The RAE and the research strategy of the School of CIS

Nigel Ling

Summary

There is evidence that RAE scores depend on some quantitative indicators, especially size of submission, in the Computer Science UoA. Panel membership may also be a factor. A close look our School's performance in the 2001 RAE shows that little progress has been made in our RAE standing since 1996 (we are in the bottom 16% of submissions). In the light of this our current research policy is reviewed. Our present productivity is not necessarily related to the distribution of resources. If we wish to improve our research capability we need to encourage a healthy research culture and a rise in the number of research staff; to do this the current allowance system should be replaced with one more equitable. Should we submit to a future RAE, our prospects may be improved by directing some of our current research to other UoAs.

Introduction

This is a paper in which I consider the Research Assessment Exercise, our results within the Computer Science Unit of Assessment, and the implications for the way we approach research, apportion resources and decide future strategy. I have two motivations for this short study. Since the School's research is now circumscribed by the RAE on the grounds that it is essential to improve our score, it makes sense to review the efficacy of this strategy from time to time. To date there has been no serious attempt to do so. Second, I spent some time in the early 1990s working on a major project that investigated the use of objective indicators to measure research quality and how these correlated with RAE scores. This study showed that RAE scores do not necessarily reflect the standard of research output as measured by these indicators. Although it is not practicable to conduct an investigation on this scale, I have compared some of these indicators (where they exist in the publicly available data) with current scores in the Computer Science UoA; the results raise questions about our approach to the RAE.

The following section makes some general observations on the RAE and others' research on the subject. Subsequently I look at the results for the Computer Science RAE and our own performance and research strategy.

The Research Assessment Exercise and its impact

The Government established the RAE in 1992 to provide a measure of research quality for the purpose of allocating resources. This followed on from an earlier research selectivity exercise in the eighties, part of a general strategy to target (and cut) public spending in higher education. There is a surprisingly large body of academic literature on the RAE, covering the social and quantitative aspects and implications of the exercise. It is also perhaps no surprise that much of this literature is critical and it has been widely noted how the RAE has changed the nature of research activity within universities. Indeed, academic research in the UK is now almost entirely constrained by the exercise; academics are forced to compete to win a share of a relatively small pool of funds with the majority of the spoils going to a

limited number of high performers as determined by the RAE grade awarded. Current Government thinking is towards making this division more trenchant. In future low scorers may be forced to abandon research altogether. Broadhead and Howard (1998) suggest that 'the RAE is … an exercise in disciplinary power'.

The RAE operates by peer review: *The panels use their professional judgement to form a view of the overall quality of the research in each submission within their unit of assessment, using all the evidence presented in the submission* (www.hero.ac.uk/rae). Since the RAE was instituted there has been pressure to be 'research active' inside a UoA; this has stimulated growth in the number of papers published and a consequent growth in the number of journals and conferences to accommodate them. This does not necessarily mean an increase in quality of work, however: there is a strong motivation to be repetitive and publish the same work in more than one outlet in an effort to appear more productive. But journal impact factors suggest that much published work has little or no impact on the community. Research has become more like a commodity, with publications as the base currency. Furthermore, the four-year frequency of the exercise discourages longer term projects or studies that may not lead immediately to publishable work. Risk is being expunged and this must have implications for the health of UK research.

Peer review is fundamental to the RAE's operation. Its integrity depends on the effectiveness of the reviewing, but peer review bias has been observed in some fields (McKay 2003). Bias has been noted in favour of prestigious institutions, obscurity and complexity of the work, the profile and even the sex of the author. Yet more troubling are suggestions that an institution's score may be enhanced by having a member on the UoA panel (Roberts 1999).

The RAE is undoubtedly having an effect on university research, at least part of which is potentially damaging. But there also appears to be some impact on teaching, not least the reduction in time available for student contact as the stress builds on academics trying to fulfil the demands of the RAE (Broadhead and Howard 1998). Teaching is being devalued because it tends to be regarded as of secondary importance, even though the majority of income in many institutions comes from student places. Government policy may be encouraging this, given the proposals to concentrate research in the 'best' universities, leaving teaching to the rest. Shaw's epigram may yet be realised.

Quantitative indicators of research quality

Over the years several indicators have been suggested as suitable measures of research quality: publication counts, possibly adjusted for size (numbers of academics in a department) or quality (e.g. journal prestige), citation counts, journal impact factors, etc. When I worked at the Science Policy Research Unit, Sussex, I contributed to a time-consuming and expensive project that looked in detail at the outputs and citation rates of all UK universities over ten years (Martin *et al* 1992, 1993). The values of these objective indicators were compared against the RAE rankings. Results were contradictory. Notably, size adjusted indicators (e.g. publications per person) were only weakly correlated with RAE scores. The strongest predictor was the size of the submission; thus the greater number of contributing staff in a UoA the better the score was likely to be; but there was no significant correlation with productivity. This

size relationship has held generally true since the RAE began, including the 2000 Computer Science UoA data (see below).

