Ranking the outcomes from the assessment of research in the UK

Anthony Kelly

School of Education,
University of Southampton,
Highfield, Southampton SO17 1BJ, UK
E-mail: a.kelly@soton.ac.uk

Abstract: The Research Excellence Framework (REF), previously called the Research Assessment Exercise (RAE), is an assessment undertaken on behalf of the UK government to gauge the quality of research in UK universities. It is a public display of reputation that carries with it a considerable risk for participants in terms of losing face and funding. The previous exercise, the 2008 RAE, focused on three categories of research output – publications, environment and esteem – quantified on a common scale and aggregated into a single spread of research activities for each submitting department. (The upcoming 2013 REF proposes something very similar, replacing 'Esteem' with 'Impact'). The outcomes were then ranked against each other by the national media (and internally by universities themselves) using an arbitrary weighting system. Using data from 'Education' in the 2008 RAE as an example, this paper develops an alternative 'justified' weighting system, adaptable to REF outcomes, to map each department's distribution of activities to a single scalar Grade Point Average (GPA). While acknowledging the shortcomings inherent in any single-criterion weighting system and the disputed value of the REF itself, the probable use of bibliometric citation data in some units of assessment is more likely to make a theoretically justified weighting system even more important to the credibility of emergent rankings.

Biographical notes: Anthony Kelly is a Professor and Director of Research at the School of Education in the University of Southampton. He lectures and researches in the general area of education policy and school improvement/effectiveness, and is particularly interested in the development of theory and methodology in these fields. He is founding editor of the journal *Education*, *Knowledge and Economy*.

[Author: Please reduce abstract of no more than 100 words.]

1 Introduction

In the UK, Research Assessment Exercises (RAEs) are undertaken every few years[1] on behalf of the government to gauge the quality of research undertaken by its universities.

The next one, the REF, is scheduled for 2013 and is very similar to the previous (2008) one in structure and purpose. Submissions from each academic discipline are ranked by peer-review Panels, and these rankings are then used to allocate funding from the National Funding Councils.

The exercises have always been controversial and some vice-chancellors have suggested that they are "outdated, expensive, time-consuming and unfair" (Rothwell, 2008). The 2008 RAE, which focused on outputs rather than on individuals (as was the case in 2001), used the same principles of peer assessment, but the results were presented as a 'graded profile' rather than as a number on a fixed seven-point scale. This was intended to allow the four UK Funding Councils to identify 'pockets of excellence' wherever they might be found (RAE, 2008a). All subjects were assessed against agreed quality standards within a common framework that recognised appropriate variation between subjects and included both quantitative and descriptive elements (RAE, 2005).

Alongside every other subject or 'Unit of Assessment' (UoA), of which there were 67, the 2008 RAE assessed each of 81 university Departments of Education[2] in terms of research publications produced by staff in the period 1 January 2001– 31 December 2007, research students and income for the period 1 January 2001– 31 July 2007, and information on staff in post on the census date 31 October 2007.[3] The 'Education Panel' categorised and graded overall research activity under three component outputs: publications;[4] environment (students, income, structure, strategy and staffing) and staff esteem. Each of the three was graded on a scale of 1* to 4*, where 1* represented research that was 'recognised nationally in terms of originality, significance and rigour'; 2* was research that was 'recognised internationally';[5] 3* represented 'internationally excellent' research, but which fell 'short of the highest standards of excellence'; 4* was 'world-leading' (RAE, 2008b). A fifth category, '*Unclassified*', represented work that fell 'below the standard of nationally recognised work' or did not 'meet the published definition of research'. From this, an overall research quality profile was then calculated as the *aggregate* of the three components, having weighted each in advance, and the results ranked nationally.[6]

The validity of attempting to gauge the value of scholarship in such a crude manner, even when undergoing 'an expert review process' (RAE, 2008c), has been debated elsewhere (e.g., Stronach, 2007 vs. Brown[7], 2007), but evidence suggests that all 81 submitting departments took the 2008 exercise very seriously (Harvey, 2008) – how could they not when the Funding Councils were to use the results to calculate future research grants – and made sometimes fractious 'political' decisions as to which colleagues were submitted and which were excluded.[8] All the leading research-focused universities had 'trial runs' in advance of the census, all had nominated 'champions' to guide submissions, and most had predicated their recruitment and promotion policies on achieving a favourable outcome. The situation is identical today in advance of the REF. All in all, it was and remains a huge investment of time and resources, which critics have suggested might have been better spent actually doing research, but whose mechanisms are guided in several respects by academics themselves (Wooding and Grant, 2003) and by the international market created by universities for their own purposes.

