On promoting rigour in educational research: the example of the RAE?1
Ian Stronach

With a response from Margaret Brown (RAE 2008 Chair for Sub-panel K45: Education)

Abstract

This article offers a deconstruction of the RAE Education sub-panel’s rubrics, drawing also on the broader RAE regulations, procedures, and associated documentation and research. It seeks to tease out the sorts of covert epistemologizing that may (or may not) be likely to take place. The theoretical ambition is to take a Derridean approach to acts of ‘translation’, reworking that metaphor as a way into the acts of translation that the RAE undertakes in placing numerical values on the quality of research outputs. How do we measure our ‘pounds of flesh’? As a practical ‘hermeneutics of suspicion’ it seeks a ‘provocative validity’ in that its aim is to provoke clarity and reassurance from the authors of such RAE documents. In that ambition it has had some success, as the response from Margaret Brown, Chair of the Education sub-panel, demonstrates.

‘One shouldn’t complicate things for the pleasure of complicating, but one should also never simplify or pretend to be sure of such simplicity where there is none. If things were simple, word would have gotten round, as you say in English.’ (Derrida, 1988:119)

One might argue that the ‘simple’ word did get round, in relation to the UK RAE excellence criteria:
<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4*</td>
<td>Quality that is world-leading in terms of originality, significance and rigour.</td>
</tr>
<tr>
<td>3*</td>
<td>Quality that is internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence</td>
</tr>
<tr>
<td>2*</td>
<td>Quality that is recognised internationally in terms of originality, significance and rigour.</td>
</tr>
<tr>
<td>1*</td>
<td>Quality that is recognised nationally in terms of originality, significance and rigour.</td>
</tr>
<tr>
<td>U/C</td>
<td>Quality that falls below the standard of nationally recognised work. Or work which does not meet the published definition of research for the purposes of this assessment.</td>
</tr>
</tbody>
</table>

The purpose of this paper is to deconstruct the categories, criteria and rubrics of the RAE particularly as they apply to Panel K, sub-panel UoA 45, Education. It is hoped that such a critical consideration can help participants of all shades enter a debate that will inform decision-making and critique. It is also a hope – though less likely - that such reconsiderations can help inform future policy development in the area of research performance appraisal more generally. It should be stressed, for the benefit of those unfamiliar with ‘deconstruction’, that such an undertaking seeks out the unspoken, the implicit and the contradictory in these rubrics. In so doing it aims for a ‘provocative’ validity, and the performance of alternative interpretations. These do not constitute a new bedrock on which appraisal can confidently be founded. Instead, they unsettle conventional readings in ways which open up new understandings and richer appreciation of the dilemmas of this sort of appraisal, in terms of its ambitions to ‘correspondence’, ‘consensus’ or – as I will eventually argue – ‘translation’, drawing on Benjamin and Derrida.
To begin at the beginning, the criteria seem to have emerged more from a political than an academic process. In identifying criteria that promised to distinguish ‘world-leading’ research (4*) from ‘internationally excellent’ (3*) and work ‘recognised internationally’ (2*), they raised the issue of ‘internationality’ in relation to research. They did so in regard to three categories of evaluation – ‘originality, significance and rigour’. These were the givens of the RAE process.

They reflected government perceptions that previous RAE appraisals had been subject to credentialist inflation, and so replaced the old grades (1,2,3b,3a 4,5,5*)². In specifying three grades of internationality, the government met grade inflation with criterial hyperinflation.

First of all, there are a number of technical and logical points that need to be made. The criteria were incompetent as such in that the descriptors gave no indication as to how they might be measured or defined. Subsequent attempts by Panel K to offer ‘expanded definitions of the quality levels’ ( merely offered further undefined attributes: ‘ highly significant contribution’; ‘significant contribution’; ‘recognised contribution’, and the loser’s 1* of ‘limited contribution’. The ‘expansion’ amounted to the displacement of one unspecified performance vocabulary with another. At the same time the tautological relation is obvious: what’s world-leading? Answer: whatever’s highly significant. What’s highly significant? Whatever’s world-leading. The ‘expansion’ of the quality levels, in short, is ‘gaseous’ in its rhetorical nature (Novoa & Yariv-Masha 2003). Finally, the envisaged system was normed in relation to an absolute rather than a distribution: ‘The definition of each quality level relies on a conception of quality (world leading quality) which is the absolute standard of quality in each unit of assessment’ (RAE 01/2005), my emphases). So quality ‘relies on’ quality and ‘is’ an absolute in and of itself, shortly before disappearing up its own benchmark. Indeed, this is the ‘tautology of redundant specification’ (Chambers 2003: 181).