The question of size and productivity of an academic unit is important; the notion of economies of scale is implicit in the goals of the RAE and Government views on research policy. SPRU at the University of Sussex has published a body of research on this subject, most recently a review commissioned by the DTI (Tunnzelmann et al 2003). They found that despite some evidence that productivity improved within research teams as size increased up to a limit of 6 or 8 individuals, there is none to suggest that larger departments or universities are more efficient in this respect. The report concludes that there is little if any convincing evidence to justify further research concentration on grounds of economic efficiency; on the contrary, smaller units appear to be more cost efficient. Government policy to encourage research concentration in top ranked departments appears to be based on a single HEFCE report (Adams *et al* 2000) who argue that they are more productive. However, Tunnzelmann *et al* point out that these departments have received much higher funding and their productivity may actually be lower. This clearly is an important conclusion for institutions like ours.

The Computer Science UoA and our performance

In this section I present the RAE 2000 results for Computer Science as graphs of several quantitative indicators against scores. The indicators are adjusted for size where appropriate by dividing by the submission size as indicated by the number of full time staff. There is a clear relationship between some indicators and RAE score but the usual warnings about correlation and causation of course apply.

The history of a university has been found a good predictor of score in many UoAs. Figure 1 is a bar chart of counts against the respective scores of universities. The 'new' universities shown in light blue are emphatically towards the lower end of the rankings. This may be an indication of bias in the RAE process but the counter argument is that new universities have yet to develop a sufficiently high standard of research. Only one creeps into the 5 rating, Plymouth University. This is an anomaly of interest to us and I pay special attention to it below.

Figure 2. shows the relationship between size and RAE scores. Kingston University's position is indicated on the graph. The widely noted bias towards size is apparent. Score increases with mean unit size, although the spread is very wide. We are one of the smallest submissions but no smaller than the slimmest grade 5. The most obvious feature is the 5* units which are on average considerably larger than the lower ranked, although again the spread is large.

To see whether the level of activity within a unit is related to score, Figure 3. plots the ratio of submitted staff to the total. Both Kingston and Plymouth's points are labelled. Here we see a very broad spread of values but some indication of a relationship between the ratio and RAE ranks. A little surprising are the figures for some of the 5*; as few as half the total staff are submitted. This may indicate a selection process within the unit to try and maximise score. A few institutions submit a very low proportion of staff, as low as 10% in one case; these all fared poorly in the 2001 exercise. Plymouth, on the other hand, have a high submission ratio, more than double

Kingston's. Their complement of staff is lower, however, and all but one were submitted; that nearly all staff are considered research active may influence the score received. It does appear that a low proportion of submitted staff implies a low score in the RAE.

RAE submissions are judged on four criteria, principally 'the quality of publications and other public output'. In the Computer Science submissions, the majority of the outputs were comprised of journal and conference articles and chapters in books; other work such as patents, reports, software, etc. are rather fewer. Figure 3 shows no clear relationship between the proportion of journal articles submitted, although those with fewest received low RAE scores. Plymouth's figure is only slightly more than Kingston's, and 5* institutions do not stand out in this respect either. This suggests that type of publication has only minor bearing on scores, as one would expect if the quality of the work is being judged.

Postgraduate student activity is one of the criteria used by RAE panels, although of less importance than publications and evidence of research culture. The graph of PhD completions (adjusted for size) in Figure 5 is fairly flat but contains some interesting outliers, especially the value of 4 for a 3a rated university (Nottingham Trent) which may be an error in the data. Plymouth has a higher figure than Kingston, but many universities with a higher RAE score than us have no larger a completion rate. Again the 5* universities seem to be in a class of their own, not by virtue of a high PhD output but with a consistent figure of around 2 — i.e. two completions per person over the 4 year period. The reason for this consistency probably has much to do with steady income over many years. Figure 6, income per person, supports this view: even the lowest funded 5* institution is well above the majority of other universities; there is a virtuous loop between RAE scores and income. Among the lower scores the link with income is less obvious. Kingston's is particularly low but the data show that a low income is not automatically penalised with a low score. And clearly the policies of funding councils means there is a strong feedback mechanism at work.

If the peer review of the panels genuinely recognises quality we might expect to see this reflected in the esteem of the journals in which articles are published: higher graded universities should have more articles in the best journals. It is not possible to perform a full analysis here but Table 1 compares Kingston and Plymouth's submissions with one another, using the journal impact factor as an objective measure of quality. The impact factor of a journal is the total number of articles it publishes in a given period divided by the total number of citations received for those articles. For most journals the factor is quite small, often less than 1. Comparing Kingston and Plymouth, there is no great difference: the medians are similar and both have one article in a higher impact journal. The main difference is in the subject matter: Kingston has nearly half in biomedical applications with the rest in assorted imaging applications; while Plymouth's output is mainly in the field of neural networks and related areas. This may or may not have a bearing on a panel's deliberations, depending on what they believe constitutes 'computer science'; it may be pertinent that the most prevalent journal in submissions to the Computer Science UoA appears to be *Theoretical Computer Science*.

