The Use of Shadow Metrics in the 2008 Business and Management Studies Research Assessment Exercise (RAE) and beyond: a consideration of the implications of applying a proposed metric to the RAE2001 BMS data.

Aidan Kelly, Huw Morris and Charles Harvey

The aim of this note is to provide some insights into the possible consequences of the use of shadow metrics in the 2008 RAE and the use of metrics more generally after the next 2008 RAE. It is important to anticipate and consider these consequences in order to inform the choice of appropriate metrics and the implications for research funding. Before presenting the data analysis it is necessary to discuss some of the issues involved in designing a funding metric.

1. Research metrics and recent DfES proposals

Funding mechanisms are to be viewed as not only rewarding performance but also shaping future performance. The design of metrics involves both a choice of variables to be included and decisions as to the relative weight given to them in the summary metric. A metric should have an appropriate balance between indicators that reward achievement in the production of quality research and, at the same time, shift research activity in desirable directions. Weighting decisions may allocate differential or equal importance to the various indicators that make up the metric. For instance, weighting might be used to give emphasis to the high quality published outputs, to research council funded research or towards collaborative research funded by business partners. Such weightings are key to the shaping of research activity in each unit of assessment. In order for the metric to present clear implications to those whose activity is to be assessed, complex technical procedures such as regression weighting or factor analysis are best avoided.

A funding metric for research is likely to include measures of the scale or volume of activity and the quality of published outputs as measured by a journal-ranking scheme or a citation impact measure. In Science, Technology, Engineering and Medicine, so-called STEM subject areas it has been suggested that researching income alone provides a metric for assessment of research volume and quality (THES, 2006). In Appendix 2 of the recent DfES consultation document it has been suggested that the non-STEM subject areas, the Arts and Humanities, develop shadow metrics using information from a wider range of data sources:

- **Input metrics**
  Research Council income
  Other research income
  User-led income
  Research Council success rate

- **Volume metrics**
  PhD numbers/completions
  Staff with measurable outputs

- **Quality/output metrics**
  Bibliometrics
  User impact
  Research Council evaluation
  Peer esteem

Although the consultation document goes on to say that "this model is confined to the metrics in data sets that are already collected and data sets that exist and may be routinely collected". And also indicates that within the model "there is no separate peer review of outputs, though there is panel determination of output metrics". It concludes by suggesting that, "the model is intended to exemplify one approach that could potentially be adopted for subjects where a research income-based model is not fully satisfactory. The paper then concludes with a recommendation that "this sort of approach could be developed further."

2. Towards a simple model for shadow BMS RAE2008 metrics and beyond

A review of the available datasets covering research activity in BMS reveals that there are no commonly agreed metric based data on user impact, research council evaluation or peer esteem. However, there is a growing body of opinion that established metrics for research income, PhD completions, staff outputs and bibliometrics might usefully be used to inform decision-making in the RAE and in any subsequent metric based assessment of research activity in BMS. Following this line of thought, this note uses a simple two-dimensional metric combining equally volume and quality indicators to reinterpret data from the RAE2001 and produce a sample set of ratings.[1]

2.1 Volume

The volume of research activity can be measured by input and throughput variables taken from the 2001 RAE data. Input volume, was constructed to give equal weight to income from OST and Non-OST sources. Whilst total research income is an obvious unitary measure of volume, the decision to include separate indicators for OST and other income is made on the assumption that government might wish to support those institutions developing collaborative relationships with other, non-OST, funding sources such as industry, government, and charities.

[1] Evidence for the validity of a combined volume and quality approach is provided by the authors in an article to be submitted to the BJM.

Throughput volume indicators are the number of academic staff with measurable outputs and the number of doctorates completed. The input and throughput volume measures form standardised components of the summary volume metric.[2]

2.2 Quality

The quality indicators used were the BBS2004 journal-ranking scheme and the citation impact factor (2003). These indicators are highly correlated (r=0.83) but measure different dimensions of the quality of published outputs (the standards applied by referees and the use made of research by other authors respectively). The measures form the standardised components of a measure of quality.

When deciding to combine the quality and volume measures several weighting possibilities suggest themselves:

a) equal weighting is given to both
b) volume should be weighted by quality
c) weights are allocated to reflect policy preferences.