The RAE results for each department were released in mid-December 2008[9] as spreads of percentage research activity in each of the five categories UC to 4*. In their submissions, university departments sought to achieve a skew at the lower-quality end or at least a normal distribution – Figure 1 shows the curves for a range of

departments[10] – but in any case to avoid submitting activity likely to be graded UC and 1*. In this respect, as Table 1 (and Table 6) shows, some relatively successful departments failed properly to gauge the standard, raising the question of whether the 2008 RAE Education Panel saw greater praiseworthiness in 1* activity than other Panels, thereby lowering the relative performance of the whole Education sector, or judged research activity generally more harshly. Figure 1 seems to support this view since the Education UoA overall has a slight skew at the high-quality end (and the two departments with normal distributions are ranked 35 and 36 out of 81),[11] though the Panel itself asserted in its final report that the results were in line 'with other *comparable* disciplines' (RAE, 2009, p.6, my emphasis).

Table 1   Top 22 ranked departments

University	FTE*	4*	3*	2*	1*	UC	'Common' weighting score	Rank
Institute of Education	218.03	35	30	25	10	0	2.90	1
University of Oxford	36	30	35	30	5	0	2.90	2
University of Cambridge	49.6	30	35	25	10	0	2.85	3
King's College London	34.15	30	35	25	10	0	2.85	3
University of Bristol	42.56	25	35	30	10	0	2.75	5
University of Leeds	33.6	20	40	35	5	0	2.75	5
University of Exeter	23.83	20	40	30	10	0	2.70	7
University of Warwick	35.43	20	40	25	15	0	2.65	8
Manchester Metropolitan	22.8	20	35	35	10	0	2.65	8
University of York	13.49	15	45	30	10	0	2.65	8
University of Durham	30.8	20	35	30	15	0	2.60	11
University of Sussex	21.5	20	35	30	15	0	2.60	11
University of Stirling	20.6	15	40	35	10	0	2.60	11
University of Manchester	40.6	20	35	30	10	5	2.55	14
University of East Anglia	14.6	15	35	40	10	0	2.55	14
University of Bath	32.7	15	30	40	15	0	2.45	16
University of Newcastle	11.1	10	40	35	15	0	2.45	16
London Metropolitan	10.8	10	40	35	15	0	2.45	16
Open University	77.01	10	35	40	15	0	2.40	19
University of Nottingham	51.2	20	25	35	15	5	2.40	19
University of Sheffield	24.2	15	25	45	15	0	2.40	19
Lancaster University	17.9	15	30	35	20	0	2.40	19

*FTE = full-time equivalents.
        Source: Guardian (2008)

The Times Higher Education, and most newspapers and universities (internally), to rank submitting departments, calculated GPA as follows: 1* activity was weighted 1; 2* activity was weighted 2; 3* activity was weighted 3 and 4* activity was weighted 4 (Corbyn, 2008). The results were added together and divided by 100 to give a weighted

mean for each department of between 0 and 4, as shown in Table 1 for the top 20 or so departments (Guardian, 2008). This 'common' weighting system is about to be used again for the 2013 REF and its simulations, but how 'fair' is it? Is it really the case that 4* activity is twice as 'good' as 2* activity?

One month after the publication of the 2008 RAE results, the Higher Education Funding Council for England (HEFCE, 2009) published its own weighting system for Quality Related (QR) funding purposes: 7, 3 and 1 for 4*, 3* and 2* activity, respectively, with zero weighting for 1* and Unclassified (UC).[12] Not surprisingly, the 'common' weightings and the 'QR' weightings produce different outcomes. Table 2 shows two Education departments, X and Y, with their respective distribution of activity, and their ranks using each of the two sets of weights.

Table 2   How 'common' and 'QR' weightings can produce different rankings

Activity grade	*4**	*3**	*2**	*1**	*UC*	
'Common' weightings	4	3	2	1	0	GPA
Department X	15%	30%	55%	0%	0%	2.60
Department Y	25%	25%	30%	20%	0%	2.55
'QR' weightings	7	3	1	0	0	GPA
Department X	15%	30%	55%	0%	0%	2.50
Department Y	25%	25%	30%	20%	0%	2.80

Figure 1   Submission curves for a range of departments (see online version for colours)
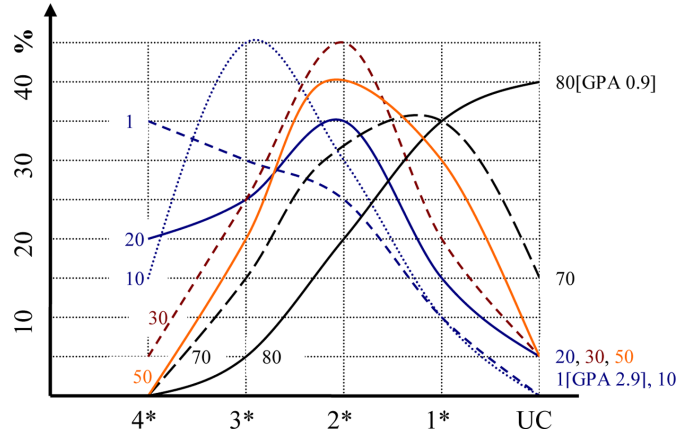
The results are significantly different under the two systems, most notably in the fact that their relative rankings are reversed. The issue is *not* that different weighting systems produce different outcomes, but that the systems (particularly the 'common' one) appear to be random. The following section, therefore, explores, using data from the 2008 RAE, the basis for developing a 'justified' weighting system for the 2013 REF – that is to say, one conceived and operationalised by theory – under several different conditions, for what is admittedly an 'imperfect and imprecise' exercise (Brown, 2007, p.356).