The last and most controversial indicator I will consider here is that of panel membership. The apparent correlation between this and RAE score has been

suggested before, including by our previous Head of School. Figure 7 is a reprise of Figure 2 but now indicating those universities who have a member on the panel. No less than four out of the six universities rated 5* have panel membership; three are from 5s, two from 4s and one straggler from a 3b. Most interesting of all is that Plymouth has a member on the panel, Mike Denham in fact, a previous Head of School at Kingston. What this tells us about panels and bias is a point to debate, but it is I think a matter of concern, especially to us, that so many panel members come from the high ranked traditional universities.

Discussion on the RAE and our strategy.

This study is necessarily limited to the data easily available. But it is able to show that as a new university we may suffer from at least some bias in the current system of the RAE. There are many questions on whether the RAE is beneficial to UK research and yet more serious questions about the conclusions the Government is attempting to draw from it. At present there is a suspicion that the RAE functions as a vehicle to justify the concentration of resources into a handful of institutions, inevitably at the expense of those in our position.

How well the RAE's peer review method works is not clear. Analysis of some indicators here and elsewhere suggests that there is an inherent bias in the rankings and universities starting from a low point could have great difficulty catching up. On the other hand, apparently indifferent performance in measures such as doctorate production and income does not necessarily imply a poor final RAE score. This is an encouraging sign that the panels, as the criteria state, really do assess quality of work without being too influenced by numerical factors.

In the introduction I stated that I wished to examine our performance in the last RAE. In 2001 we moved to a score of 3b from 2 in 1996. However, the grade scale was changed between the two exercises and we can only sensibly compare performances by examining the grade definitions. Below is an extract of the 2001 grading criteria from the HERO website:

> 4 Quality that equates to attainable levels of national excellence in virtually all of the research activity submitted, showing some evidence of international excellence.
>
> 3a Quality that equates to attainable levels of national excellence in over two-thirds of the research activity submitted, possibly showing evidence of international excellence.
>
> 3b Quality that equates to attainable levels of national excellence in more than half of the research activity submitted.

From this we can see our submission was judged to have attained national excellence in between half to two-thirds of the research. Comparing the above definition of 3b with that from 1996:

> 3b Research quality that equates to attainable levels of national excellence in the majority of sub-areas of activity. (Research quality that equates to attainable levels of national excellence in a majority of the sub-areas of activity, or to international level in some.)

This definition is closer to 4 in 2001 than 3b, showing that grade inflation has taken place. (This is likely to continue as an additional grade has been mooted for the next RAE.) Thus, despite the one grade point increase, the quality of our research has not improved appreciably as judged by the RAE. No doubt there are reasons to dispute the RAE panel's judgement but the fact remains that we are close to the bottom of the RAE pile and left with the implications.

Given this result a close review of our research strategy is now overdue. Following the 1996 exercise it was decided to direct most of the resources in the old CSES school to a single research group; this was expected to lead to the best improvement in the next RAE. The drawback to this decision was that other research in the School was starved of resources. Meanwhile, research in the old School of Information Systems was given no University resources.

Given our focus on the RAE, this strategy has clearly not worked well. How should we proceed from here? One of the main RAE criteria is 'evidence of research culture, strategy and vitality in the Department…'; by any measure this would seem to be a prerequisite for any successful research effort. This I believe is where we need to focus most of our attention. At present the School of CIS has a work allocation model based on one used in the old CSES. It contains an element for those staff engaged in research, or more specifically those considered to be 'RAE research active'. The method of allocating research time has not been discussed or agreed within the School but, like the rest of the work model, drawn from the old. Research time is awarded in the model according to the following criteria: authorship of four papers, receipt of research income, supervision of graduate students, management of a research project. There are a number of problems with this. The four papers refers to the maximum size of submission per person; having less than four papers with at least two years to go is irrelevant. This criterion also perpetuates the four-paper myth: the RAE allows for less than four outputs per person when '*engaged upon work which does not produce early outcomes but is likely to lead eventually to an outcome of high quality'*, and for other reasons. Moreover, it makes no allowance for quality. Nor is adjustment made for double counting: multiple authorship, multiple applicants for grants, etc. Research income, students and projects tend to go together; thus allowances for these three items are another form of double counting. The method encourages manipulation of figures.

Obtaining objective measures of research quality is difficult but it might be illuminating to examine one of the above indicators, publications per person, for our School. Figure 8 plots the number of articles in ISI journals since 2001 (eligible for the next RAE), adjusted for number of internal authors, plotted against research

allowance. I use ISI journals because the data are independently compiled and the list of journals is of a recognised standard. The graph does not include those with zero ISI publications. There is a cluster of points in the middle, with two individuals who appear to receive too little allowance relative to their output, and two who have a comparatively generous allowance. The correlation between papers and allowance is negative, but of course the data points are few and I place no great emphasis on the statistics. Other factors should also be considered, but given that quality research usually appears in good journals, this graph contains useful information and suggests the distribution of research allowance is not necessarily a reflection of productivity.