Giving weight to volume is conservative in its implications since it will reward those larger in the scale of activity, this scale itself part of their inheritance from previous funding decisions. The more weight that is given to quality, the more radical any reallocation of funding is likely to be as it may reward institutions with smaller scale higher quality research. For the purposes of this exercise it was decided to give equal weight to quality and volume indicators in order to see what impact that would have had on the 2001 RAE ratings.

3. Data Analysis

In this section, a metric is proposed based on an equal weighting of volume and quality indicators. This is then applied retrospectively to the 2001 RAE data, and the ability of the metric to determine RAE 2001 ratings is assessed, alongside a consideration of the consequences of the metric the institutions affected.

First of all an item analysis was conducted to reveal how, empirically, the summary metric was correlated with items included and excluded in it. The proposed metric is a multidimensional scale weighted equally for volume and quality indicators. Decisions have been made to treat quality and volume equally and within these component indices to weight each component equally to form an additive index. It is taken as read that each measure of volumes measures a different aspect of volume and these are not necessarily correlated, hence the application of measures of consistency or reliability is not necessary. The purpose of the item analysis is to reveal the characteristic profiles of the universities that would score high or low on the proposed metric.

[2] As the metric combines indicators with different units of measurement, some form of standardisation is necessary. In the analysis that follows, all indicators are transformed into percentile ranks (100 is the higher ranked score on an indicator).

Table 1 shows the correlation between the component items of the metric and the metric itself.[3] These correlations show that the measures of volume and quality have been given equal weight in the summary metric score and in practice would not favour those institutions with high volume relative to quality and vice-versa. The variables that make up the volume and quality measures are also correlated with the summary metric. Institutions with higher OST research income would not find this as clearly reflected in their metric quality rating as the other variables listed. This might be surprising in view of the plausible likely association between OST income and staff appointments, publication records etc. The decision to give non-OST income equal weighting to OST income accounts for it higher correlation with the summary metric. Of the volume indicators, institutions with a high number of doctorates would perform relatively better on the summary metric. These differences are relatively minor, but the do show empirically the profile of performance that would be actually be rewarded if the proposed metric had been used to determine ratings in the 2001 exercise. On balance, although volume and quality are treated equally in the mathematical construction of the metric, in practice it favours institutions with high quality publications. We shall explore the consequences for institutions in a later section, but here it is worth noting the difference between weighting decisions made in the abstract design process and how these may work out when applied to data.

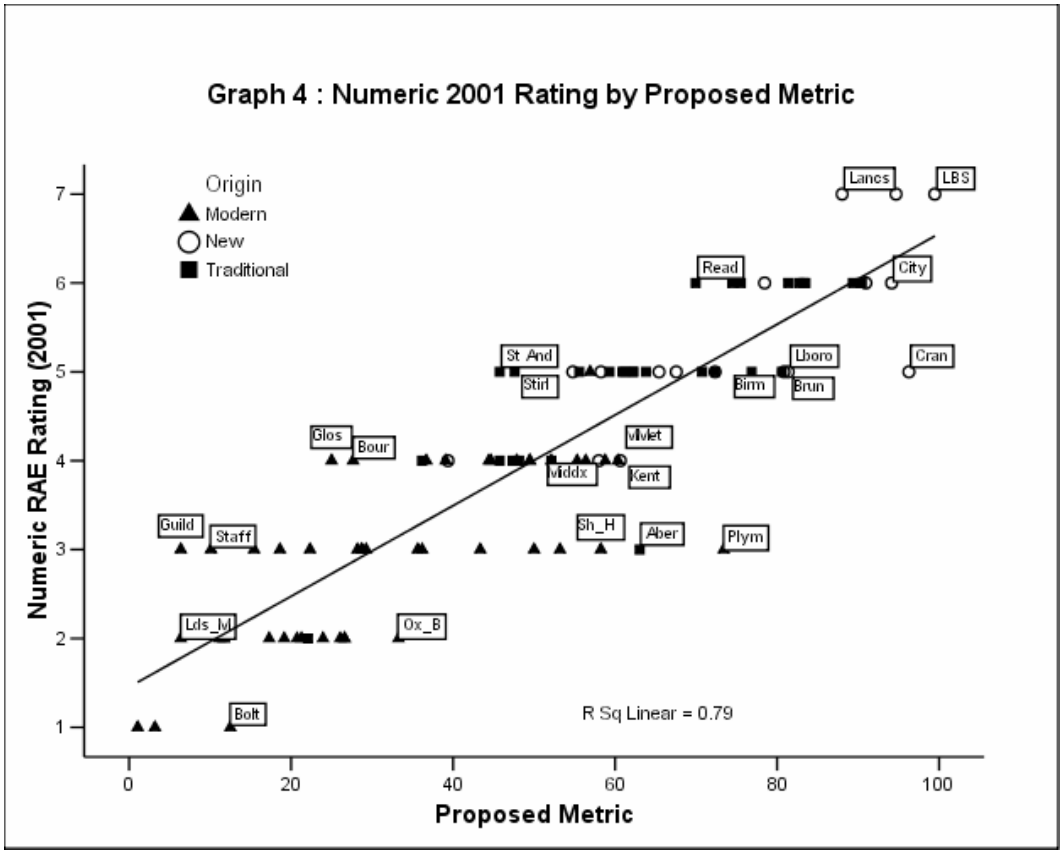
Table 1 Item Analysis of variables Included in Proposed Metric	Pearson's 'r'
Quality	.876
Volume	.881
OST Research Income	.589
Non-OST Income	.717
No. of Staff	.723
No. of Doctorates	.821
BBS04 Ranking	.851
Citation Impact	.851
Grade 4 articles per capita	.813

