

## Scale economies in academic excellence: an exploratory analysis of the United Kingdom's 1992 research selectivity exercise

ANTHONY G. HOARE

*Senior Lecturer in Geography, University of Bristol, Bristol, BS8 1SS, U.K.*

**Abstract.** Taken in aggregate, bigger university departments did disproportionately well in the 1992 U.K. Higher Education Research Selectivity Exercise (RSE). A number of reasons are reviewed whereby such an 'economies of scale' effect might apply both in general and with respect to the RSE. A methodology is developed whereby the RSE performance of the UK's universities across academic units can be attributed to 'size' and 'non size' components, the relative importance of which are then calculated for each of the 'old' universities, paying attention to their independent designations as 'research' and 'teaching' institutions. Possible implications for academic planning by universities are drawn out.

### Introduction

The subject of this paper can be introduced diagrammatically by means of Figure 1. This shows the relationship among the United Kingdom's 56 'old' universities<sup>1</sup> between their research rating, on the latest (December 1992) peer group review from the Universities Funding Council (UFC) ('the review' hereafter), and the average size of their academic subject groups ('departments') assessed in this same exercise, as measured in full-time academic staff (*University Statistics, Vol. 1, Table 31: USR 1992*). The clear message from the scatter plot is of a positive relationship between the two. Many universities with high research ratings, as a percentage of the maximum attainable given the number of academic 'units' over which they were assessed, also display high average departmental sizes, while a number of others lie at the low ends of the same two spectra.<sup>2</sup> Oxford, Cambridge, University and Imperial Colleges London exemplify the first, and Aberdeen, Queen's Belfast, St. David's Lampeter, Keele, and Bangor the latter. On the other hand some universities, such as Birkbeck College, London and the London School of Economics have performed well in the review despite relatively small departments while some like Aston, Bradford and Ulster have achieved somewhat lower UFC ratings than expected, given their average departmental size.

Although a useful scene-setter, this scatter-plot is necessarily an over-simple representation of the relationship to be explored further in what follows – between ratings in the review and sizes of assessed departments. One obvious limitation of it is that, although the same rating scale (from a maximum of 5 to a minimum of 1) was applied to each academic unit, the 56 universities plotted were evaluated over very different subsets of units,<sup>3</sup> and average departmental size is likely to vary substantially by academic unit. According to the data in *The Higher* of 18 December 1992 for example, the unweighted average size of 'old university'

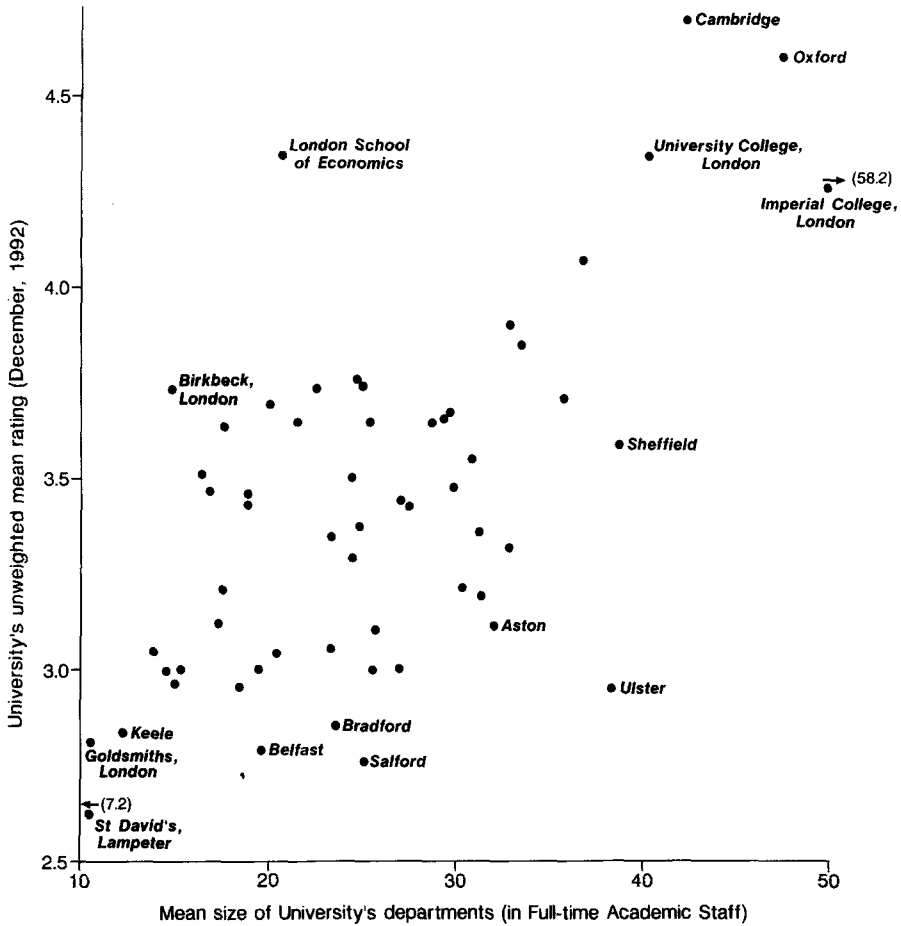


Fig. 1. General all-university relationship between departmental size and UFC research rating.

assessed departments in Economic and Social History was 6.9 full-time equivalent active research staff and that for Italian 4.1, while Civil Engineering averaged 11.9, Chemistry 19.5 and Veterinary Science 54.5. Hence a university with a bias towards science and engineering (such as Imperial College, London) will not only to be assessed over a very different subset of units than, say, St. David's College, Lampeter, with its emphasis on humanities, but also one with much higher average departmental sizes nationally.

It follows that in comparing university dots on Figure 1 we are not comparing like with like. Nevertheless, it helps pose some interesting questions which can be explored below with due account taken of such unit-to-unit, as well as place-to-place, variations in department size and research rating. First, to what extent nationally do the inter-university UFC ratings for any given subject (academic unit) mirror variations in department size, rather than representing variations with other possible causes? Second, how far do different universities vary in the relative

importance of these 'size' and 'non-size' effects? Third, how closely do such variations map onto the increasingly divergent paths that universities have pursued in the search for public funding maximisation in a financial climate that has tightened almost to strangling point? These questions form the substance of the empirical section III below, before which, in Section II, the rationale for the to-be-tested role of department size in research performance and review rating is examined more thoroughly.