2 Preliminary theoretical discussion

The most obvious way to rank submitting departments is simply on their percentage of 4* activity, and then to 'split ties' on the basis of the percentage of 3*, 2* and 1* activity in turn; if two departments have the same 4*, 3*, 2* and 1* activity, then they are ranked based on the total number of FTE entered. Such a system acknowledges the RAE's aspiration to identify excellence, but it only rewards *a certain type of* excellence – 'world-leading' excellence – and does not capture the relative merit of having a high percentage of internationally (and nationally) excellent outputs. In effect, such a system assigns a weight of zero at every iteration to the grade below the one being used, which seems counterintuitive when grading scholarship.

Assuming a non-iterative holistic ranking system, the critical issue then becomes whether 1* activity should be given zero weighting.[13] On the one hand, it can be argued that since it is known that 1* activity will receive no QR funding (HEFCE, 2009), and since submitting departments tried (in vain, as it turned out) to avoid having any such activity, it would seem perverse to reward its inclusion. On the other hand, there should be some reward for departments that submitted more staff,[14] even when those colleagues were not (yet) producing activity of international significance, and since 1* activity is most likely (in theory at least!) to come from staff on the borderline of submission, there should be some (small but non-zero) recognition of their contribution.

Table 3 shows the case for Nottingham and captures the effect of both perspectives on GPA.

Table 3 An example (Nottingham) of the effect of zero-weighting 1*

Activity grade	*4**	*3**	*2**	*1**	*UC*	
Percentage in each category (FTE = 51.2)	20%	25%	35%	15%	5%	GPA
A weighting system (4,3,2,1,0) with zero weight for UC	4	3	2	1	0	2.40
A weighting system (4,3,2,0,0) with zero weight for 1* and UC	4	3	2	0	0	2.25

Since nearly half (45%) of the submissions made by Nottingham were 'internationally excellent' or 'world-leading', and since it entered a relatively large number of staff in 2008 (51.2 FTE),[15] it could be argued that it is unfair to penalise Nottingham 'doubly' by assigning a zero weight to its 15% of 1* activity since its overall GPA has already 'suffered' from including it.

Clearly, there is merit in both arguments – on the one hand that the stated aim of the RAE is to identify excellence and support the allocation of QR funding, and on the other that it unduly penalises 'nationally recognised' scholarship – so this paper now considers a more robust ranking system for both cases: where *three* grades (2* to 4*) warrant weightings and separately for the case where there are *four* non-trivial grades (1*4*).

3 Developing theory for a justified weighting system for the REF

This model assumes *initially* that only 2*, 3* and 4* activity has value and that an overall scalar can be attached to each submitting department so that all 81 Education departments

can be ranked relative to each other. Since 1* and UC activity is zero-weighted in this scenario, it is simply ignored, except in the case of a tie when the total number of FTE submitted at grades 1*–4* is taken into account.

Consider a ranked distribution of three non-zero weights of the form:

$$w(4^*, 3^*, 2^*) \rightarrow]0\%, 100\%]$$

that is to say, where each weight is a non-zero scalar between 0 and 100%, and where all three weights sum to 100%:

$$w(4^*) + w(3^*) + w(2^*) = 100\%$$

and where

$$w(4^*) > w(3^*) > w(2^*).$$

Consider the set of *all* distributions of three non-zero weights as a triangle, as shown in Figure 2. The three boundary lines are not included since $w(4^*) \neq w(3^*) \neq w(2^*)$. The vertices are (100, 0, 0), (0, 100, 0) and (0, 0, 100) and the area indicated by the arrow represents the subset of all distributions where $w(4^*) > w(3^*) > w(2^*)$.

Figure 2   The set of *all* distributions of three non-zero weights (see online version for colours)
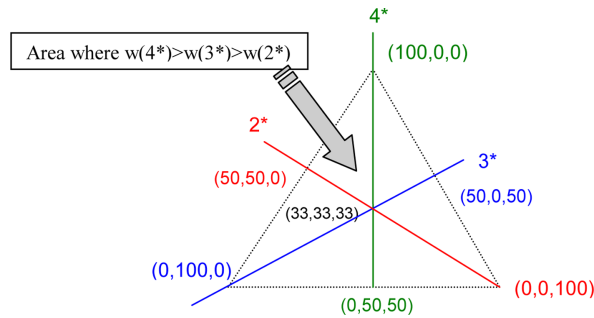
If the same set of weights is to be applied across all activity within a department and across all submitting departments, it is best estimated as the centroid, representing the mean, of the 'bounded area' arrowed in Figure 2 whose three vertices are (100, 0, 0), (50, 50, 0) and (100/3, 100/3, 100/3).

