in modern universities. Such units are unlikely to attract world class researchers: that is not the way the world works. The threshold for entry to the assessment of impact is 2* underpinning research, and in most cases this was 'comfortably passed', according to the panel. There are longstanding relationships with users of our output of graduates, who may also have offered training placements to students, and research in partnership with them can build on those to provide added value, an enhanced service, an arena for testing results and a bridge to implementation and impact, developing case studies of impact for assessment in a five year horizon. So, building impact in as a project feature at the design stage, and working in partnership with those who will implement research findings in contexts where impact will be evident *and measurable* is one approach to countering the bias in the exercise.

The more drastic approach is to see such partners and stakeholders as the key people from whom we seek 'fame and fortune', as one former VC summarised the aims of research assessment. If working with them enhances reputation (fame) and even generates some funding (fortune) through a consortium of organisations with common agendas, that may be more beneficial, in reward for effort than playing a game that cannot be won. Their staff would be involved in defining projects where evidence informed practice is needed. They can also be involved as co-workers enhancing partnerships, and with university staff giving support

without being overloaded by a one-person project Perhaps, for REF 2020, there will be even fewer submissions, and the downward spiral will continue, while another activity domain, linked to professional practice, develops, to create a binary funding system for an under-rated form of research.

Several learning points arise from this small study:

There needs to be a fuller, clearer statement about the ends of the REF, within the guidance on the REF, not on funding council general pages on funding, since the purpose and funding are seen as inter-twined. This should set the exercise within objectives for overall research and development policy, and links with other activities, such as teaching or regional development.

Provision for research quality assessment should put more emphasis on enhancement – a formative and forward-looking approach to developing capacity and capability as well as quality.

The process needs to be better led and managed, at both system and institutional level to overcome widespread feelings of depersonalisation by staff involved, which can lead to de-motivation, risks disengagement and may result in lower productivity and poorer quality.

Varieties of excellence need to be recognised. Research is done for a variety of reasons, with different objectives and different means/methods deployed. If 'fitness for purpose' is a core criterion, small scale local partnership approaches may be fit means for some of the ends identified purpose. The BERA/RSA report (2014) comes to a similar conclusion.

That recognition should be used to re-balance the distortion that successive UK exercises have introduced in promoting an isomorphic model of research quality and its outputs (Adams and Gurney, 2014). That may 'change behaviour' in a positive way.

The relationship among the three elements of evaluation needs reviewing.

Address for correspondence
Ian McNay, Centre for Leadership and Enterprise, University of Greenwich, London, UK. email: i.mcnay@gre.ac.uk

References

Adams, J and Gurney, K (2014) *Evidence for Excellence: Has the signal overtaken the substance?* London, Digital Science

Australian Research Council (ARC) (2014) *The Excellence in Research for Australia (ERA) Initiative*, www.arc.gov.au/era/ Accessed January 2015

BERA/RSA (2014) *Research and the teaching profession. Building the capacity for a self-improving education system,* British Education Research Association

Boyer, EL (1990) *Scholarship reconsidered: priorities of the professoriate,* Jossey-Bass/Carnegie Foundation for the Advancement of Learning

Edgar, F and Geare, A (2013) 'Factors influencing university research performance', *Studies in Higher Education,* 38(5): 774-792

Else, H. (2015) 'Don't forget humanities when dividing REF spoils', *Times Higher Education,* 1 January 2015: 6-7

Franco-Santos, M, Rivera, P and Bourne, M (2014) *Performance management in UK higher education institutions: the need for a hybrid approach,* London, Leadership Foundation for Higher Education

HEFCE (2009) *Research Excellence Framework: second consultation on the assessment and funding of research*, HEFCE 2009/38, Bristol: Higher Education Funding Council for England

HEFCE (nd) Mainstream QR research funding: note on QR data. www.hefce.ac.uk/whatwedo/rsrch/howfundr/mainstreamqr/ Accessed 3 December 2013

HEFCE (2014a) 'Feedback on REF 2014', Letter from David Sweeney and Professor Anne Greenhough, 18 March 2014

HEFCE (2014b) *REF 2014 shows UK university research leads the world*, News release, 17 December. Bristol, HEFCE.

HEFCE (2015) 2014 *Research Excellence Framework (REF) reports published*, News release, 27 January. www.hefce.ac.uk/news/archive/2015, accessed 28 January 2015.

Jump, P (2014) 'It may not be a revolution, but impact's effect is felt', *Times Higher Education,* 18/25 December 2014: 6-7

Kallio, M and Kallio, TJ (2014) 'Management-by-results and performance measurement in universities – implications for work motivation', *Studies in Higher Education,* 39(4): 574-589

Leathwood, C and Read, B (2013) 'Research policy and academic performativity: compliance, contestation and complicity', *Studies in Higher Education*, 38(8): 1162-1174

McNay, I (2003) 'Assessing the assessment: an analysis of the UK Research Assessment Exercise, 2001, and its outcomes, with special reference to research in education', *Science and Public Policy*, 30(1)

McNay, I (2009) 'Research Quality Assessment: objectives, approaches, responses and consequences', in Brew, A and Lucas, L (eds) *Academic research and researchers,* Maidenhead, SRHE/Open University Press

McNay, I. (2014) *Ends and means in RQA*, Presentation to the HE Close Up conference, Lancaster, June, and London International Conference on Education, London, December

Ministry of Education (2013) *In pursuit of excellence: analysing the results of New Zealand's PBRF Quality Evaluation*, Auckland, Ministry of Education

Ministry of Education (2014) *Review of the Performance-Based Research Fund,* Auckland, Ministry of Education

Rebora, G and Turri, M (2013) 'The UK and Italian research assessment exercises face to face', *Journal of Research Policy* 42(9): 1657-1666 http://dx.doi.org/10.1016/j.respol.2013.06.009

REF (2014a) Submissions, www.ref.ac.uk/subguide accessed March 2015

REF (2014b) REF2014 Key Facts, www.ref.ac.uk accessed January 2015

REF (2014c) *Results and submissions*, http://results.ref.ac.uk/. Accessed December 2014

University Grants Committee (2012) *Research Assessment Exercise 2014: Guidance Notes*, Hong Kong, University Grants Committee

Yokoyama, K (2006) 'The effect of the research assessment exercise on organisational cultures in English universities: collegiality versus managerialism', *Tertiary Education and Management,* 12(4): 311-32