### **Academic scale economies and research performance**

For most universities, the review's units of assessment match fairly closely their own structures of academic departments. And these in turn serve as the basic building blocks for universities' internal planning purposes, including the allocation of resources and of student numbers. Since the days of their foundation, when many may have been no more than one-person departments, most have grown significantly, as part of the general expansion of the university sector. But in addition to this system-wide growth, three particular factors have driven departments to become larger, both for teaching and research. First, the breadth of scholarship now required in teaching many subjects to degree level, whether for single or combined honours, has outpaced the capacity of any one scholar to embrace more than a subset at acceptable levels of competence (aside from the time s/he would need to devote to teaching if covering the entire syllabus!). Where degrees have to be validated by some outside body for professional purposes, as with medicine, law, architecture, planning, dentistry, so the need to cover a range of specialisms with expert teaching in each is ramified further. Second, the fixed cost outlay of equipment required for training, say, physicists, engineers and medics is only justified by a large throughput of students, and hence of teaching and support staff. Third, the 'academic mass' arguments apply to research as to teaching, though probably more so in science, engineering and medicine than in the humanities and arts. Much research in the physical sciences is essentially long-term group work, often involving a number of 'tenured' academics, but where this is not so the shorter-term, more fluid patterns of collaboration among researchers as they team up in different ways still gain some advantages from size. Even academics who usually work alone (like the present author) benefit from informal discussions with colleagues, in-house advice on papers in early draft, reaction through seminars on ideas still being formulated, or the competition and rivalry of colleagues in proximate fields.

These last arguments suggest a 'community of scholars' research advantage to academics in large departments, derived from interpersonal interaction. And even those choosing to keep themselves entirely to themselves in their academic endeavours can still profit from departmental size. The more numerous one's colleagues, the more frequent the chance for time release for sabbatical or study leave and the less the recurrence interval of time- and mind-clogging major departmental chores such as admissions, examinations and the headship.

So however they may impact, there clearly seem a number of benefits of size, particularly for academics who are active researchers. These may attract a disproportionately high-quality field of applicants for posts in large departments over small, which reinforces, through enhanced reputation and research productivity, their previous advantages as places in which to work.

One consequence of these influences, for many academic subjects, is that a minimum size exists for a university department, below which it is difficult to deliver a currently acceptable level of 'output' (research and teaching). I term this the 'minimum academic size' (m.a.s.) for any academic discipline. This basic idea appears in many of the surveys commissioned by the University Grants Committee of specific subjects nation-wide, between 1986 and 1989. Table 1, drawing on my synthesis of these surveys (Hoare 1991a), shows their albeit tentative estimates of m.a.s., and how they vary significantly from one subject to another.<sup>4</sup> By implication, any department falling below its relevant m.a.s. is likely to be performing less than optimally both for its institution, for its discipline and for the university system more widely.

This approach is developed further in two ways in this present paper in the specific context of the December 1992 review. First, attention is inevitably confined to the benefit of 'size' for *research*, since, whatever, the links between quality research and quality teaching, it was only the former that the review explicitly examined. Hence no parallels can necessarily be drawn between what follows and the provision of effective teaching at university level. Second, I suggest that the m.a.s. concept can be extended to a family of such values, each specific to a particular level in the hierarchy of research excellence as specified by the review, and represented by the 5-point scale it then applied to each assessed university department. As departmental size increases so the opportunities to produce research of higher quality and over a wider range of areas within each discipline should increase too, following the previous arguments, even if, as with the data in Table 1 the actual sizes appropriate to these hierarchical levels will vary among academic subjects. Note that this does not assume that each department reaching a particular

Table 1. UGC reviews: suggested sizes of single honours academic departments

Subject	Size threshold
Veterinary education	36 clinical staff as minimum
Earth sciences:	
Level 1 (teaching and research)	30 academic staff
Level 2 (honours teaching)	> 15 academic staff
Level 3 (limited teaching)	Perhaps 6 or less academic staff
Chemistry	20 academic staff (including 3 professors) as minimum
Physics	20 academic staff as minimum
Accountancy	10 academic staff as minimum
Philosophy	9 academic staff as minimum
Sociology	8 academic staff as minimum
Social policy and administration	5 academic staff as minimum
Dentistry	50 undergraduates per year as minimum

m.a.s. will *necessarily* produce research of the corresponding quality, merely that evidence from the university system as a whole suggests that such a level is *possible* with that size of department. Indeed, to the extent that departments fail to achieve the research rating implied by their m.a.s. then size alone cannot explain research excellence, thereby perhaps contributing to the scatter of points evident in Figure 1.

Before converting these arguments into an operational research methodology consider how such 'benefits of size' arguments relate explicitly to the mechanics of the previous (1986, 1989) University Grants Committee and current UFC research reviews, in two further ways. First, as Johnes and Taylor (1990) have argued, large departments benefit from the greater chance that review panel members have personal acquaintance with colleagues there and the department's work, as research collaborators, external examiners, academic visitors or even former members. Second, in both 1989 and 1992 research excellence took into account the performance of each department across a series of academic 'areas' within the relevant parent unit, in each of which research was judged as of international, national or less than national significance. These 'areas' were defined for each submitted department partly by the categories into which the relevant university chose to subdivide its work and partly through the judgement exercised by the appropriate review team. Hence this number of areas identified could well vary among the universities submitting within any one academic unit. Each was then assessed solely with respect to the areas where they were thus deemed to be active, with no penalty for an absence of submitted research in other areas significant in the equivalent departments in other universities. However, the earlier scale economies arguments would suggest that bigger departments should not merely be active in more areas than small but should also have more researchers within each of its areas as defined for the review, all disproportionately productive of good research as compared to the smaller numbers in equivalent departments elsewhere.

So, on this basis, the larger departments could have a higher 'visibility' with the reviews teams, benefit from the way they go about their job and also produce more good quality work to set before them, both overall and in each of the 'areas' on which their evaluation was based. In 1986 and 1989, every academic member of staff was entered in the review exercise, and the departments as a whole could enter a pre-specified maximum number of pieces of work from the staff in aggregate. And in 1992, when departments could choose which colleagues to submit for review, bigger ones enjoyed the same benefit of having more choice not just of which research to present but also of which colleagues to select.

It comes as no surprise to find that some single-subject post-mortems of the earlier reviews identified departmental size as one factor in the variation among institutions in research ratings. Thus a critical mass of 21 full-time staff necessary for a 4 or 5 rating emerged from analysis by the History at the Universities Defence Group (*Times Higher Education Supplement*, 7 September 1990), while equivalents by Edwards (1991) for the 1989 review and Gleave *et al.* (1987) for that of 1986 suggested the same general benefits of size for Geography departments, as well as underlining some serious implications for smaller departments, disadvantaged by the then prevailing methodology. Similar 1986 'size-rating' relationships identified

in other subjects are summarised by Platt (1988) although, as she notes, the ratings themselves were often disputed, and bore little or no similarity to differences in research output *per caput* among the departmental academic staff so assessed. Taken together, at the very least this would imply that the rating benefit enjoyed by larger departments does not hang solely on 'objective' measures of academic output and performance.

So to summarise so far, a number of *prima facie* reasons suggest not only that larger departments will have the advantage in aggregate and *per caput* research output over small, but also that for these and other reasons, associated with the workings of the research selectivity reviews, this advantage should translate into higher peer group ratings. But while the latter seems supported by a number of independent subject-specific post-mortems of these reviews, this literature also finds little convincing explanation for this 'big department' success in terms of their higher research productivities in reality, despite their supposed advantages in this respect.