In the case where four activity grades (1*, 2*, 3* and 4*) have value, by analogy with the three-grade case above, and where there is a ranked distribution of four non-zero weights of the form

$$w\{4^*, 3^*, 2^*, 1^*\} \rightarrow]0\%, 100\%]$$

$$w(4^*) + w(3^*) + w(2^*) + w(1^*) = 100\%$$

$$w(4^*) > w(3^*) > w(2^*) > w(1^*)$$

the set of distributions is represented by a tetrahedron whose vertices are (100, 0, 0, 0), (50, 50, 0, 0), (100/3, 100/3, 100/3, 0) and (25, 25, 25, 25).

And in the (more unlikely) case where all five RAE grades (UC, 1*, 2*, 3* and 4*) have value, and where

$$w\{4^*, 3^*, 2^*, 1^*, UC\} \rightarrow]0\%, 100\%]$$

$$w(4^*) + w(3^*) + w(2^*) + w(1^*) + w(\text{UC}) = 100\%$$

$$w(4^*) > w(3^*) > w(2^*) > w(1^*) > w(\text{UC})$$

the set of distributions is represented by a 5-shape whose vertices are (100, 0, 0, 0, 0), (50, 50, 0, 0, 0), (100/3, 100/3, 100/3, 0, 0), (25, 25, 25, 25, 0) and (20, 20, 20, 20, 20).

Reverting to simpler situations for completeness: in the case where only two grades (3* and 4*) have value, the set of distributions is represented by a half-line whose end points are (100, 0) and (50, 50); in the case where only one grade (4*) has value, the weight is trivially represented by the point (100, 0).

All these solutions are represented in Table 4.

Table 4　　　All solutions

Vertex where 1 weight is needed	**100**	0	0	0	0
Vertices where 2 weights are needed	**50**	**50**	*0*	*0*	*0*
Vertices where 3 weights are needed	**100/3**	**100/3**	**100/3**	*0*	*0*
Vertices where 4 weights are needed	**25**	**25**	**25**	**25**	*0*
Vertices where 5 weights are needed	**20**	**20**	**20**	**20**	**20**

Since the centroid for any given 'bounded area' can represent the mean of the vertices, when *three* weights are needed (for the three grades 2*–4*), they can be got by adding each column in turn, down to row 3, and averaging:

$$1/3[100 + 50 + 100/3] = 61.11\%$$

$$1/3[50 + 100/3] = 27.78\%$$

$$1/3[100/3] = 11.11\%.$$

We refer to this as the 'justified 3-weight' system.

When *four* weights are needed (for grades 1*–4*), the corresponding scalars are:

$$1/4[100 + 50 + 100/3 + 25] = 52.08\%$$

$$1/4[50 + 100/3 + 25] = 27.08\%$$

$$1/4[100/3 + 25] = 14.58\%$$

$$1/4[25] = 6.25\%.$$

We refer to this as the 'justified 4-weight' system.[16]

4　Analysis of 2008 RAE outcomes using 'justified 3-weights' and 'justified 4-weights'

Table 5 shows the justified weightings, rounded to the nearest integer, for the 3-weight and 4-weight scenarios: namely, where the UC grade *only* is ignored (as in the 'common' system currently used by newspapers and internally by universities) and where *both* UC and 1* grades are ignored (as in the HEFCE 'QR' system).

Table 5   Summary of two justified weighting systems

Activity grade	4*	3*	2*	1*	UC
Weights when only the UC grade is ignored: the 'justified 4-weight' system	52	27	15	6	n/a
Weights when both UC and 1* grades are ignored: the 'justified 3-weight' system	61	28	11	n/a	n/a

Table 6 shows the entire UoA cohort of 81 submitting Education departments ranked using the 'common' system, and ranked again using the 'justified 4-weight' and 'justified 3-weight' systems. The shaded cells show those departments whose ranks have changed by two or more places as a result of the new weightings. Clearly, there is little variation at the upper end of the rankings – no change at all, in fact, for the top nine institutions – so it can be said that the 2008 RAE was fairly secure in its judgement as to which departments are performing best (accepting that there are limitations to the assessment in the first place).