Second, there is an oddity about the categories of excellence, ‘originality, significance and rigour’. The first two sound like maximum competencies, but the third suggests a minimum competence which makes peculiar sense if one tries to align it with the evaluation criteria: what would ‘world-leading rigour’ look like? Would it be a good thing, or would it imply the rigidity of a rigor mortis?

Third, research quality was to be appraised in terms of ‘originality, significance and rigour’. This was the generic phrase across the generic RAE documentation and main panel rhetorics³. It is
noteworthy that Panel K is unusual in that it sometimes reverses this order and refers to ‘rigour, significance and originality’ (p19, 32, 46). The Education sub-panel followed suit (UoA draft, p5; Education sub-panel document p23). There is no such reversal in related subjects (eg UoA 41, Sociology). The Education sub-panel also spells out its take on ‘rigour’ quite closely in offering guidance on how each piece of research might present itself to the panel – ‘in particular how the criterion of rigour is met’, along with ‘methodological robustness’ and ‘systematic approach’. Without being over-suspicious, such words as ‘systematic’ ‘robust’ and ‘rigour’ belong most readily to neopositivist forms of research narrative that many have criticised as ‘scientism’ rather than Science (Erickson & Gutierrez 2002). One might wonder if the panel is expressing, or being required to express, a coded predilection hereabouts. The sub-panel is clear elsewhere, however, that there have to be methodological horses for courses even if it offers this reassurance in a subordinate clause:

‘...but rigour can best be assessed on a case by case basis using whichever dimensions are most appropriate’

Other, more science-related panels are less concerned with ‘rigour’. Engineering does not mention the term and spells out its priorities in terms of advancing knowledge and understanding, originality and innovation, impact on theory, methodology, policy, practice etc. Physics expects to see ‘some of the following’: ‘agenda-setting, research that is leading or at the forefront of the research area, great novelty in developing new thinking, new techniques or novel results ..’ (p42). Again, no mention of rigour. Linguistics on the other hand, takes that term and disseminates it across a whole range of research qualities: ‘intellectual coherence, methodological precision and analytical power; accuracy and depth of scholarship; evidence of awareness of and appropriate engagement with other work in the field or sub-field’. So the notion of ‘rigour’ can either be dropped or exploded, and by disciplines more readily called ‘Sciences’ than educational research or indeed Psychology.

Whence the Education sub-panel’s additional concern for ‘rigour’, and the special emphasis expressed in Panel K? Is this the semi-sciences reflecting the same obsessive concern for respectability as the semi-professions? Put provocatively, is ‘Science’ mobilised as Rigour within the RAE sub-panel process in order to make an honest woman of Educational Research?
Thus far, in relation to ‘rigour’, we have some straws in the wind – no more – a shift in word order, a foregrounding in terms of specification, a hypothetical bias towards particular ‘scientific’ approaches to social inquiry – a possibility noted by Armstrong & Goodyear (2005). The rest of this paper seeks to say more about the location and import of this notion of ‘rigour’ in the hope that its opening-up will lead to more clarity for those who orchestrate, execute or endure the RAE process.

There is a research basis for the RAE categories of appraisal. Wooding and Grant (2003) conducted research on behalf of HEFCE across a broad range of disciplines in order to establish the key categories against which the research community would wish to see their work appraised. They did so via a number of workshops spread across the country. It is reasonable to assume that their work is sufficiently representative across most disciplines concerned in the RAE. In their Executive Summary they reported that there was a consensus that high quality research was based on ‘rigour; international recognition; originality; and the idea that the best research sets the agenda for new fields of investigation’ (p3, my emphasis). So there we have a vindication of Panel K’s and the Education sub-panel’s re-prioritisations – ‘rigour’ is the paramount concern.

But when we examine the full research text of which the Executive Summary claims to be an accurate summary, we find the following statement:

‘The concept of defining the research agenda by framing new research questions and advancing a field into new areas was seen as the most important characteristic of high quality research’ (p14, my emphasis).