However, as well as being confined to individual subjects, none of these previous studies has attempted to distinguish quantitatively a size and non-size component in the research review outcomes in the way possible using the m.a.s. concept, and not just for individual subjects but also for whole institutions. This approach also serves to shed a new and different light on the size-research rating relationship which, difficult though it may be to account for, is clearly still part of the conventional wisdom of peer group reviews. I now turn to the methodology adopted for such a detailed analysis of the 1992 results, to explore thereby the significance of departmental size among these latest data.

## Analysis

### *Methodology*

The methodology adopted can be understood through the research ratings for any given academic unit across a number of individual universities, as shown in the hypothetical example of Figure 2. These ratings broadly increase with departmental size, but within any one assessment level a spectrum of sizes still emerges. Equally, certain departments of similar size achieve different ratings. Assume now that the m.a.s. for a given rating level is identified by the *smallest* such department submitted by any institution in the review<sup>5</sup> which attains that rating. Hence, in Figure 2, a size of 25 staff is the necessary minimum to attain a '5', one of 17 for '4' and so on, with everything below 4 in size (the m.a.s. for '2') receiving a '1' rating.

Some university departments (examples U1–U5) attain the rating appropriate to their size: all '5' scorers like U5 come into this category by definition, whereas the other four (U1–U4) all fall below the m.a.s. necessary for a higher rating. These could not reasonably have achieved more than they did, given the performance of the competitor departments elsewhere. Conversely, some other departments fail to attain the attainable rating for their size, and fall below it by one, two, three or even

### HYPOTHETICAL ACADEMIC UNIT

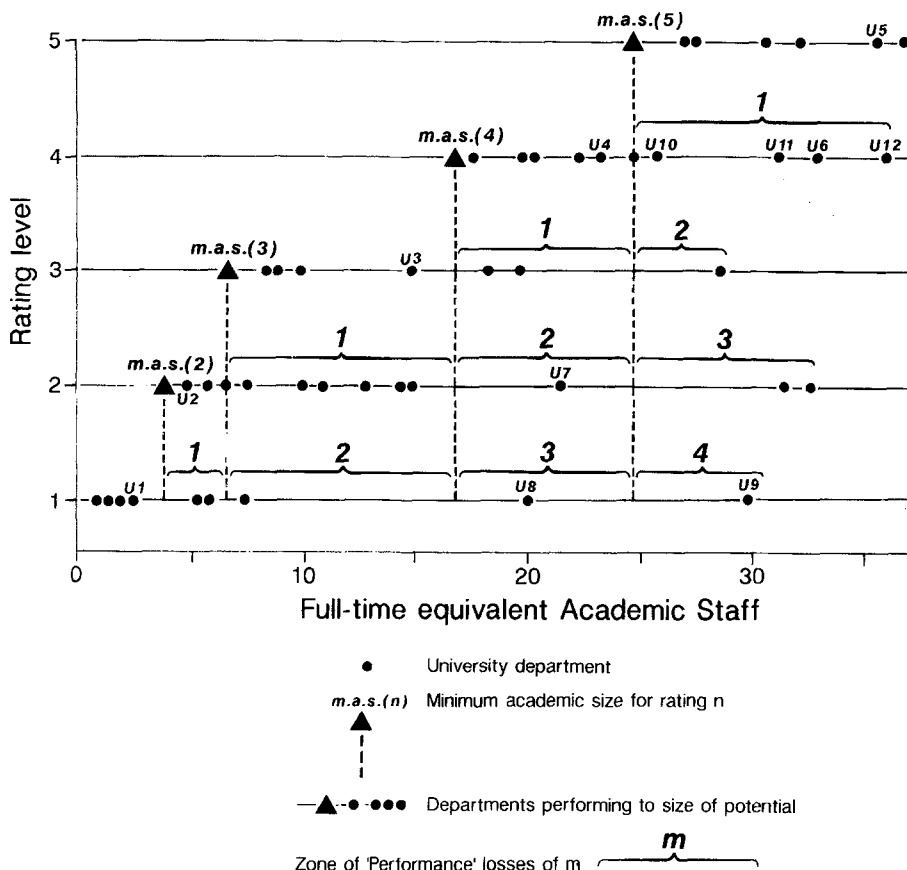


Fig. 2. Hypothetical academic unit to demonstrate methodology employed.

four levels (examples, U6, U7, U8, U9 respectively). If we assume that any rating less than 5 represents a 'shortfall' from the maximum attainable, we can disaggregate this further into these two components, 'size' and 'performance', corresponding respectively to the circumstances just described. Thus university departments U2, U7, U8 all have shortfalls of three, but whereas in U2 this is entirely attributable to size (it could not reasonably have done better), and in U8 to performance (it could reasonably have done much better), U7's shortfall of 3 can be attributed to one rating level lost due to size (the maximum attained elsewhere for a department of that size being a '4') plus a further performance loss of two. In reality, not all academic units were as well behaved as that of Figure 2, as I show below, but it establishes the general principles which can be applied to a range of actual experiences.

By aggregation over all the relevant academic units on which it was assessed, a given university's overall attainment in the review can be separated into the

contributions made by performance and by size to its shortfall, the amount by which it fails to attain the maximum score of '5' for all its assessed units.<sup>6</sup> At one extreme this could be entirely attributable to size, if all its departments perform up to their size potential. At the other, all could be of a size equal to or greater than the relevant academic unit m.a.s. value for a '5', but fall below this on performance. In reality, of course, some intermediate position is more likely, representing a mixture of these two components. More formally, the relevant indices for any given university are defined as follows.

Shortfall:

$$[(N_u \times 5) - \sum_u (R_u)] / (N_u \times 5) \times 100\%$$

Size:

$$([\sum_u (Max_u)] / \text{Shortfall}) \times 100\%$$

Performance:

$$([\sum_u (Max_u - R_u)] / \text{Shortfall}) \times 100\%$$

When:

$N_u$  is the number of assessed units  $u$  (and departments) in a particular university

$R_u$  is the rating accorded to department  $u$  in that university

$Max_u$  is the m.a.s. rating equivalent to the size of department  $u$  in that university

Clearly, the methodology just outlined takes a particular stance on the question of the most useful index measure of the size profile of university departments to employ when looking at its pattern of variation with research ratings. It might seem to some readers that the mean size per rating level would be a more obvious and suitable criterion than the m.a.s., paving the way for some statistical analysis of the size/rating relationships. However, this would need to be on a unit-specific basis (pooled across-unit data fall foul of inter-unit variance in the relevance of 'size', as discussed above), and it is unlikely that many such unit distributions would meet the technical preconditions of further categorical data analyses. Furthermore, and more serious, are the practical implications, for university management. Thus if mean size was the chosen index a department such as U10 in Figure 2 could argue for more resources to bring it up to the mean size of the higher (5) rating level – a mean size that is probably a statistical artefact, occupied by no department anywhere in practice. However, in practice its university is likely to be more 'impressed' by the fact that it failed to match the grade of the *actual* department which forms the m.a.s. for the 5 level, on which basis its scarce resources might be