Table 6   The effect of weighting systems on the entire cohort

University (ranked by order as per Table 1) All ties split on #FTE	Rank using common weights	4*	3*	2*	1*	UC	Score using 'justified 4-weights'	Rank	Score using 'justified 3-weights'	Rank
Institute of Education	1	35	30	25	10	0	30.65	1	32.50	1
University Oxford	2	30	35	30	5	0	29.85	2	31.40	2
University Cambridge	3	30	35	25	10	0	29.40	3	30.85	3
King's College, London	4	30	35	25	10	0	29.40	4	30.85	4
University Bristol	5	25	35	30	10	0	27.55	5	28.35	5
University Leeds	6	20	40	35	5	0	26.75	6	27.25	6
University Exeter	7	20	40	30	10	0	26.30	7	26.70	7
University Warwick	8	20	40	25	15	0	25.85	8	26.15	8
Manchester Met	9	20	35	35	10	0	25.70	9	25.85	9
University York	10	15	45	30	10	0	25.05	12	25.05	13
University Durham	11	20	35	30	15	0	25.25	10	25.30	10
University Sussex	12	20	35	30	15	0	25.25	11	25.30	11
University Stirling	13	15	40	35	10	0	24.45	14	24.20	14
University Manchester	14	20	35	30	10	5	24.95	13	25.30	12
University East Anglia	15	15	35	40	10	0	23.85	15	23.35	15
University Bath	16	15	30	40	15	0	22.80	17	21.95	17
University Newcastle	17	10	40	35	15	0	22.15	20	21.15	19
London Met	18	10	40	35	15	0	22.15	21	21.15	20
Open University	19	10	35	40	15	0	21.55	23	20.30	24
University Nottingham	20	20	25	35	15	5	23.30	16	23.05	16
University Sheffield	21	15	25	45	15	0	22.20	19	21.10	21
Lancaster University	22	15	30	35	20	0	22.35	18	21.40	18

Table 6   The effect of weighting systems on the entire cohort (continued)

| University (ranked by order as per Table 1) All ties split on # FTE	Rank using common weights	4*	3*	2*	1*	UC	Score using 'justified 4-weights'	Rank	Score using 'justified 3-weights'	Rank
University Birmingham	23	10	35	35	20	0	21.10	25	19.75	25
Queen's University Belfast	24	10	40	30	15	5	21.40	24	20.60	23
University Edinburgh	25	15	30	30	20	5	21.60	22	20.85	22
University Southampton	26	10	25	45	20	0	19.90	26	18.05	26
Roehampton University	27	5	25	50	20	0	18.05	31	15.55	33
Canterbury Christ Church	28	10	25	35	25	5	18.70	27	16.95	27
University Ulster	29	10	15	50	25	0	18.25	28	15.80	32
University Leicester	30	5	25	45	20	5	17.30	35	15.00	36
University Plymouth	31	10	20	40	25	5	18.10	30	16.10	30
University Brighton	32	10	25	30	30	5	18.25	29	16.40	28
University Winchester	33	5	35	30	20	10	17.75	32	16.15	29
University Glasgow	34	10	20	35	30	5	17.65	33	15.55	34
University Strathclyde	35	5	25	40	25	5	16.85	38	14.45	39
Oxford Brookes University	36	5	25	40	25	5	16.85	39	14.45	40
University Reading	37	10	20	35	30	5	17.65	34	15.55	35
Glasgow Caledonian Uni.	38	0	25	50	25	0	15.75	45	12.50	47
University Lincoln	39	5	30	40	10	15	17.30	36	15.85	31
University Hull	40	5	20	45	25	5	16.25	42	13.60	43
Goldsmiths College	41	10	20	30	35	5	17.20	37	15.00	37
Staffordshire University	42	5	30	25	35	5	16.55	40	14.20	41
University Aberdeen	43	5	20	40	30	5	15.80	43	13.05	44
Birmingham City	44	5	20	40	30	5	15.80	44	13.05	45
Sheffield Hallam	45	5	20	35	35	5	15.35	47	12.50	48
University West England	46	5	10	50	35	0	14.90	48	11.35	50
University Gloucestershire	47	10	15	40	20	15	16.45	41	14.70	38
University Huddersfield	48	5	10	50	30	5	14.60	49	11.35	51
University Wolverhampton	49	5	15	35	40	5	14.30	51	11.10	52
Liverpool John Moores	50	0	20	40	35	5	13.50	56	10.00	60
University Sunderland	51	0	25	30	40	5	13.65	55	10.30	56
Bangor University	52	10	20	20	35	15	15.70	46	13.90	42
Newman UC	53	0	25	35	30	10	13.80	54	10.85	54
University Dundee	54	5	15	35	35	10	14.00	53	11.10	53
Coventry University	55	0	20	40	30	10	13.20	59	10.00	61
University East London	56	5	20	30	30	15	14.30	52	11.95	49
University West Scotland	57	5	25	25	25	20	14.60	50	12.80	46
University Northampton	58	5	10	40	35	10	13.40	57	10.25	59
Leeds Met	59	0	20	35	35	10	12.75	63	9.45	64

Table 6   The effect of weighting systems on the entire cohort (continued)