A fundamental requirement for a thriving research culture must be an equitable distribution of resources. Currently more research time is allocated to the members of one grouping (Figure 9). The reason is partly historical: this group received most of the resources previously. Given the lack of success in the RAE there is no justification for this situation to continue. Although there may be an argument for transferring the majority of resources from this group to another, this would be perpetuating the same problem and I would argue for a strategy that encourages everyone rather than create division. The distribution of individual research time in Figure 10 shows a very inequitable spread of resources, even excluding those staff members with zero allowance, some of whom are very active researchers. We cannot expect to encourage more research activity if many staff are denied a proper share of resources. Apart from nourishing a research culture, so important if we are to improve the long term quality of our work, it will provide the opportunity to submit more staff to future RAE exercises, which as I have noted above, may be one of the simplest ways of improving our score.

It is of course appropriate to consider whether we should play the RAE game at all, and to some extent we treat it in a tactical fashion by debating issues like adding more names to papers. Ultimately we hope the reward for persevering with the RAE will be more central funding, but if future money goes to the two highest grades we have a steep hill to ascend, one where the summit may continually move upward. The mechanical approach, boosting indicators such as postgraduate student numbers, for example, will continue to be difficult in a climate in which even universities with studentships have trouble filling them.

Whether to submit to the RAE is probably a decision out of our hands. Whatever happens, our present approach is undermining the health and breadth of research in the School. Regrettably we seem to have developed something of a punishment culture, even down to one group's website highlighting in red its members who do not yet have the requisite number of papers. Good research is about generating new knowledge and understanding; this requires adequate resources and a supportive environment. Research worth doing also carries the risk of failure and it is vital that this is recognised and not penalised.

If we do remain with the RAE we should consider submitting to other UoAs. Information Systems research does not necessarily fit naturally into the Computer Science Unit, nor indeed do some other areas of research in the School. It may prove more efficient, for instance, to submit the biomedical research to Other Studies and Professions Allied to Medicine, a UoA for which Kingston already has a grade 3a. Similarly some of the more mathematical research could be entered in one of the

mathematics UoAs. There are no doubt other suitable strategies we could adopt but the fundamental aim must be to provide the opportunity for everyone to fulfil their research potential. And if we choose to persist with a game strategy we could do worse than put our Head of School up for RAE panel membership.

References

Broadhead L, Howard S. (1998) "The art of punishing": The Research Assessment Exercise and the ritualisation of power in higher education. *Education Policy Analysis Archives* 6 (8)

Martin BM, Skea JEF, Ling EN (1992) Performance indicators for academic scientific research. End of award report to the Advisory Board for the Research Councils and to the Economic and Social Research Council for project number A418254009.

Martin BR, Hicks D, Ling EN, Skea JEF (1993) The Effects of Size and Other Factors on the Research Performance of University Departments. Report to the Australian Centre for Innovation and International Competitiveness

McKay S. (2003) Quantifying quality: can quantitative data ("metrics") explain the 2001 RAE ratings for Social Policy and Administration? *Social Policy and Administration*. 37(5) 444-467

Roberts C. (1999) Possible bias due to panel membership in the 1996 Research Assessment Exercise. Manchester University: School of Epidemiological and Health Sciences. www.medicine.man.ac.uk/epidem/biostats/raepanel.rtf

Tunzelmann N, Ranga M, Martin B, Guena A. (2003) The effects of size on research performance: a SPRU review. Report to the Office of Science and Technology, Department of Trade and Industry.