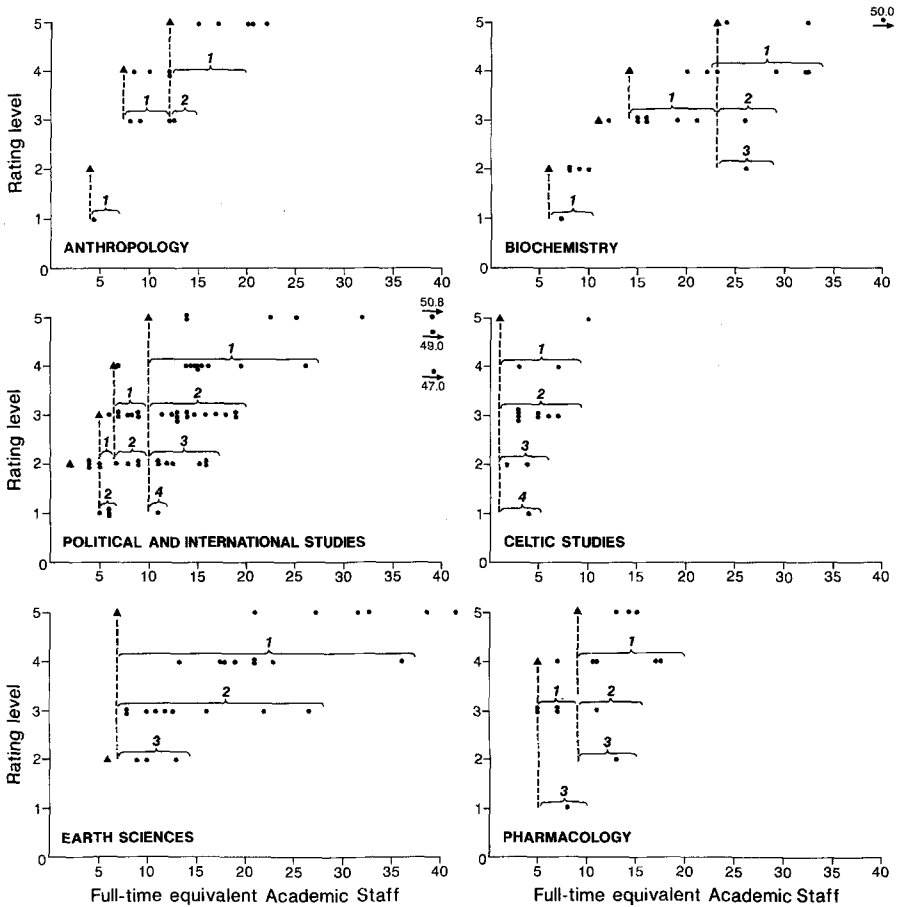
better directed elsewhere – to departments falling below the m.a.s. thresholds for a higher performance – and with very different questions asked of the U10 management. This is not to argue that some readers and institutions might not find a ‘mean’ index value more useful in their own contexts, but rather that the practical value of the approach adopted here is still considerable.

Before moving to the results some further complications, limitations and variants of the methodology should be noted. The first *complication* is that not all rating levels can be identified in every academic unit. Not only do the review panels not always use each point on the 1–5 scale in their assessments, but also the m.a.s. score of a higher level sometimes absorbs all the individual university departments accorded lower ones. Figure 3a shows a sample of the academic units to give a flavour of the variety that emerges, ranging from the well-behaved Anthropology, Biochemistry and Political and International Studies to the more eccentric Celtic, Earth Sciences and Pharmacology where middle rating tiers are squeezed or the hierarchy truncated. Second, within some academic units further subdivisions are provided for a few universities where the importance of the relevant discipline makes a finer-grained subdivision appropriate. In almost all cases I treat these subdivisions as part of the same assessment system as the parent unit as a whole, though a small minority are excluded as seemingly very specialised and potentially distorting to the analysis.<sup>7</sup> However, where separate ‘basic’ and ‘applied’ assessments are provided within one parent academic unit each is treated as a different, self-contained academic unit for purposes of analysis.

Next, some general *qualifications*. First, the institutional measurements above are unweighted in the sense that they make no allowance for the different sizes of those departments within each university. In practice, when, as later, government cash was disbursed partly on the basis of the review, universities benefited *ceteris paribus*, where their shortfall came primarily from their small departments. Second, and more serious, is the assumption that the smallest performing department on any rating level represents its m.a.s. It is unrealistic to expect such a conveniently-sized department to be present for each and every rating level and academic unit. In practice, the ‘true’ m.a.s. for a given level could be lower than that identified in the methodology, but remain hidden as no university has entered such a smaller department, working to its research potential. The effect will be to overestimate size relative to performance. At other times the opposite might result if the m.a.s. identified for a particular level is set unrealistically low.<sup>8</sup> There may be additional factors uniquely underpinning the research output in one institution which it is unreasonable to expect other institutions to emulate. Without inside knowledge of each and every academic unit there is no way to take this further, but at least such *a priori* aberrant cases are relatively uncommon in the raw data, and unlikely to disrupt the overall results. Third, no distinction is made in the methodology among the performance losses of the sort illustrated by examples U10, U11 and U12 in Figure 2. Each receives a performance score of 1 even though the ‘failure’ in the first is a very narrow one and of the others, on the face of it, progressively less excusable.

As a final qualification, of course, the review teams may simply have ‘got it wrong’. Their judgements are not open to appeal but there will be some claims of

a) Mode 1



'Foul', especially from units receiving the lower ratings. This has to remain unproven: such peer group reviews are obviously controversial and their methodologies liable to heavy criticism (Johnes 1989, Johnes and Taylor 1990), but, equally, little independent evaluation of the work of the 72 review panels is available to the author (or to any of their critics).<sup>9</sup> Furthermore, the assessments will form a major influence upon money disbursements among institutions so, right or wrong, it makes sense for universities to understand how their general performance has been derived, notwithstanding any qualms they may retain about some of their ratings.

An important *variation* is introduced into the analysis once we consider the crucial, but hitherto glossed-over, question of what we mean actually by 'size'. As indicated earlier, the 1992 review differed from its predecessors in allowing universities to enter less than the full academic complement of staff for assessment. It permitted universities to decide how many of their staff in each department they considered as 'active researchers', and to enter only those who met this criterion.