The enumerations the authors give in Fig 11 confirm this. The priorities are reported in terms of respondents’ views as (1) defining the research agenda (99 attributions), (2) rigour (71 attributions), (3) international recognition (66 attributions). So rigour is not the leading requirement, as their executive summary claims. Struck by this slip, I thought that an examination of the research annexes might be interesting. ‘Defining the research agenda’ (DRA) had been one of the HEFCE researchers’ emergent categories, and items like ‘advancement of the field’, ‘potential to move discipline forward’ had typically been included. But an item ‘advance body of knowledge’ was attributed to a separate category called ‘scholarship’, which seemed indefensible since it fitted perfectly into DRA – ‘advancement of
the field’ and ‘advance body of knowledge’ are synonymic. Again, the category of ‘rigour’ had been awarded the response item ‘depth’ – more reasonably attributable to the notion of ‘originality’. Finally, there was a remarkable accident to one of the workshop’s appraisal of the importance of ‘originality’. Although originality had been highly rated by other workshops, there was a zero score for one workshop. It seemed unlikely that a group of academics would be so uninterested in ‘originality’. Further exploration revealed an explanatory footnote:

‘The sticky hexagon fell off the chart and was not available for voting’ (p5).

A case of the ‘hanging chad’, as a colleague remarked. An averaging of the other workshop scores for that category (which would have been 7.75) was not undertaken, and instead the presumption of a 0 score was made. Taking these sorts of anomalies into account a re-analysis of the actual workshop data would read as follows:

Defining the research agenda – 120
Originality – 69
International recognition – 66
Rigour – 62

So the category of rigour was 1st in the Executive Summary, 2nd in the main text, and 4th in relation to the data in Annex II. ‘Rigour’ was a shifting if not shifty signifier, whose analysis had certainly not been rigorous. Now, it is almost always a good idea in relation to UK governance to adjudicate cock-up/conspiracy theories in favour of the former, but it does seem possible that ‘rigour’ might be experiencing some unwarranted promotion.

Given the Education sub-panel’s unusual concern for ‘rigour’ it is worth raising some questions that the panel might want to consider and even address publicly, in a spirit of reassurance. They are: is there a particular promotion of the notion of ‘rigour’? Is that promotion, if such it is, part of paradigmatic prejudice? Will ‘rigour’ as a minimum competence be used to police the other two categories? Are covertly hegemonic moves being made in relation to RAE documentation?

In relation to all of the above questions, it is relevant to note that the panel has detailed requirements for members to declare interests. All of these are ‘familial’ in nature – same
institution, former student, collaborator etc: none concerns paradigmatic prejudice, and yet as a BERJ editor of 10 years standing, I know that such bias is far more likely than any other. So why that gap? This is a particularly acute omission when one considers how explicitly other panels have marked out excellence as a matter internal to specific fields and concerns:

‘In view of the diverse nature of the discipline of sociology, the sub-panel understands the quality descriptors to relate to indicators within fields, sub-fields and cognate areas’ (UoA 41, p47).

Philosophy too invokes immanent criteria: ‘..will judge submissions against the best work in the field’ (p40)

To return to the specifics of the Education sub-panel’s guidelines, they are unusual in that they offer a very specific example of what might count as a justifying rhetoric for any single publication (150 words are permitted in order to provide evidence of the claim to originality, significance and rigour). It reads as follows:

‘Hypothetical Example:

Humanities in primary schools – short booklet containing advice to teachers and policymakers based on a synthesis of international research. This provides an innovative conceptualisation of the field and has been referred to by the TDA for Schools (2007) as the basis for its criteria for CPD in this area. The review considered 1250 references of which 41 met the criteria for inclusion. The full review has been accepted by Springer in 2008. The distillation of the implications of the literature was done by a working group of five researchers and five teachers. The draft was refereed by two international referees and piloted by ten teachers and four policymakers to ensure it was appropriate and user-friendly. It is cited by two researchers because of the considerable work involved in its production; they contributed equally and co-directed the project, financed by a £15,000 grant from Tesco.’ (p32)
There are a number of points to be made here. First, this is the only example. In the draft, there were five such examples (UoA 45, draft as at 16.7.05, p4-5), and one wonders why that plurality was narrowed down, making the example much more likely to be read as an exemplar or even as a template. Second, the earlier draft had taken a very different tack – emphasizing contributions to theory or methodology, prestigious keynotes, international awards and so on (ibid., p4). Would it be unreasonable to envisage the heavy hand of the State intervening in this peer-led process of consultation, promoting itself and its agencies as the User/Methodologist Who Mattered? Third, the narrative is almost exclusively focussed on local and national users, and on outlining the detail of a research process in accordance with one particular notion of rigour. The evidence for originality is very limited (a claim of ‘innovative conceptualisation’), and the significance is clearly circumscribed by the national. So that’s a clear 1*? Yet, it would be a foolish reader who thought that it was included to illustrate the bottom end of the range. It is also hard not to take the methodological orientation of that narrative and associate it with the foregrounding of a particular approach to educational research with its characteristic notions of rigour, robustness, and systematicity - the features we earlier wondered about in relation to panel predilections. If we then ask ourselves ‘what kind of writing is this?’ then there is a simple and clear answer. This is the kind of ‘structured abstract’ that Sebba has been calling for since 1998, and which is currently based on an extension of the work of Hartley (2003). It performs and validates a kind of neopositivism that has been quick to dismiss almost all other approaches to educational inquiry as invalid and above all, lacking in rigour. As in the fictional abstract, they fail to meet ‘the criteria for inclusion’. And it is ‘rigour’ that demarcates that exclusion.

To conclude, the categories and criteria for RAE appraisal in educational research are far from clear. There seems, prima facie, to be cause for serious concern in relation to the promotion of certain kinds of research and certain kinds of outcome. No doubt, the panel will argue that they are their own masters and will come to their own judgements- and the element of judicious peer review and consensus will of course be present. Indeed, three of the Education sub-panel members said so when this paper was presented at BERA, September 2006. Provided they make full disclosure of paradigmatic prejudices, no one will doubt the panel’s integrity. But the discourses around which they construct their deliberations also surround them. The extent to which they will be prisoners, warders or governors remains to be seen: escapees they will not be.
This brief account began with the claim that the orienting metaphors of appraisal depended on notions of ‘correspondence’ and ‘consensus’. In the shortest of shorthands these represent different epistemologies – one could plausibly toss Hammersley into the first box and Habermas into the second, albeit not with the same reverence. But both projects suppress a kind of impossibility. All translations, to move to our third metaphor of evaluation, express what Derrida has called ‘an economy of in-betweenness’ (2001: 179). In discussing The Merchant of Venice, Derrida argues that the economy of translation (ducats/pound of flesh; christian/jew) circles through both a ‘proper meaning’ and a ‘calculable quantity’ (ibid). Letter and spirit are irreconcilable yet inevitable and necessary to each other:

‘This relation of the letter to the spirit, of the body of literalness to the ideal interiority of sense is also the site of the passage of translation, of this conversion that is called translation’ (p184).

The RAE process is just such a plural act of conversion/translation. Quite explicitly, it re-expresses the local as a globally normative value, one that invites international appraisal, investment and emulation. That is one version of the fantastic ‘passage’ of translation. It is part of the global phenomenon of ‘comparison, defining a new mode of governance’ (Novoa & Yariv-Mashal; 2003: 428; Stronach 1999). It involves the incommensurable translation of a ‘field’ into a ‘harvest’, a transubstantiation that is no less mystical than that of ‘bread’ or ‘wine’ in Christian ritual, to return to Benjamin’s analogy for the notion of ‘translation’. In addition, it effects the parallel construction of the ‘self-auditing academic’ (McWilliam 2004: 162), and with it another ‘passage’, this time of identity rather than status or nature. Of course one could go on, but already it can be seen that the singular body of a spirit of inquiry is thus translated into sundry commodities of appraisal – individual, institutional, disciplinary, national - and various ‘pounds of flesh’ are weighed and translated in terms of an ‘impossible but incessantly alleged correspondence’ (Derrida 2001: 184). These require certain kinds of passage which cannot be legislated for, any more than law can guarantee justice in any single case. Justice, like mercy, is above the law, and can have no a priori regulatory transparency, any more than can virtues such as tact. The future RAE moves to ‘metrics’ will inevitably make even more injudicious such translations, by obscuring the ‘objectified subjectivity’ of the exercise (Velody, cited in Bence 2005: 151; see also von Tunzelman & Mbula 2003: 15). The translation of our pounds of flesh into RAE ducats is inherently incommensurable, and that is the scandal of assessment which the RAE and even more its future metrification seek to suppress. If such appraisal has an essence it is the absence of all possible rigour.
'If the problem, the *aporia*, in any of these cases is resolved not through *experience*, through the “ordeal of the undecidable”, but through recourse only to calculation or a formula – which is always somebody’s formula – then it will have been a sell-out, a set-up by and for one economy, “fixed” from the start’ (Derrida, cited in Davis 2001: 95; see also Derrida 2005, Richter 2002).