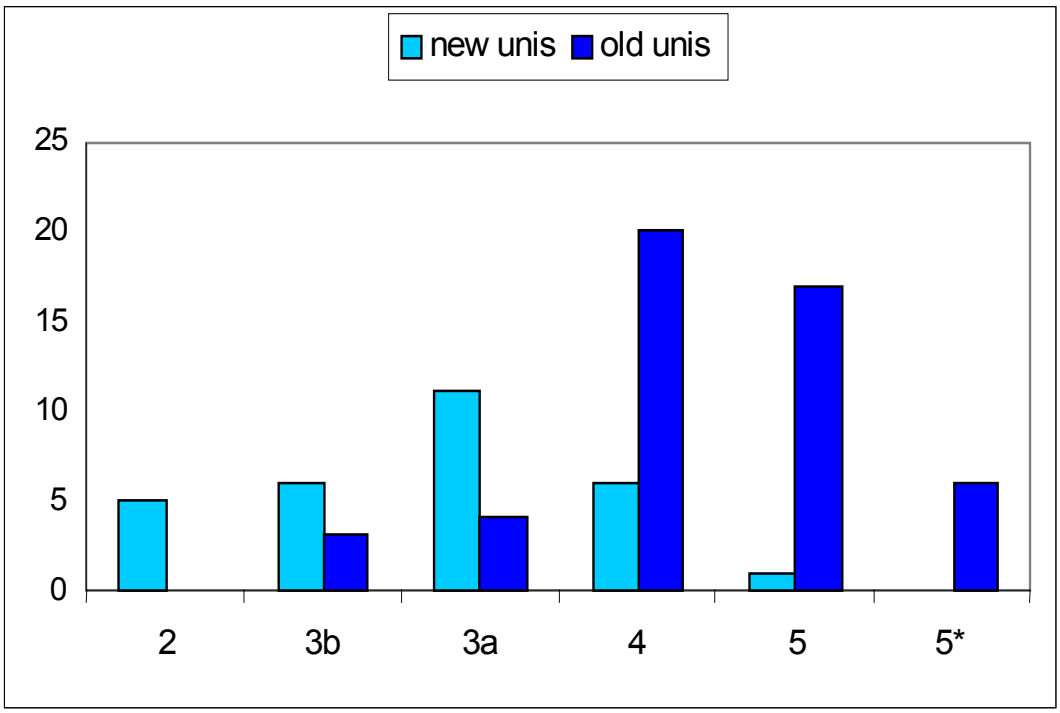

Figure 1. Rankings for Computer Science UoA 2001

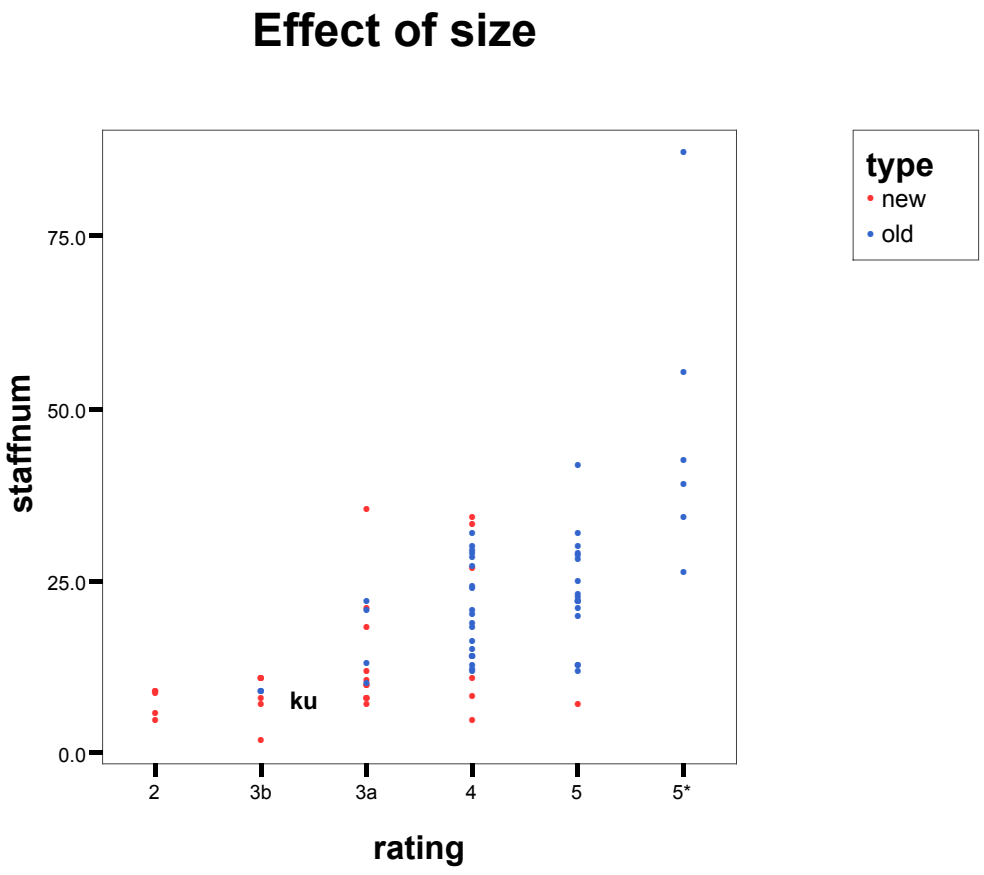

Figure 2. Number of full time staff versus RAE score. New universities in red.

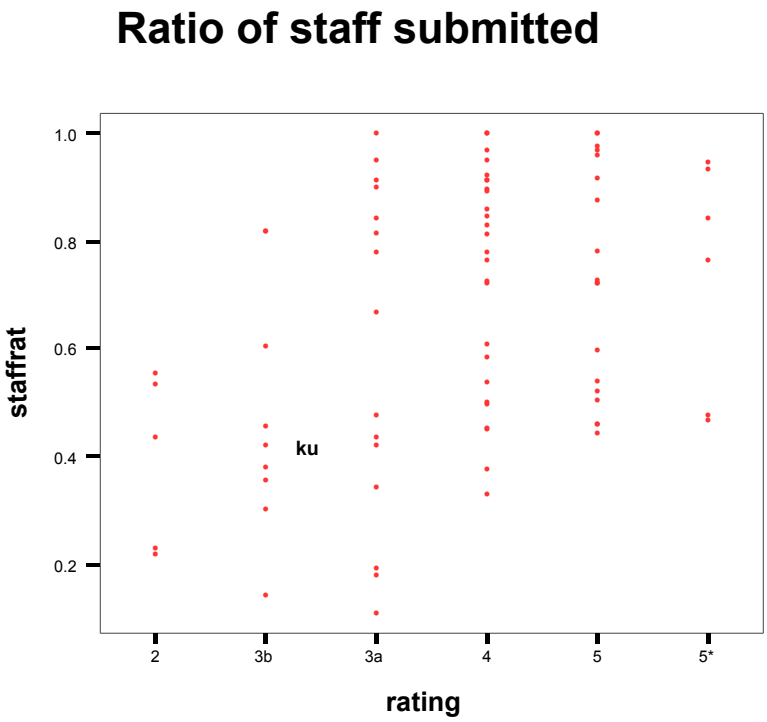

Figure 3. Ratio of full time staff in submission versus RAE score.