## b) Mode 2

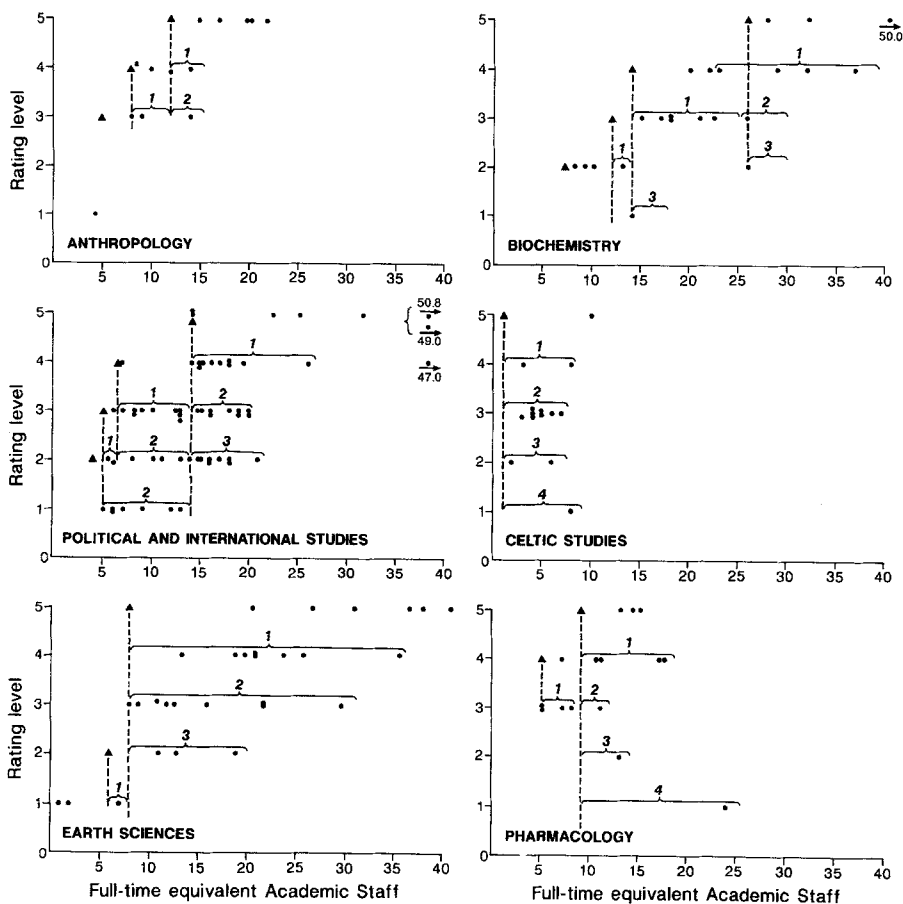


Fig. 3. Case study academic units analysed in Modes 1 and 2.

The official end-of-exercise report from the UFC (1992) records three things for every university entry under each academic unit:

- the rating allocated to each department,
- the departmental academic staff (in full-time equivalent terms) entered, and
- the proportion b) represents of all the eligible staff in that department. This last is measured on an A-F scale, mostly representing 20-percentage point bands of the fraction of entered staff.

Hence, one measure of department size in the terms of the previous discussion is that reported under b), i.e., the 'active researchers' as defined by each university. Given that 16 active researchers in Chemical Engineering at Cambridge and at University College London can (and do) attain a '5', the fact that Birmingham's 29 only attain a '4' and Bradford's 31.5 a '3' shows these latter as falling short, in

performance terms, of what a corpus of researchers of their size could attain. This will be termed the Mode 1 approach in the following analyses, and is the basis for the m.a.s. model as illustrated in Figure 3a.

However, this also raises two problems. First, no independent check is made on how each university/department defines its 'active researchers', so it is quite possible, as in Figure 4a, for two departments, otherwise identical in their mix of active and non-active researchers, to convey very different impressions of research excellence. Each attains a grade '4', but Y's entered size places it above the '5' m.a.s. attained by Z, 'identifying' it, harshly in the event, as falling short on performance terms. But X, though identical to Y for all operational purposes, escapes this opprobrium. Second, as Figure 4b suggests, Mode 1 is inappropriate when reviewing the research rating of an entry against the total departmental resources it can bring to bear on its research. Hence Y now attains a '4' with a research team not only larger than X's but backed by many more colleagues and resources to undertake the teaching programme and departmental chores. In some ways X's performance might seem the more commendable, but the Mode 1 approach will label it as falling short on performance, by setting the common submitted size of X and Y as the m.a.s. for a '4' rating.

The solution, the Mode 2 approach, is to measure size by the total number of academic staff in each entered department, irrespective of whether these have been included in its 'research active' entry. Unfortunately, this total size has not been made public, nor does UFC intend to make it so (Higher Education Funding Council for England; personal communication), leaving as the second-best option that of grossing-up the size figure reported under b) by the midpoint of the percentage point range identified under c), and then rounding to the nearest whole number. For entries in the A category (95–100% of staff entered) I made no changes of this sort, whereas, for example, those coded as F (under 20% of staff entered) were increased some six-fold. The Mode 2 equivalents of Figure 3a appear alongside in 3b. A minority of m.a.s. points now disappear completely or move to the right, as where the Mode 1 m.a.s. sizes were based on universities with less than complete (i.e., non-A) staff entries. The effect on other points is also mixed: of those that do change their performance most move 'upwards', but a minority move the other way.

In their different ways both Modes have a role to play. Mode 1 shows the excellence attainments of the research teams of different sizes in different institutions considered on their own terms (*pace* the reporting inconsistencies noted above, about which we can do nothing with the information to hand): Mode 2 indicates, albeit for a less accurate measure of 'size', how the total person-power a department translates into research excellence ratings more effectively in some universities than others.

## Results

Given the previous discussion of the benefits of size, the first and most general observation is as important as it is surprising. The results over 56 institutions and

72 academic units, in both Modes 1 and 2, show how the greater part of the shortfall in the 1992 review derives from performance rather than size. In both Modes the median institutional performance index is about 68% of the shortfall, but it is above 50% for Mode 2 in slightly more universities (49 of the 56, compared to 45 for Mode 1).