Derrida ends his ‘translation’ argument concerning Shylock and *The Merchant of Venice* by arguing that in the end it is mercy (‘It droppeth as the gentle rain from heaven’) rather than justice for which we ought to pray. Maybe that’s right.

‘Though justice be thy plea, consider this,
That in the course of justice none of us
Should see salvation: we do pray for mercy
And that same prayer doth teach us all to render
The deeds of mercy.’ (*The Merchant of Venice*, Act 4, Scene 1)
On promoting rigour – a response
Margaret Brown (RAE 2008 Chair for Sub-panel K45: Education)

I enjoyed Ian’s paper although I think he sometimes sees conspiracies where I don’t believe they exist. I will not respond to all his points – indeed there are several with which I agree – but only try to explain where I think he might be on the wrong scent.

I should make it clear that I am writing in a personal capacity and nothing in this response should be taken as either having the agreement of the rest of the subpanel or any official status at all.

Most academics do a great deal of assessment – of articles, research bids, promotion applications, student essays, etc. - and are thus familiar with its inherent imprecision and with the difficulty of both devising and applying general criteria e.g. to arrive at essay marks. Some BERA members, like Dylan Wiliam and Harvey Goldstein, are experts in educational assessment and can tell us rather more, for example that standards tend to be developed and maintained by communities of examiners and that the complexities of shared understanding among such practice communities cannot readily be encapsulated in written prose of a general nature. There is also some evidence that in GCSE examining, criterion-referencing is less reliable than norm-referencing, i.e. even when an experienced group of examiners seem to be agreed over where to set grade boundaries, it is actually more reliable in the sense of maintaining consistent standards across year-groups to ask them to put the papers in order and to tell them roughly what percentage of candidates to put in each grade rather than leaving it to their professional judgement.

It is perhaps a pity that HEFCE did not consider making use of such assessment expertise in setting up the RAE methodology; but what we have is nevertheless a familiar model in academia of grading against fairly vague criteria. In some disciplines there may be greater clarity in status of publication routes than there is in education, but in few are the distinctions likely to be easy to make. In evolving our own gloss on these in the guidance specific to Education (Panel Criteria and Working Methods: Panel K, Ref RAE 01/2006(K)) we initially hoped we might make them more precise and meaningful, but with such a broad field and time for nothing more than a quick dry run on a varied set of articles submitted in 2001, as Ian has detected it proved impossible to realistically go much further in describing the boundaries
between the categories. It would be much easier and probably more honest to norm-reference, working to a fixed percentage in each category, but that would not be politically acceptable as the percentages would either have to be common across disciplines, obscuring any differences of standards in research quality between them, or differences would have to be predetermined based on insufficient evidence.

On the matter of the three given characteristics of rigour, significance and originality, the sub-panel have generally been happy to accept that these are useful indicators, not perhaps surprisingly as they are often used in relation to other academic judgements. Ian claims to detect a bias towards rigour, and what’s more to a neopositivist interpretation of rigour. I do not accept that the three characteristics are different in type, in relation to maximal or minimal, and I can think of outputs which I would regard as exceptionally rigorous in different research paradigms (the subpanel for example in one of its practice exercises has recently agreed that an overseas publication in philosophy of education deserves a 4* rating on the basis of its outstanding rigour as well as because of its originality and significance). However Ian rightly picked up our suggestion that rigour is probably a key hurdle for recognition of a research output as qualifying for the 1* category. However apparently original or significant, we would not expect to give much recognition to research which is clearly sloppy to the extent that any conclusions are likely to be suspect, whether that is interpreted as providing a poorly defined and incoherent argument, failure to search out and understand relevant work, lack of clarity in its design and implementation, incorrect use of procedures or measures, or any other indication of lack of rigour which might be appropriate to different paradigms. (Science subpanels incidentally may feel it is less necessary to draw attention to the rigour criterion than the social sciences as their methodologies are more standard and may better embody it; the linguistics expansion would certainly be appropriate to some educational research).