| University (ranked by order as per Table 1) All ties split on # FTE	Rank using common weights	4*	3*	2*	1*	UC	Score using 'justified 4-weights'	Rank	Score using 'justified 3-weights'	Rank
University Glamorgan	60	0	20	35	35	10	12.75	64	9.45	65
University Hertfordshire	61	0	20	35	35	10	12.75	65	9.45	66
University Greenwich	62	5	5	45	40	5	13.10	60	9.40	67
Thames Valley University	63	0	25	30	30	15	13.05	61	10.30	57
Brunel University	64	5	10	35	40	10	12.95	62	9.70	63
St George's Medical School + Kingston (Joint)	65	0	25	25	35	15	12.60	67	9.75	62
Loughborough University	66	5	15	30	35	15	13.25	58	10.55	55
Swansea Met	67	5	10	30	40	15	12.20	68	9.15	68
Stranmillis UC	68	5	20	15	40	20	12.65	66	10.30	58
Liverpool Hope University	69	0	20	30	25	25	11.40	69	8.90	69
University Bolton	70	0	15	30	40	15	10.95	70	7.50	70
Bath Spa University	71	0	15	30	35	20	10.65	71	7.50	71
Edge Hill University	72	0	5	25	50	20	8.10	72	4.15	74
University Central Lancs	73	0	5	25	50	20	8.10	73	4.15	75
University Wales Institute	74	0	0	30	50	20	7.50	75	3.30	78
St Mary's UC	75	0	10	20	40	30	8.10	74	5.00	72
University Wales, Newport	76	0	0	25	55	20	7.05	78	2.75	80
University Cumbria	77	0	5	25	35	35	7.20	76	4.15	76
UC Plymouth St. Mark & St. John	78	0	5	15	50	30	6.60	79	3.05	79
Bishop Grosseteste UC	79	0	10	15	35	40	7.05	77	4.45	73
Glyndŵr University	80	0	5	20	35	40	6.45	80	3.60	77
York St. John University	81	0	0	15	55	30	5.55	81	1.65	81

Figure 3(a) and (b) shows the *extent* and *spread* of this statistical 'churn' over the whole cohort. It can be seen that the variation (from the 'common' rankings) is more exaggerated in the case of the 'justified 3-weight' system, which assigned zero weighting to both UC and 1* outputs.

Comparison between Figure 3(a) and (b) also reveals that the greatest difference between the 'common' weighting system and the two justified weighting systems occurs *not* in the top or bottom quartiles of performance, but in the middle, with the greatest variation occurring in the second-from bottom quartile (i.e., among those departments ranked between 40 and 60 on the 'common' scale). However, the figures also show that there is no significant difference between the 'justified 3-weight' system and the 'justified 4-weight' system in terms of how each differs from the 'common' weighting system, so that criticism of HEFCE for ignoring 1* activity is *not* supported, at least not in relation to how the RAE/REF defines its mission.

Figure 3   (a) Difference between the 'common' and 'justified 4-weight' systems and
(b) difference between the 'common' and 'justified 3-weight' systems (see online
version for colours)
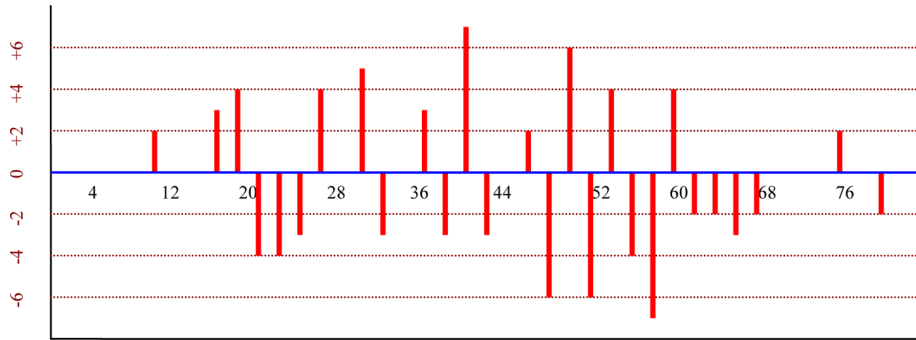
(a)
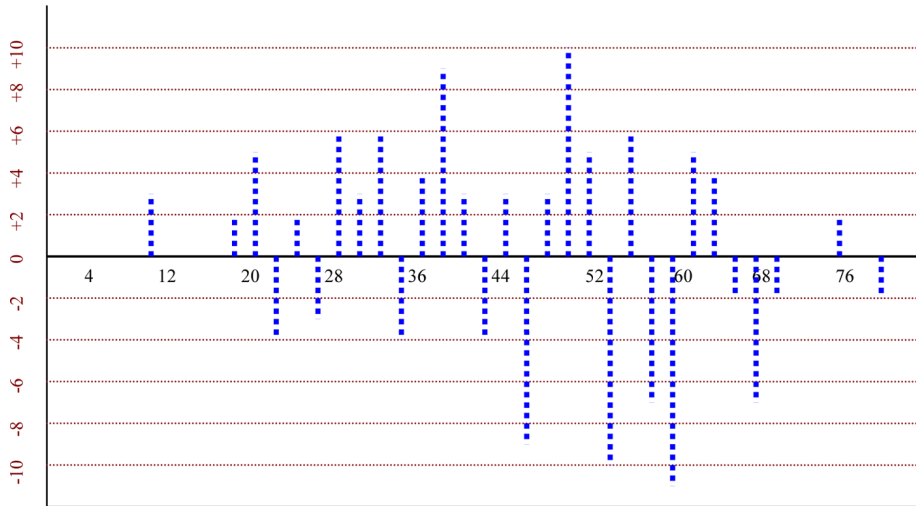
(b)