The correlations between other variables not included in the scale and the proposed metric are presented in Table 2. The proposed metric does not seem to favour institutions with high proportions of income from the various specific sources in the 2001 submission data. It is moderately correlated with employed of research staff as would be expected. However, it is clear that had the proposed metric been the only measure of research performance in the 2001 exercise, it would have benefited those institutions whose published output in the form of journal articles. Indeed the highest correlation is between the metric and grade 4-rated journal articles per academic in the submission. The

Table 2 Correlations for variables not included in Proposed Metric	Pearson's 'r'
Total Research Income	.538
OST Income %	.198
Charity %	-.038
Central Government %	-.114
Industry %	-.019
Overseas %	-.112
Other Income %	.153
No of Research Staff	.627
No. of Research Assistants	.540
Authors %	.029
Editor %	-.080
Chapters %	-.163
Articles %	.180
Other forms %	.236
Income per capita	.450
OST Income per capita	.299
Research Staff per capita	.048
Doctorates per capita	.381
Grade 4 articles per capita	.813

[3] These correlations are the empirical correlates in the 2001 data, the four volume indicators are weighted equally as are the two quality indicators, then the volume and quality indices are weighted equally in the summary metric.

conclusion from table two is that there is no strong case for the inclusion of additional variables in the summary metric.



The graph shows the relationship between the proposed metric scale in its percentile rank form and the RAE 2001 ratings categories expressed as a seven-point scale. The OLS regression suggests that the additive volume and quality metric explains 79% of the variation in the 2001 research rating: a high level of predictive power for the shadow metric.[4] The quality weighted volume measure accounted for slightly less (72%) variation and was not preferred to the simpler additive measure.

What would have been the implications of using this metric in the 2001 exercise? To assess the consequences the metric was used to allocate ratings, on the assumption that the 2001 panel was correct in the proportion of universities it allocated to each rating. Thus it awarded three '5*' and the metric was used to determine which three Universities has the highest rating. According to the 2001 metric the '5*' Universities were:

[4] Using the SPSS PLUM procedure, comparing actual and metric predicted ratings for 2001, 63% of cells had zero frequencies and casts doubt on any results obtained. For information, the Cox-Snell pseudo R square was 0.86.

London Business School
Warwick
Cranfield

Applying the proposed metric to the 2001 data produced the following ‘5 rated’ Universities:

City
Bath
Oxford
Cardiff
Imperial College
Lancaster
Nottingham
Manchester
Cambridge
Leeds
Loughborough
Brunel
Birmingham

The results for all universities are list in the appendix.

Where would most of the movement taken place? Would the modern universities have fared better? Our analysis suggests that the 1960s New Universities would have lost the most, but the gain to modern universities is very selective.

Four Universities would have moved up two ratings: Cranfield; Plymouth; Aberystwyth; and Sheffield Hallam. The table displays the lists of Universities ‘gained’ or ‘lost’ one rating when the proposed metric was applied to the 2001 data.