Nevertheless, as a second point, in this and other respects the range of results among institutions is very wide. On the performance index it extends between 91%

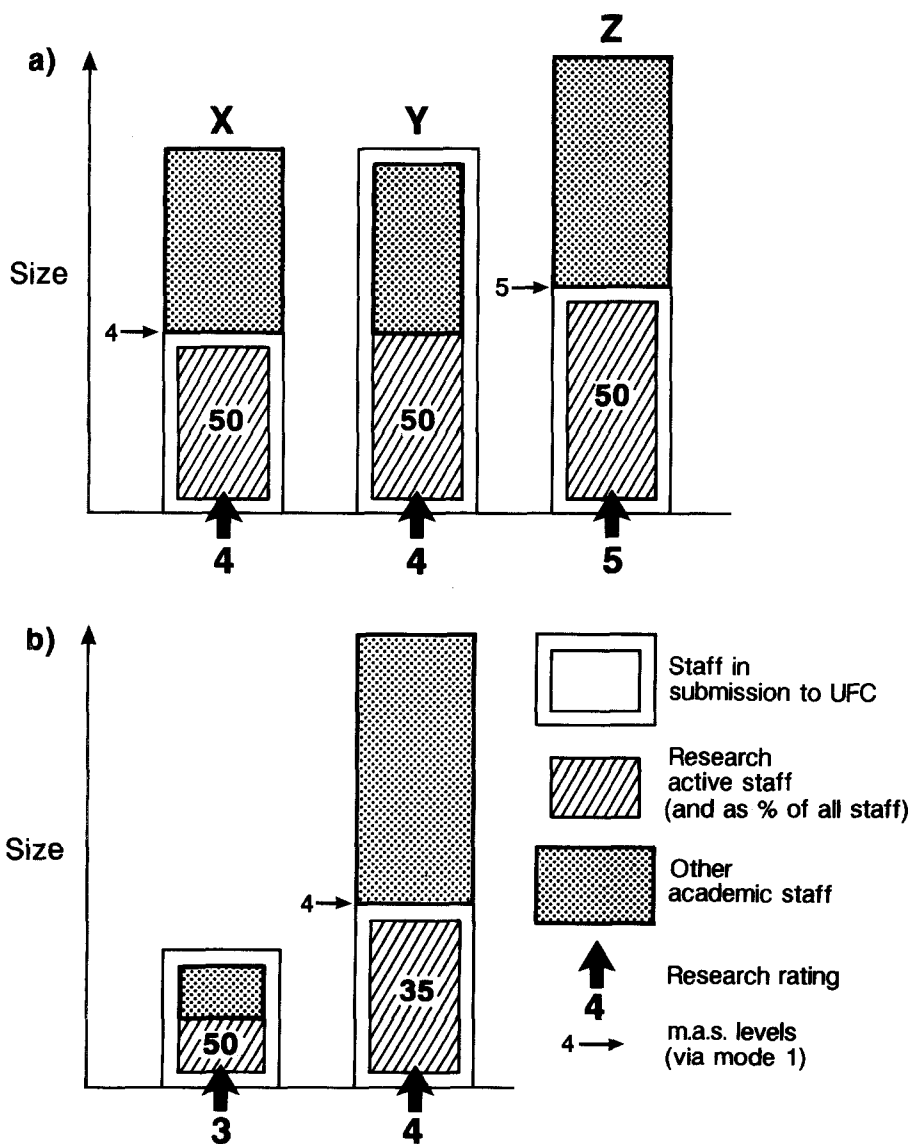


Fig. 4. Problems in measuring 'size' via Mode 1 (see text).

and 38% for Mode 1 and 100% and 38% for Mode 2. Further variation arises when examining the spread of institutional performance losses across academic units, about the median value of 1.5 rating points lost for every unit with a performance-induced shortfall. Table 2 summarises the extreme scores in these respects. Not surprisingly, universities where the performance loss overall is large have some tendency for the loss-per-department rate also to be high (in Bradford, Salford and the City University it is comparatively rare for any department's performance loss to be limited just to one rating level), but the experience over all 56 universities suggests this as not a particularly strong relationship. Equally, no obvious mapped pattern results either, with each geographical region containing a mixture of universities with high and low performance scores and loss-per-subject rates.

Another more interesting non-relationship is that between the overall shortfall of each university on the 1992 review and its relative disaggregation by performance and size. The rank correlation coefficients between the two series are virtually zero, with universities faring very differently in the rating exercise as a whole having very similar performance indices, and *vice versa*.

So, were this not a matter of common sense already, the message from these analyses is that each university intent on understanding its shortfall should analyse its own results for itself (as via the simple methodology outlined above, with such modifications in the light of local circumstances as it thinks fit), rather than looking for any broad-brush lessons-to-apply from the behaviour of the university sector as a whole.

This is reinforced by one further aspect of the analysis. Under pressure to maximise their share of increasingly scarce government funding universities have been 'encouraged' to become more competitive, one with another. Some have opted to follow government exhortations to emphasise their *teaching* ('T') function by increasing student numbers, largely through 'fees only' students paid at below the normal (if still arbitrary) fee income levels for the subjects concerned, while others have opted for a *research* ('R') path, designed to maximise their income from government and other research funding sources. The first UFC annual grant allocation publicly to distribute money to universities separately under 'teaching' and 'research' heads (February 1992) identifies a clear divide, with a group, largely of southern universities, receiving relatively more from the research budget, and the teaching funds being directed disproportionately towards the geographical peripheries. The broader structure of the geography of national higher education of which this is a very salient feature has been explored elsewhere (Hoare 1991b), but it is relevant to ask here how far this research/teaching funding split accords also with the performance in the review some 10 months later, as analysed above.

By and large, those universities performing best in the latter were also those moving in a research direction already. However, when considered against performance and size we find this second discriminant runs across the grain of the research/teaching divide, as Figure 5 clearly shows.<sup>10</sup> Given that many universities have pinned their colours to the research funding mast it is particularly important for that group to understand how their recent assessments came about. This is

Table 2. The extreme performers on this analysis

	Highest	Lowest
Performance as % shortfall		
Mode 1	Bradford (91%) Loughborough (84%) Manchester (82%)	St. David's Lampeter (38%) Aberdeen (39%) Birkbeck College (42%)
Mode 2	Aston (100%) Birmingham (91%) Loughborough (87%)	St. David's Lampeter (38%) Birkbeck College (41%) Aberystwyth (43%)
Performance loss per relevant unit <sup>1</sup>		
Mode 1	Bradford 2.4 Salford 2.1 City 2.0	St. Andrews 1.2 Bath 1.2 Imperial College 1.2
Mode 2	Bradford 2.4 Salford 2.2 Heriot-Watt 2.1	Imperial College 1.0 Bath 1.1 St. Andrews 1.1

<sup>1</sup>unit on which some performance loss occurs.

obviously the more so for those 'R' universities with the higher 'shortfall' scores, where the likely future penalties in research income lost are greatest.

The 'R' universities as shown in Figure 5 fall into two groups. In the 'western' one, containing Exeter, Aberystwyth, East Anglia and Bath for example, such shortfalls have arisen through an approximate balance of size and performance. So any desire on their part to raise their research ratings next time round could lead them to examine the continued justification both for some of the relatively small departments which performed true to type but seem below the m.a.s. for a higher rating, and also others which failed to meet their potential. In some cases, as shown above, departmental ratings could reflect the combined effect of the two, of course. Perhaps some departments do justify an increase in size and, hopefully thereby, of research stature. But also the failure of others to reach their size potentials needs attention. Why was this? What are the realistic prospects of achieving their potential in the next assessment? Already, by January 1993, some universities were introducing very firm steps to learn from the experience of this recent review.