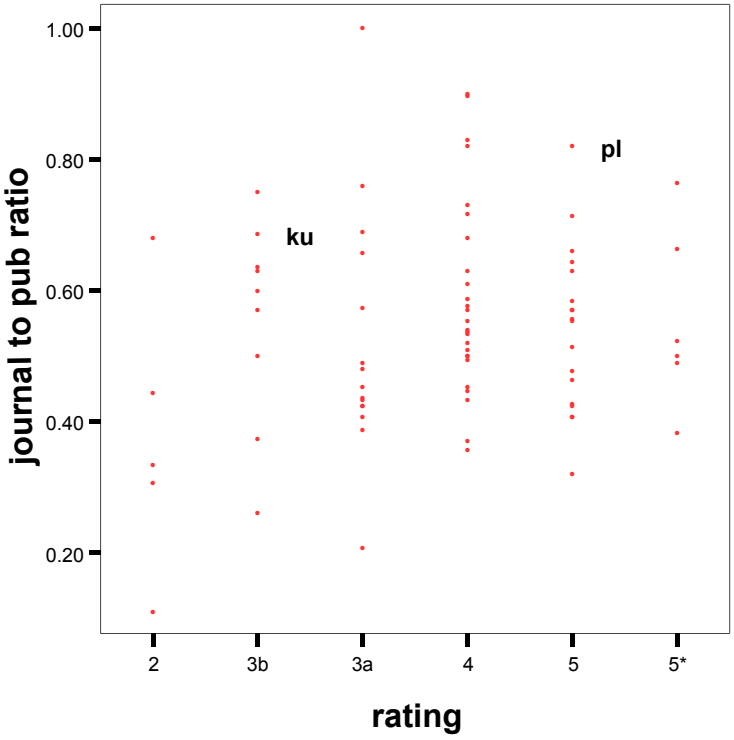

Figure 4. Proportion of articles in journals to total submitted outputs.

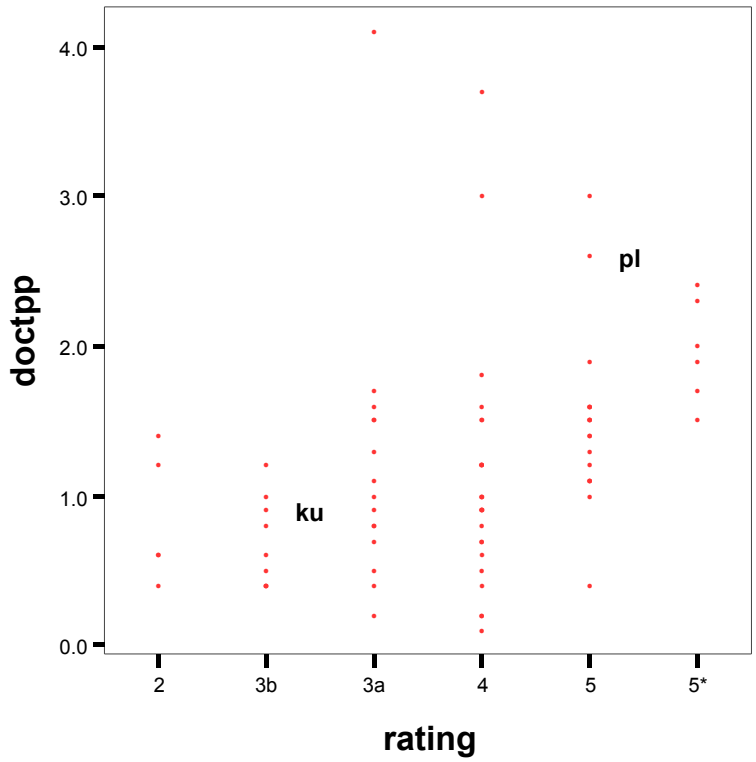

Figure 5. Number of PhD completions per staff member.