Table 3 Universities whose metric rating change by one rating point	
Losers	Winners
Lancaster	Loughborough
Aston	Brunel
UMIST	Birmingham
L.S.E.	Kent
Reading	Manchester Metropolitan
Portsmouth	Middlesex
Exeter	London Metropolitan
Surrey	Nottingham Trent
Stirling	Westminster
St Andrews	Oxford Brookes
East Anglia	Coventry
Hertfordshire	Derby
Glasgow Caledonian	East London
Wales, Swansea	Sunderland
Bournemouth	Bolton Institute
Gloucestershire	
Wolverhampton	
Greenwich	
Abertay Dundee	
Staffordshire	
London Guildhall	

For over half the Universities, the use of metrics would have made no difference to their 2001 RAE rating.

As a final indicator of the implications of the proposed metric, the change in the institutions 2001 rating from the one judged to that of the metric can be correlated with the component variables of the metric.

Table 3 suggests that those institutions that would have gained had shadow metrics of the kind proposed in this note been used were those with the highest OST income (non-OST income is neutral in its effects on changes in ratings). Quality as measured by the quality index and the component quality indicators are correlated with improvements in metric rating over actual rating awarded in 2001. The highest correlation in Table 3 shows that it was the institutions with the lowest awarded rating that had most to gain from the use of metrics.[5] Interesting as these findings are, none of the correlations is sufficiently large as to suggest that the proposed metric is unduly biased in its implications for changes in rating allocations. The shift to metrics will undoubtedly mean change for most institutions and it is clearly important for subject areas to take a view on the metrics they feel are most appropriate for their unit of assessment.

Table 3 Correlates of improved 'metric' rating

	'gain'
OST Income	.238
Non-OST Income	.036
No. of Staff	-.114
No. of Doctorates	.014
Journal Ranking	.193
Citation Impact	.211
Input Volume	.165
Throughput Volume	-.054
Volume index	.068
Quality Index	.208
2001RAE Rating	-.275

Concluding remarks

This note has shown that it is possible to construct a relatively simple and transparent metric made up of four measures of volume and two measures of output quality. The proposed metric makes explicit assumptions about the appropriate balance of volume and quality indicators and includes a balance between income from OST and non-OST sources that might be in line with government plans for the future of research funding in general.

The analysis shows that the proposed metric is capable of reproducing the panel judgements made in 2001, although, as we will see, there are significant differences in the ratings awarded by the panel and those predicted by the metric. These differences highlight the possible implications of the use of the proposed metric and similar metrics in the 2008 exercise.

The use of the proposed metric would ensure a considerable degree of continuity in funding should it influence decisions made by the 2008 panel. This stability in funding flows from the decision to give equal weight to quality and volume. Additional continuity would result from a decision to apply the formula to only 75% of the budget, using the remaining funds to ‘dampen’ the impact of any reallocation.

[5] Empirical analysis suggests that the ‘gap’ between peer review outcomes and implied metric outcomes was greatest amongst those intuitions grades ‘3a’ and ‘3b’ (Kelly et al Forthcoming.)

Inevitably, there will be winners and losers in the shift towards the use of metrics. Overall, there would seem to be some 'gain' for the lower rated institutions and some 'loss' for some of the new Universities. Overall, the item analysis suggests that the proposed metric is relatively neutral in its impact on a group of Universities with very different research profiles.

The proposed metric would encourage institutions to increase the scale of their activity and improve the quality of published outputs. On the volume side, the equal weight given to non-0ST sources despite being relative smaller in total that OST sources, will possibly encourage additional collaborative research with partner organisations in the public and private sectors. The proposed metric would provide clear guidance to institutions on what is needed to improve their research rating. It implies that a relative increase in the volume of research activity as measured by the four indicators, and an increase in the number of articles published in top rated and cited journals will lead to a higher research rating and an increased proportion of the available research funds.

It is hoped that the proposed model and the analysis of the 2001 RAE data will stimulate further discussion about the use of metrics in the BMS RAE2008 and subsequent future rounds of research assessment.