On the positive side, it is worth noting that the 'QR' weighting system (7, 3, 1, 0, 0) is much closer to the 'justified 3-weight' system (61, 28, 11, 0, 0) than any other weighting system. On the negative side, Figure 3(a) and (b) suggests that since there is approximately the same number and extent of winners as losers in any change of weighting system, RAE/REF ranking is essentially a zero-sum game.[17] This adds support to those who feel instinctively that the (sometimes aggressive) competition that develops in academia in the period running up to research assessments (Curran, 2000, 2001; Sikes, 2006), not just in the UK but elsewhere (Orr, 2004; Harvey, 2008; Ranking Forum of Swiss Universities, 2008), runs counter to the intrinsically cooperative nature of scholarly endeavour. It is not clear that policy-makers have taken this into account in the run-up to the REF: there is a difference between competition, which they see as a 'good thing' *per se*, and zero-sum assessment, which has an adverse impact on cooperation between institutions and 'linking pockets of excellence'.

5   Conclusions

The 'Education' UoA descriptor for the 2008 RAE described research in the field as 'multidisciplinary and closely related to a range of other disciplines with which it shares blurred boundaries' (RAE, 2006, p.29). It is perhaps fitting then that Education accepts some blurring at the boundaries of performance assessment – indeed the chair of the Education Panel acknowledged, in advance of the 2008 RAE, its 'inherent imprecision' and 'the difficulty of devising and applying' 'fairly vague criteria' (Brown, 2007, p.353) – but with reputation, funding and the ability to retain and recruit research-active staff more likely to (re)emerge as functions of the 2013 REF, it seems careless to accept the indiscriminate use of random weighting systems when 'justified' ones are available.

The UK government announced in 2006 that *after* the 2008 RAE, a new system of metrics (the REF) would be developed to inform future QR funding mechanisms and which would include, to varying extents depending on the academic discipline, new bibliometric indicators alongside existing peer-review processes (HEFCE, 2008). Though peer review will remain the more significant of the two for subjects like Education, the REF generally will be a mix of quantitative and qualitative data across all UoAs, so the demand for a justified weighting system can only increase in significance as far as ranking outcomes is concerned. What has been proposed in this paper is a utility function that maps the distribution of percentage grades to a single scalar, and although this is a more justified weighting system and more in line with the expectation of the 'Berlin Principles' (CHE, 2006) than both the 'common' one used internally in universities and the 'QR' one used by HEFCE, *any* single-criterion system will inevitably attract criticism. There may be more complex Bayesian models available as alternatives, but since any assessment exercise depends for its credibility on transparency and interpretability, the cost of adding greater complexity is sometimes not so much a gain in robustness as a loss in understanding among practitioners.

Acknowledgements

References

Brown, M. (2007) On promoting rigour – a response', *Journal of Education Policy*, Vol. 22, No. 3, pp.353–356.

Centrum für Hochschulentwicklung (CHE) (2006) *Berlin Principles on Ranking of Higher Education Institutions*, Available online at: http://www.che.de/downloads/Berlin_Principles_ IREG_534.pdf, Accessed 21 February 2009.

Corbyn, Z. (2008) *RAE 2008: The Results*, Times Higher Education, 18 December, Available online at: http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=4047 86, Accessed 16 February 2009.

Curran, P.J. (2000) 'Competition in UK Higher Education: competitive advantage and Porter's diamond model', *Higher Education Quarterly*, Vol. 54, pp.386–410.

Curran, P.J. (2001) 'Competition in UK Higher Education: applying Porter's diamond model to Geography departments', *Studies in Higher Education*, Vol. 26, No. 2, pp.223–251.

Guardian (2008) *RAE 2008: Education Results*, The Guardian, 18 December, Available online at: http://www.guardian.co.uk/education/table/2008/dec/18/rae-2008-education/print, Accessed 16 February 2009.

Harvey, L. (2008) 'Editorial – rankings of higher education institutions: a critical review', *Quality in Higher Education*, Vol. 14, No. 3, pp.187–207.

HEFCE (2008) *Research Excellence Framework: Outcomes of Consultation and Next Steps*, Circular Letter Number 13/2008, Available online at: http://www.hefce.ac.uk/pubs/circlets /2008/cl13_08/, Accessed 6 June 2008.

HEFCE (2009) *Funding for Universities and Colleges in 2009–10*, Published Electronically on 28 January and available at: https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind0901&L= ADMIN-HEFCE&T=0&F=&S=&P=5163, Accessed 16 February 2009.

Lovegrove, R. (2008) 'Olympic games' medals tables', *Mathematics Today*, June, pp.121–122. AUTHOR PLEASE CITE THIS REFERENCE IN TEXT.

Orr, D. (2004) 'Research assessment as an instrument for steering higher education – a comparative study', *Journal of Higher Education Policy and Management*, Vol. 26, No. 3, pp.345–362.