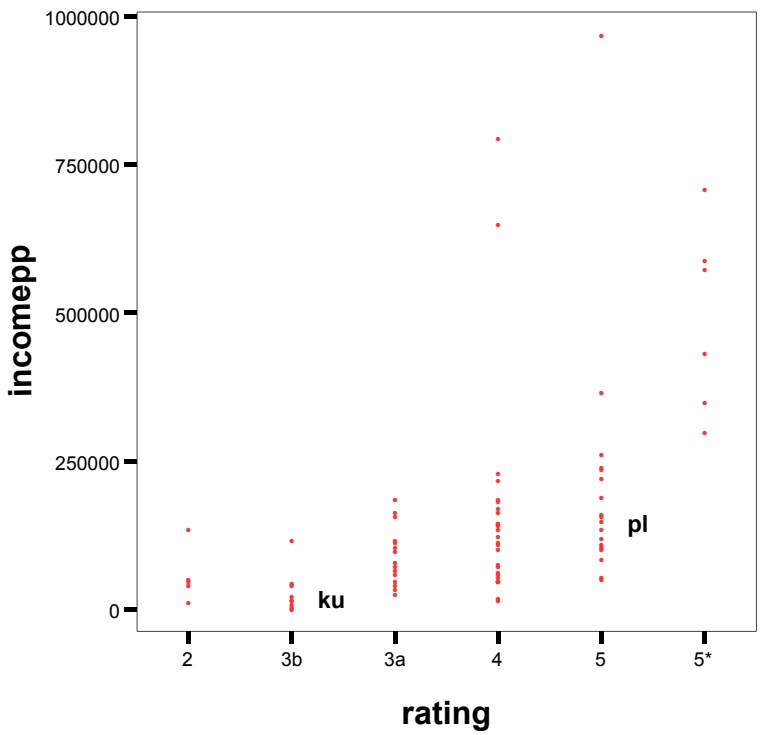

Figure 6. Total income per staff member.

Plymouth	Impact	Kingston	Impact
Biological Cybernetics	1.473	American Journal of Ophthalmology	1.828
Biosystems	0.736	Investigative Ophthalmology and Visual Science	4.172
Cognitive Systems		Applied Optics	1.459
Connection Science	0.964	Ophthalmology	3.066
Design Studies		Behaviour and Information Technology	0.603
Hippocampus	4.333	Journal of Electronic Imaging	0.723
Journal of the Acoustical Society of America	1.44	Anticancer Research	1.416
Nature Neuroscience	15.668	Medical and Biological Engineering and Computing	1.172
Neural Computation	2.727	Computer Vision and Image Understanding	1.298
Neural Networks	1.431	Geographical and Environmental Modelling	
Neural Processing Letters	0.379	Electronics Letters	0.97
Neurocomputing	0.534	Image and Vision Computing	0.893
Optical Society of America		Skin Research and Technology	
Perception & Psychophysics	1.492	The Lancet	13.251
Personal Technologies		European Journal of Thermology	
Reviews in the Neurosciences	3.794	International Journal of Remote Sensing	0.827
		International Journal of Geographical Information Systems	0.905
Sum	34.971		32.583
Median	1.4565		1.235

Table 1. List of journal titles in 2001 submissions for Plymouth and Kingston and the corresponding journal impact factors.

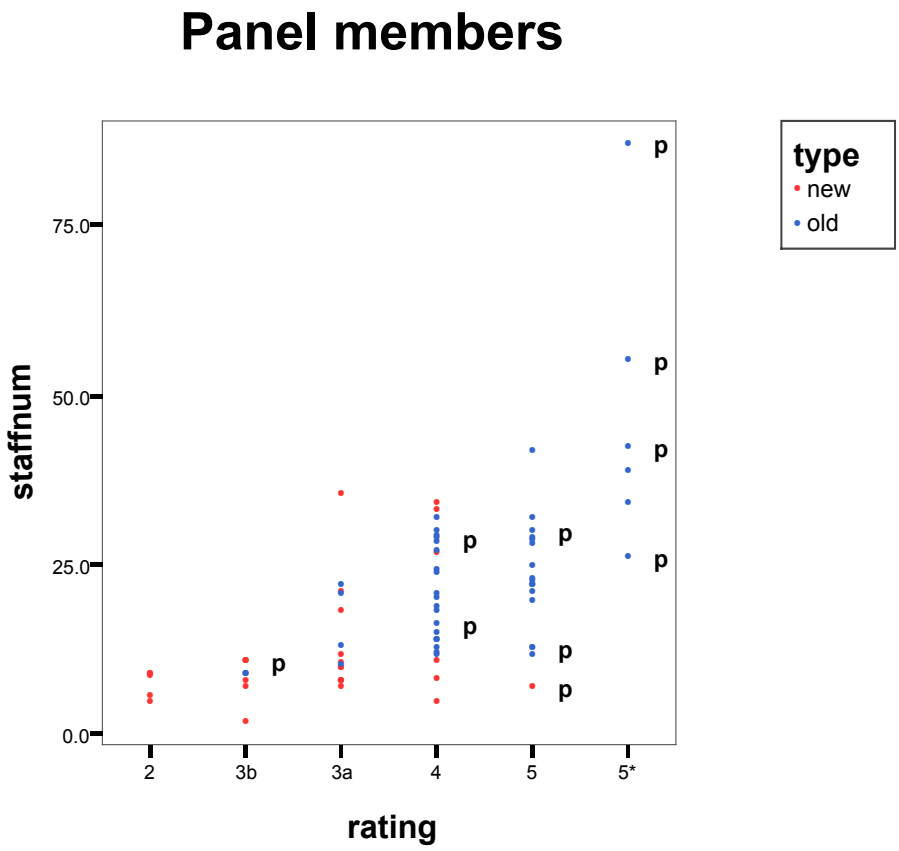

Figure 7. Plot of size indicating universities with a member on the Computer Science panel.