Appendix: Percentile ranks for variables in the proposed metric by institutions ranked by proposed metric.															
	OST income	Other Income	Academic Staff	PhDs	Mean BBS 2004 Rank	Mean Citation Impact	Input Rank	Throughput Rank	Volume Rank	Quality Rank	2001 rating	2001 Num. Rating	Metric Rank	Metric Rating	Rating gain
London Business School	95	100	99	93	100	100	97	96	98	100	5*	7	100	7	0
Cranfield	82	96	89	90	93	98	89	90	94	95	4	5	99	7	2
Warwick	100	98	100	100	90	85	99	100	100	88	5*	7	98	7	0
City	84	84	66	89	99	99	84	78	86	99	5	6	97	6	0
Bath	99	89	91	75	87	81	94	83	91	84	5	6	96	6	0
Oxford	85	83	76	73	95	94	84	74	84	94	5	6	94	6	0
Cardiff	19	95	98	98	97	95	57	98	81	96	5	6	94	6	0
Imperial College	81	93	78	91	76	86	87	85	90	81	5	6	93	6	0
Lancaster	90	62	95	97	88	79	76	96	88	84	5*	7	91	6	-1
Nottingham	19	87	85	78	86	97	53	81	72	91	5	6	90	6	0
Manchester	97	94	87	95	60	74	95	91	96	67	5	6	89	6	0
Cambridge	19	85	86	85	81	93	52	86	72	87	5	6	88	6	0
Leeds	19	73	90	88	89	89	46	89	69	89	5	6	87	6	0
Loughborough	87	90	69	80	68	80	89	74	85	74	4	5	87	6	1
Brunel	96	63	17	68	98	96	79	43	64	97	4	5	85	6	1
Birmingham	86	67	55	80	77	83	77	67	75	80	4	5	84	6	1
Aston	80	79	94	82	62	70	79	88	88	66	5	6	83	5	-1
Glasgow	64	82	38	58	85	87	73	48	64	86	4	5	82	5	0
UMIST	19	99	96	99	66	66	59	97	82	66	5	6	81	5	-1
LSE	98	64	88	87	46	64	81	88	89	55	5	6	80	5	-1
Plymouth	67	80	24	62	94	84	73	43	61	89	3b	3	79	5	2
Birkbeck College	72	71	46	54	61	91	72	50	64	76	4	5	77	5	0
Strathclyde	79	78	97	94	43	59	78	95	88	51	4	5	77	5	0
Bradford	70	69	67	95	74	51	70	81	79	63	4	5	76	5	0
Reading	19	74	72	82	78	69	47	77	64	73	5	6	74	5	-1
Hull	44	86	49	87	70	57	65	68	72	64	4	5	73	5	0
Keele	69	35	82	58	57	68	52	70	65	63	4	5	72	5	0
King's College	74	88	52	24	79	53	81	38	63	66	4	5	71	5	0
Aberystwyth	94	3	34	49	91	65	48	42	46	78	3b	3	70	5	2
Southampton	66	28	65	65	80	55	47	65	59	68	4	5	69	5	0

	OST income	Other Income	Academic Staff	PhDs	Mean BBS 2004 Rank	Mean Citation Impact	Input Rank	Throughput Rank	Volume Rank	Quality Rank	2001 rating	2001 Numeric Rating	Metric Rank	Metric Rating	Rating gain
Edinburgh	78	81	71	58	44	40	79	65	76	42	4	5	69	5	0
Queen's Belfast	19	59	77	49	64	72	39	63	52	68	4	5	67	5	0
Sheffield	19	72	60	61	63	71	46	60	55	67	4	5	66	5	0
Kent	89	37	40	72	65	49	63	56	64	57	3a	4	65	5	1
Manchester Met	63	76	60	77	40	43	69	68	73	41	3a	4	64	5	1
Royal Holloway	49	7	70	46	84	67	28	58	41	76	4	5	63	5	0
Middlesex	59	60	81	70	35	48	59	76	72	41	3a	4	62	5	1
Sheffield Hallam	56	13	17	65	71	90	35	41	35	81	3b	3	60	5	2
Heriot-Watt	68	55	55	54	56	54	62	55	62	55	4	5	60	5	0
Open U.	77	39	80	41	50	45	58	61	65	47	3a	4	59	4	0
Portsmouth	88	68	84	68	26	38	78	76	80	32	4	5	57	4	-1
Ulster	83	77	93	68	32	19	80	80	85	26	3a	4	56	4	0
Exeter	19	11	13	65	96	88	15	39	20	92	4	5	55	4	-1
Kingston	62	51	46	41	69	46	56	44	54	57	3a	4	54	4	0
Surrey	19	97	64	80	52	29	58	72	69	40	4	5	53	4	-1
London Met	65	22	31	2	82	78	44	17	28	80	3b	3	52	4	1
South Bank	19	65	40	7	83	73	42	24	28	78	3a	4	51	4	0
Salford	91	27	38	37	55	47	59	37	50	51	3a	4	51	4	0
Nottingham Trent	47	47	78	54	49	44	47	66	59	46	3b	3	49	4	1
De Montfort	45	41	73	41	48	50	43	57	49	49	3a	4	47	4	0
Northampton	76	32	11	18	73	56	54	15	36	65	3a	4	47	4	0
Durham	73	53	51	46	47	35	63	48	61	41	3a	4	46	4	0
Brighton	93	66	57	58	19	24	79	57	72	22	3a	4	45	4	0
Stirling	50	38	74	84	45	32	44	79	62	38	4	5	44	4	-1
Leicester	19	12	17	72	67	77	15	44	26	72	3a	4	43	4	0
Aberdeen	19	34	33	7	72	82	27	20	18	77	3a	4	41	4	0
St Andrews	60	36	30	30	39	62	48	30	39	51	4	5	41	4	-1
Luton	61	31	24	41	41	60	46	33	38	51	3a	4	39	4	0
West of England	19	52	69	24	54	39	36	47	41	47	3a	4	38	4	0
Westminster	71	91	57	24	22	26	81	41	64	24	3b	3	37	4	1
East Anglia	52	4	24	24	53	76	28	24	20	64	3a	4	36	3	-1