RAE (2005) *RAE 2008 Guidance on Submissions, RAE 03/2005*, Available online at: http://www. rae.ac.uk/Pubs/2005/03/, Accessed 16 February 2009.

RAE (2006) *RAE 2008: Panel Criteria and Working Methods, RAE 01/2006*, Available online at: http://www.rae.ac.uk/pubs/2006/01/byuoa.asp?u=k, Accessed 17 February 2009.

RAE (2008a) *About Us*, Available at: http://www.rae.ac.uk/aboutus/changes.asp, Accessed 14 February 2009.

RAE (2008b) Details available at: http://www.rae.ac.uk/, Accessed 16 February 2009.

RAE (2008c) *RAE 2008 Confirms UK's Dominant Position in International Research*, Available online at: http://www.rae.ac.uk/news/2008/results.asp, Accessed 12 January 2009.

RAE (2009) *UOA 45 Subject Overview Report*, p.6, Available online at: http://www.rae.ac. uk/pubs/2009/ov/, Accessed 12 February 2009.

Ranking Forum of Swiss Universities (2008) *On University Rankings*, Cited in Harvey (2008) and available online at: http://www.universityrankings.ch/en/on_rankings, Accessed 21 February 2009.

Rothwell, N. (2008) 'RAE reaction: go figure', *Times Higher Education*, 22 December. AUTHOR PLEASE SUPPLY FULL DETAILS.

Sikes, P. (2006) 'Working in a 'new' university: in the shadow of the research assessment exercise?', *Studies in Higher Education*, Vol. 31, No. 5, pp.555–568.

Stronach, I. (2007) 'On promoting rigour in educational research: the example of the RAE', *Journal of Education Policy*, Vol. 22, No. 3, pp.343–352.

Wooding, S. and Grant, J. (2003) *Assessing Research: The Researchers' View*, RAND, Santa Monica, California.

Notes

[1] 1986, 1989, 1992, 1996, 2001 and 2008. The outcome of the last RAE was published in December 2008 and January 2009 and is the source of the data used in this paper.

[2] Representing 82 institutions: there was one joint submission.

[3] More precisely, the data returned by each submitting department was on: staff (RA0 & RA1), research outputs (RA2), research students and studentships (RA3a–b), research income (RA4), and environment and esteem (RA5a–c). Staff were returned in one of four categories, A–D (see RAE, 2005, para 68).

[4] Basically, four research publications were expected from each full-time eligible member of staff.

[5]'International' and 'National' refer to quality standards, not to the nature or geographical scope of the subject matter or where it has been published.

[6]The ranking was not done by the RAE Panel itself, but by the media on the basis of what the RAE had produced. In establishing an overall quality profile for each submission, the RAE *Education* Panel allocated 70% weighting to publications, 20% weighting to environment and 10% weighting to esteem (RAE, 2006, p.36). Figures were rounded to the nearest 5% according to a strict rounding methodology described in paragraphs 8–11, Annex A, of RAE (2005).

[7]Margaret Brown was Chair of the Education Panel for the 2008 RAE.

[8]As with the upcoming REF, no minimum number of staff was required to be returned in 2008 and departments were *not* required to select for inclusion *all* eligible staff. The RAE Panel reported post facto that many universities were more selective in their 2008 entry than they were in 2001 (RAE, 2009, p.1).

[9]The three 'component' profiles were published on 5 January 2009.

[10]The numbers in Figure 1 refer to rankings using the 'common' weighting system (4,3,2,1,0) as used in newspapers (see Figure 1). The distribution curves for the top-ranked and bottom-ranked departments are almost mirror images of each other.

[11]It is a moot point whether one should assume *ab initio* that RAE grades are normally distributed around 2* or should be 'normalised' upon return, and whether quality standards nationally are the same for all disciplines. Interestingly, the chair of the Education Panel admitted that it would have been 'much easier and probably more honest to norm-reference, working to a fixed percentage in each category', but that would not have been 'politically acceptable' as the percentages would either 'obscure differences of standards in research quality' across disciplines or 'differences would have to be predetermined based on insufficient evidence' (Brown, 2007, p.354).

[12]QR funding is well in excess of £1.1 bn per year and is allocated to universities on this weighted basis (notified to universities on 5 March 2009).

[13]It is assumed that *Unclassified* activity should receive a zero weighting no matter what the system. This is not a necessary theoretical condition, but it seems reasonable to assume that no 'significance' or 'reward', no matter how small, should be attached to research activity that fails to 'meet the published definition of research' (RAE, 2008b).

[14]The RAE Panel reported that 'a strong positive correlation [exists] between size and quality', but that the relationship is 'complex and non-linear' (RAE, 2009, p.1).

[15]The average number of FTE was less than 21 for Education, with a median of 13, and almost 70% of submissions had fewer than 20 FTE.

[16]And in the unlikely event that *five* weights are needed (for all five RAE grades, including UC) the corresponding 'justified 5-weight' system is: 45.67%; 25.67%; 15.67%; 9%; 4%.

[17]Reflected in part by the condition that the weights must sum to a fixed 100.