However if the work is almost completely lacking in originality, or if the issue is too insignificant, it might not make the first hurdle either, however great the rigour. At higher levels it seems likely that significance and originality will have more salience, but highly graded outputs are likely to shine on all three characteristics. Thus I do not believe we are giving overmuch attention to rigour above the other two.

It may be that what we say about rigour can be interpreted as having a neo-positivist bias, perception being in the eye of the beholder, but Ian does quote selectively to prove his point -
evidence of a certain lack of rigour? For example he quotes the phrase ‘methodological robustness’ as evidence of a scientific paradigm whereas the actual phrase was ‘methodological and theoretical robustness’, which hopefully puts a rather different spin on it. Perhaps I could quote the start of the paragraph from the subpanel guidance on rigour up to the full sentence that Ian quotes so that others may judge:

‘Rigour can be judged in many ways, and can helpfully be associated with methodological and theoretical robustness and the use of a systematic approach. It includes traditional qualities such as reliability and validity, and also qualities such as integrity, consistency of argument and consideration of ethical issues. It certainly entails demonstrating a sound background of scholarship, in the sense of familiarity and engagement with relevant literature, both substantive and methodological. Different dimensions of rigour will be important in different types of research….’ (para. 28)

I can only assure everyone that no paradigmatic bias was intended; indeed a sub-group of the subpanel representing different research paradigms devised and checked the text to try to avoid this. (Incidentally if Ian studies the composition of the subpanel he is unlikely to conclude that there is evidence of a paradigmatic bias, I think, although I don’t promise to take up his suggestion of declaring our collective or individual prejudices.) I cannot state too strongly our intention to be inclusive and to accept a wide variety of research styles, judging each on appropriate criteria of quality. If no-one on the subpanel is qualified to make a judgement we will either cross-refer to another subpanel or appoint a specialist adviser. We will also carry out our own cross-moderation processes, using both qualitative and quantitative methods, to check as far as we can that no particular paradigms are receiving higher or lower ratings. It won’t be a perfect process, but it will be the best we can do given the constraints. Part of the incentive for improving it is of course that our own outputs will be part of the assessment exercise.

On the issue of the example of the up to 150 words of textual description, it really is a mistake to read too much into the nature of this. The generic reason for allowing additional text is to enable authors of non-traditional outputs (e.g. artistic or engineering artefacts or, in the case of education, teaching materials or publications aimed at practitioners) an opportunity to suggest how the output met the criteria for research. It was for this reason that an example was chosen of a practitioner-related output which was unlikely itself to contain a description of its methodology. Ian is right that in the previous draft we included a greater variety of examples
but each description was limited to 50 words; for various reasons it was decided to expand the allocation to 150 words which meant that in the final draft we could only squeeze in one example. Purely for convenience and to save my having to call on other subpanel members, the one quoted is in fact the one I had supplied, which is a considerably adapted and, I had hoped, somewhat disguised description of one of my own publications for practitioners which I won’t actually submit. (From that you can at least deduce that I don’t think it is a candidate for a 4* rating.) There should be no suggestion that this publication or its 150 word description is any kind of ideal model output or that it would be assessed at any specific level. It is illustrative only in that it includes the types of information we have asked for. I admit that one of the reasons for using this particular example was that there had been some concern expressed to me by different sources that literature reviews in particular and useful research in general were not valued by the RAE – one source of these concerns, but not the only one, was the DfES. So Ian is right that the selection was not entirely innocent and was intended to reassure the state among others, but wrong if he takes it as evidence of a bias towards neo-positivism, a predilection for systematic review (which incidentally this description deliberately does not claim to embody), or a desire to kowtow to the views of government.

However I do need to challenge his assertion that merely on the grounds that it relates to a national context the assessed research quality is limited to a 1*. The quality of an output will be determined only by the quality of the research, not by whether either the context or publication is international. Thus it would be entirely possible for a literature review of national literature intended for application nationally to gain a 4* if it were judged to be of outstanding quality.