In a different 'eastern' 'R' subset the causes for concern lie more squarely with performance. Despite orienting themselves towards research income their departments often fell short of their size-related research potential. And while certain of these universities did well in the rating exercise overall (Edinburgh, Imperial London and Warwick all being in the top ten non-specialist institutions) others such as Nottingham, Sheffield, Southampton and particularly Glasgow might have more cause for critical internal reassessment, if intent on continuing the research route to funding.

The lessons may seem less immediately relevant for the 'T' universities, though they still cannot be ignored other than by any intent on becoming purely teaching

institutions with no pretensions to research expertise. Those such as Dundee, Keele, Stirling, Aberdeen, Lampeter, and Queen's Belfast, all of which did badly in the review, owe their poor results to a mixture of size and performance. On the other hand, the likes of Bradford, Loughborough, Salford, Bangor, the City University and Cardiff can attribute their shortfalls predominantly to their departments' failure to perform to their research potential, judged by the achievements of other competitor departments elsewhere.

One final aspect of Figure 5 is its showing the extent and direction of any deviation from institutional performance scores as between Modes 1 and 2. Predictably enough from what has gone before, the majority of any shifts are towards the right, as the shortfall element attributable to performance increases once total departmental size is the yardstick. Equally, and consistent with the R:T distinction, those in the latter category not only are somewhat more emphatic in their rightwards movements – evident in 14 of 19 'T' universities compared to 11 of 16 'R' ones – but the size of those shifts is also greater, averaging 5.8 percentage points for 'R' as against 3.8 for 'T' universities. So the general message of the results – that university departmental units are more likely to fall short of their size potential under the more all-embracing definition of 'size' – is especially so for those being channelled/channelling themselves towards a teaching emphasis. Such a distinction is entirely consistent with the orientation of staff and other resources one would expect such places to show over universities where a higher proportion of departmental endeavour is being directed towards research.

## Conclusions

To begin the ending as I began the beginning, with the importance of departmental size, the evidence from this paper is that this is *not* the major control upon the research ratings achieved by different university departments. This is not to deny it plays some part, and more so in some universities and among some academic disciplines than others. It is also likely to be promulgated by some departments, in part as an excuse for their low rating and in part to justify their bidding for more resources on the basis of 'make us bigger and we'll do better'. But, overall, the evidence tells a different story. Size is less important than the inability of departments to achieve what appears to be their academic research potential, in that at least one other comparable department of equal or lesser size has achieved a higher rating than they. The way this disaggregation of the overall shortfall of departments and universities is managed analytically is described above. To the best of my knowledge no such definition of 'minimum academic sizes' for various levels of research excellence and across institutions and subjects has been attempted hitherto even if, again as explained above, the methodology rests on presumptions which may not be justified in each and every case. However, in the absence of any practical alternative, it remains the best option there is!

The second general finding is that different universities vary substantially, both in the relative importance of size and performance on their rating shortfall and in



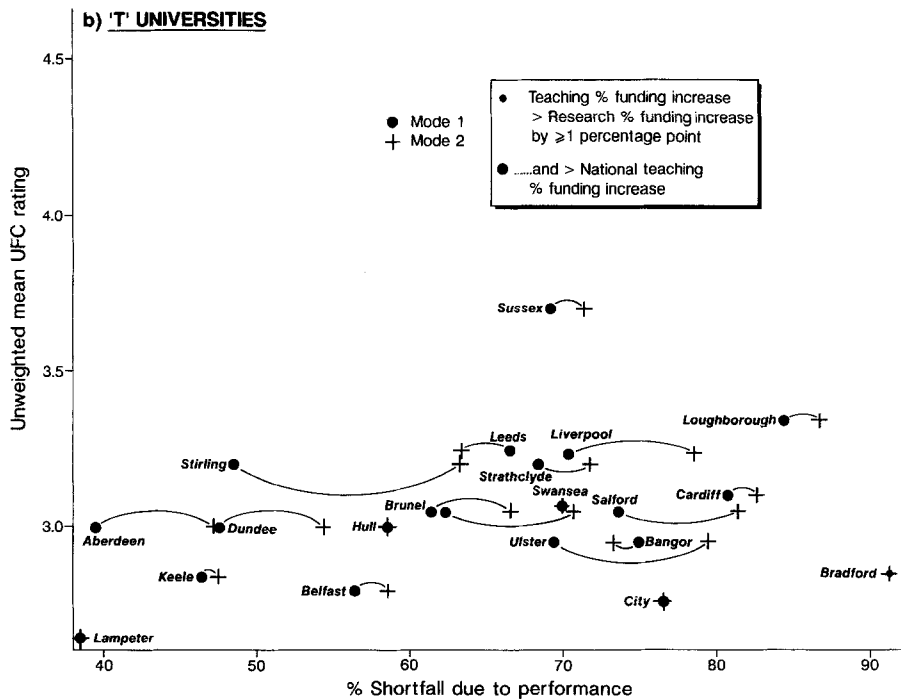
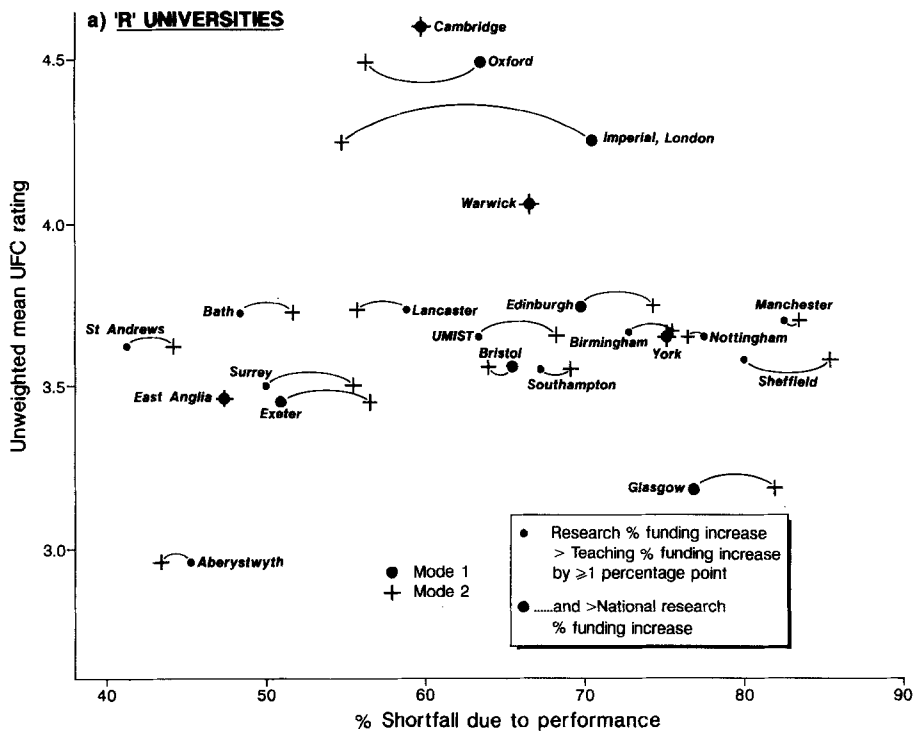


Fig. 5. Results in Research and Teaching universities.

the extent to which performance losses are spread over a number of departments in any institution, rather than concentrated in fewer, weaker, ones. As noted earlier, universities intent on using the lessons of their individual shortfalls in future university planning need to analyse their own position separately, perhaps adapting the approach here in line with local knowledge and other campus realities, limitations and opportunities.