	OST income	Other Income	Academic Staff	PhDs	Mean BBS 2004 Rank	Mean Citation Impact	Input Rank	Throughput Rank	Volume Rank	Quality Rank	2001 rating	2001 Numeric Rating	Metric Rank	Metric Rating	Rating gain
Hertfordshire	19	30	62	52	31	52	24	57	37	41	3a	4	35	3	-1
Glasgow Caledon	55	54	83	37	16	13	55	60	62	14	3a	4	34	3	-1
Liverpool J.M.	19	70	31	62	15	34	45	47	46	24	3b	3	32	3	0
Swansea	43	23	44	75	17	36	33	59	45	27	3a	4	32	3	-1
Northumbria	54	50	35	14	38	30	52	24	38	34	3b	3	31	3	0
Oxford Brookes	40	19	48	37	51	31	30	42	32	41	2	2	30	3	1
Glamorgan	19	56	62	24	7	23	38	43	40	15	3b	3	29	3	0
Huddersfield	19	29	44	52	37	21	24	48	33	29	3b	3	27	3	0
Napier	19	49	21	34	30	41	34	27	28	36	3b	3	27	3	0
Lincoln	38	61	53	18	14	14	49	36	43	14	3b	3	26	3	0
Bournemouth	19	33	28	30	59	20	26	29	22	39	3a	4	24	3	-1
Coventry	19	16	7	18	36	61	18	13	8	48	2	2	23	3	1
Derby	19	43	60	46	18	11	31	53	40	14	2	2	22	3	1
East London	19	2	14	11	35	63	11	13	5	49	2	2	21	3	1
Gloucestershire	19	44	51	34	23	18	31	42	34	21	3a	4	20	3	-1
Sunderland	57	10	21	41	24	22	34	31	29	23	2	2	19	3	1
Wolverhampton	39	5	15	37	27	37	22	26	18	32	3b	3	18	2	-1
Newcastle	48	14	41	18	20	28	31	30	26	24	2	2	17	2	0
Buckinghamshire	19	20	19	14	33	33	20	16	12	33	2	2	16	2	0
Central England i	53	48	5	30	10	12	51	18	35	11	2	2	15	2	0
Queen Margaret	41	45	36	30	6	6	43	33	35	6	2	2	14	2	0
Greenwich	19	24	24	4	28	27	22	14	11	27	3b	3	13	2	-1
Paisley	51	18	28	46	9	4	35	37	33	6	2	2	12	2	0
Abertay Dundee	19	40	11	11	29	7	30	11	15	18	3b	3	11	2	-1
Bolton Institute	19	57	10	7	3	15	38	9	20	9	1	1	10	2	1
Solent	19	17	4	24	13	16	18	14	10	14	2	2	8	2	0
Anglia Poly	19	46	27	24	2	5	32	26	25	4	2	2	8	2	0
Robert Gordon	19	26	2	18	12	17	22	10	9	14	2	2	6	2	0
Staffordshire	19	15	44	7	11	9	17	26	14	10	3b	3	5	2	-1
London Guildhall	19	9	9	7	21	2	14	8	3	12	3b	3	4	2	-1
Leeds Met	19	21	7	14	5	10	20	10	9	7	2	2	4	2	0
Trinity All Saints	46	1	1	2	4	2	23	2	8	3	1	1	2	1	0
Dartington	19	6	3	2	1	2	13	3	2	2	1	1	1	1	0