Finally all measurement reflects very limited aspects of what is being assessed and is by its nature imperfect and imprecise. We only do it because for some reason it is seen to be necessary. I am not going to attempt to justify the RAE as a whole, but only to say that given that the elected government has come up with this process then the education subpanel, whatever its collective and personal concerns, will try to grasp the nettle and carry out the task to the best of its ability, hopefully dispensing justice with mercy! I have always yearned to play the role of Portia, but had not previously connected it with the RAE, so for that, as for several other insights, I am especially grateful to Ian!
Notes (Ian’s paper)

1 Thanks for critical comment to Jo Frankham and Harry Torrance. An earlier version of this paper was delivered at the British Educational Research Association conference, Warwick University, September 2006.

2 It may seem strange that a simple numerical scale should become so contaminated with letters and stars, but it’s important also to read these scales as class markers, in the way that socio-economic classes are distinguished in the UK. To joke only a little, 3b/3a might be seen as corresponding to a distinction between ‘rough’ and ‘respectable’ working class. The economic disenfranchisement of 1/2/3b/3a in 2001 thereby enhanced the loading of funding at the top end and widened the gap between ‘have’ and ‘have-not’ in ways that might also be regarded as analogous, particularly as these cuts were conducted under the overall policy banner of ‘capacity building’. There are underlying parallels, it might be argued, with New Labour inclusionary rhetorics.

3 There is a longstanding history of ‘significance’ and ‘originality’ in relation to questions of ‘recognition’ in Science and its Sociology (Merton 1957). The priority of ‘originality’ is clear.

4 Armstrong and Goodyear argue: ‘It would not do for an assessment model to be dominated by advocates of large-scale, randomised control group experiments (or poststructuralist policy critique, for that matter)’ (2005: 21). Indeed. (See also Stronach 2005).

5 A natural riposte to such questions is to dismiss such possibilities as a slur on the integrity of reviewers. But our recent and highly analogous experience with an ESRC end-of-report review shows that such paradigm warfare does happen. The report and case study (Piper et al) was reviewed. One reviewer said good things about the research. First, ‘a qualitative approach is justified’. But ‘with some reluctance I have rated this report “problematic” rather than “good”. Why? The reviewer’s “fairly strong reservations derive from my prejudices against qualitative research”. The other 3 reviewers rated the research ‘outstanding’, which was its overall grade. The example illustrates both the possibilities of bias and its correction. It highlights the need for careful moderation across reviewers and the need to take paradigmatic bias into account.

6 Or perhaps we’re back to ‘rigor’ as in ‘rigor mortis’. MacLure has criticised the fruits of such systematic reviews as illustrated in the Hypothetical Example as ‘tiny dead bodies of knowledge’ (MacLure 2005: ).

7 Benjamin differentiates between the ‘intended object’ and the ‘mode of intention’ in making this point in relation to the meaning of ‘bread’ and ‘wine’ in different languages (Benjamin 1973: 74). Others might express a similar point in relation to connotative and denotative meanings. The RAE’s move from overall institutional assessment to individual item appraisal has its parallels in the practices of risk management: ‘The so called asset-by-asset approach dominates the scene over the portfolio-theoretical approach’ (Kalthoff 2005: 75).

8 In discussion at the BERA presentation of an earlier version of this paper, I argued that the RAE was like the ‘pool’s panel’ – which meets in the UK to decide the result of postponed football matches so that gambling results are available. But that was the 2001 RAE. There is a difference in the 2008 RAE exercise. This time the panel’s wisdom is far greater: no longer content to decide the outcome of a match that was never played, it also decides how well each player performed.

References


Research Assessment Exercise (2006), Sub-panel criteria and working methods, Unit of Assessment 45, Education.

Research Assessment Exercise (2006) Sub-panel criteria and working methods, Unit of Assessment 41, Sociology.

Research Assessment Exercise (2006), Sub-panel criteria and working methods, Unit of Assessment 32, History.

Research Assessment Exercise (2006), Sub-panel criteria and working methods, Unit of Assessment 60, Philosophy.

Research Assessment Exercise (2006), Sub-panel criteria and working methods, Unit of Assessment 62, History.

Research Assessment Exercise (2006), Criteria and working methods, Unit of Assessment 58, Linguistics.


Research Assessment Exercise (2006), Criteria and working methods, Panel D, Agriculture, Veterinary and Food Science.
Research Assessment Exercise (2006), Sub-panel criteria and working methods, Unit of Assessment 66, Communication and Media Studies.

Research Assessment Exercise (2006), Sub-panel criteria and working methods, Unit of Assessment 19, Physics.