Such future actions will also be driven by the answers to three other questions. First, what type of university is each trying to be in the plannable future? The greater the importance of research and research-derived income the more important the understanding of the inner workings of the review becomes. Second, how well did each university do in that exercise? If 'very well' there may be little that calls for immediate attention on this front<sup>11</sup>, if 'less well' then more major structural rethinking may be called for. And given that the next review period is already well under way, the sooner the better. Third, how will the new Higher Education Funding Councils choose to parameterise the research ratings in future inter-university funding decisions? Those for Scotland, Wales and Northern Ireland, where most universities did poorly in the review, might be minded to provide research funding on a more generous basis than in England, for equivalently-performing universities, lest the research base of their domains is decimated irretrievably.<sup>12</sup> In England, a decision to fund research, say, in inverse proportion to the size of shortfall as weighted by departmental size (as used as the basis for some league tables produced from this review's results by the broadsheet press) could encourage 'R' universities to restructure internally on the basis of an elite set of large, quality ('5'-grade) research departments, so endangering the viability of smaller, lesser ones.

The question of how best to optimise university strategies to maximise research income is clearly more than just that of analysing why all universities have failed to achieve a maximum score in the December 1992 review, as examined here. But the former can only come following a sound knowledge of the latter, even if it needs more besides. This paper is offered as a contribution to that end.

### **Acknowledgements**

I record with thanks the comments of Peter Haggett (Geography, Bristol) and John Wand (formerly of HEFCE) on an earlier version of this paper. The usual disclaimer applies, of course.

### **Notes**

1. Unless otherwise specified, the paper is concerned with the 56 universities existing prior to the abolition of the 'binary divide' in 1992. This includes the main constituent colleges of the universities of Wales and London, but excludes specialist institutions, chief amongst which are

- London's School of Slavonic and East European Studies and its School of Oriental and African Studies, all medical schools and the specialised Business Schools of London and Manchester. The Open University is also excluded.
2. To avoid confusion over terminology the academic subject headings within which the different submissions of each university are assessed are referred to as 'units', whereas the subsections of each university assessed under these headings are 'departments', even though some of the latter do not match precisely the local 'departmental' structure and terminology at each university.
  3. This ranges widely from Edinburgh's 56 assessed units and Cambridge's 52 to St. David's Lampeter's 11 and Aston's 9.
  4. The Chemistry review is alone in specifying an upper limit (of about 50) above which the academic equivalent of diseconomies of scale presumably may set in. Despite the general tenor of many of these Reviews the empirical basis of any 'department research scale economies' thesis is weaker and controversial. Thus Hicks and Skea's (1989) analysis of the 'staff size – research paper productivity' relationship among UK Physics departments suggested it could be accounted for entirely by an 'Oxbridge' effect: for other universities no significant residual relationship existed. (I am grateful to an anonymous reviewer for bringing this work to my attention).
  5. For this part of the exercise the new (ex-polytechnic) universities were also included, since the logic of the m.a.s. model applies equally to them as to the 'old' universities. However, as it is unrealistic to expect the 'new' ones, with their much higher emphasis on teaching, to equate with the research activities of the longer established institutions, including them along with the old universities in the subsequent analysis of shortfall, performance and size would be distorting.
  6. This ranges from a shortfall of 6.2% in Cambridge to 47.4% at St. David's Lampeter, among the 56 universities analysed in the paper.
  7. There is obviously an element of subjectivity in this, not least as I do not have first-hand knowledge of the subjects concerned. As illustration of the resulting decisions taken, the entries for German, Dutch and Scandinavian from University College, London under 'German and Related Languages' were included separately, as were 'Modern Languages' and 'European Studies' under the parent heading of the latter from Bradford. However, the 'Photogrammetry' entry for UCL and 'Surveying' for Newcastle, both under 'Civil Engineering', were excluded. Full details are available from the author.
  8. Hence, for example, the fact that General Engineering at Keele is rated '5' from an 'A' submission from just 1 staff has major effects on the assessment of the remainder of this academic unit.
  9. In the Politics and International Studies academic unit a 'shadow' panel was set up by the Political Studies Association (PSA), which received and assessed duplicates of the official university returns to the UFC. Although its ratings were not identical to those of the equivalent UFC panel in every case, the PSA's own independent assessments were sufficiently close as to allow it to view the official judgements, overall, as reasonable (HEFCE, personal communication).
  10. Not all 56 universities used previously can be included in this analysis. Some fail to meet my criteria for delimiting research or teaching institutions (see Figure 5), while for London colleges other than Imperial College no separate funding figures by these heads were published by the UFC. Equally, although technically a 'T' university on the criteria in Figure 5, Essex also performed very strongly on the 'research income' front and so has been excluded from the analysis.
  11. The reported remarks (*The Higher* 22 January 1993) of its Provost that staff contributing to the failure of some departments to attain a '5' at University College, London would be sidelined from research, in the drive to across-the-board '5' ratings next time, suggests that shortfall may be keenly examined even in very successful universities.
  12. The decision in 1993 of the Welsh Higher Education Funding Council to do just that, funding research in its universities more generously than the equivalently-rated English ones, may be an example others will follow.

**References**

- Edwards, R.A. (1991). 'UK geography departments: a perspective on UFC ratings', *Area* 23, 197–208.
- Gleave, M.B., Harrison, C., and Moss, R.P. (1987). 'UGC research ratings: the bigger the better?', *Area* 19, 163–166.
- Hicks, D., and Skea, J. (1989). 'Is big really better?', *Physics World* 2, 31–34.
- Hoare, A.G. (1991a). 'Reviewing the reviews: the geography of university rationalisation', *Higher Education Quarterly* 45, 234–253.
- Hoare, A.G. (1991b). 'University competition, student migration and regional economic differentials in the United Kingdom', *Higher Education* 22, 351–370.
- Johnes, G. (1989). 'Ranking university departments: problems and opportunities', *Politics* 9, 16–22.
- Johnes, J., and Taylor, J. (1990). *Performance Indicators in Higher Education*. Guildford: The Society for Research into Higher Education.
- Platt, J. (1988). 'Research policy in British higher education and its sociological assumptions', *Sociology* 22, 513–529.
- Universities Funding Council (UFC) (1992). *Research Assessment Exercise 1992: The Outcome* (Circular 26/92). Bristol: UFC.
- Universities Statistical Record (USR) (1992). *University Statistics 1990–91, Volume 1. Students and Staff*. Cheltenham: USR.