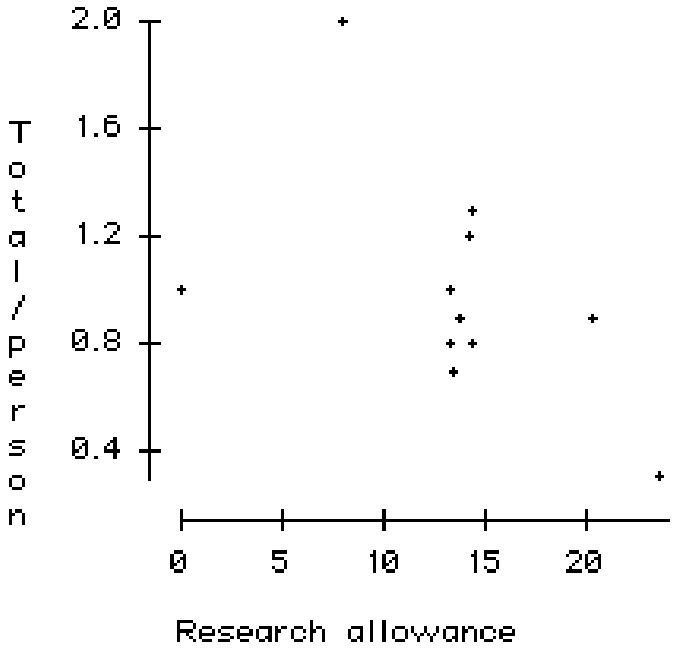

Figure 8. Total articles since 2001 in ISI journals per person (internal author) versus research allowance in 'work units'.

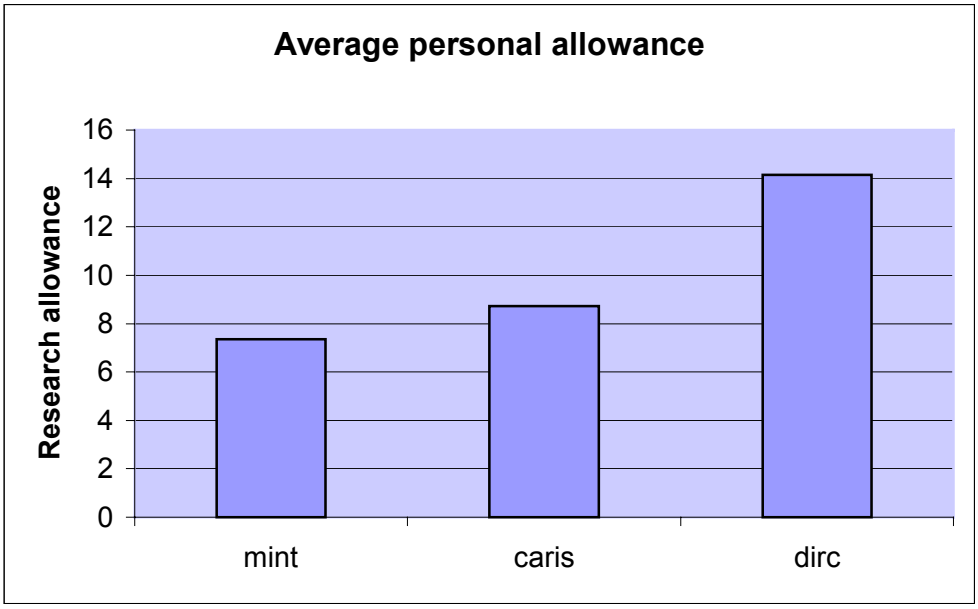

Figure 9. Average personal research allowance among the research groupings *mint, caris* and *dirc.*

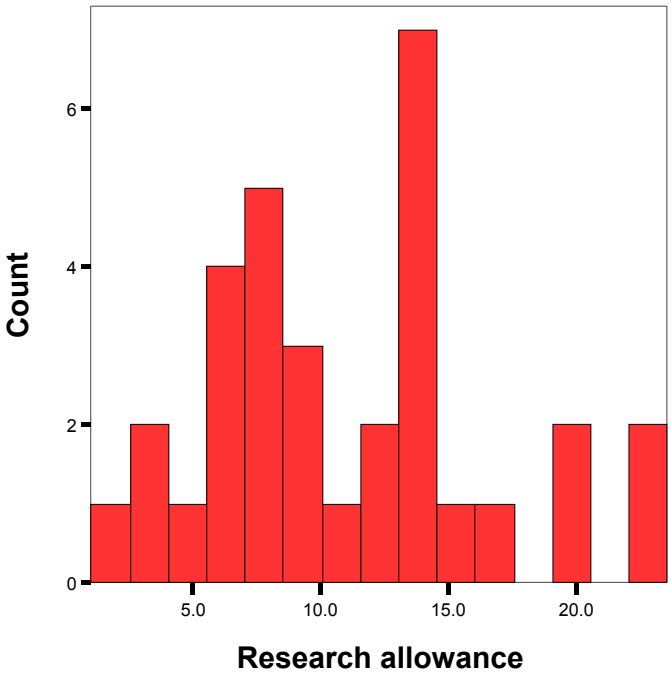

Figure 10. The School's distribution of individual research allowances in units.