Research Policy

Selectivity by the UGC: a taste of things to come

RICHARD SMITH

In 1986 for the first time the grants that universities received from the University Grants Committee were related to ratings of the research effort of their various departments. Such selectivity was effectively forced on the UGC by the cuts in its budget. The mechanism used to grade the departments was only half explained, and the results were never published collectively. Yet the results of the gradings were collected together haphazardly by some groups and publications, and many departments thought that they had been misjudged. Many academics objected particularly to the fact that it was impossible to validate the exercise because the methods used were inadequately described.

Smith, a geographer, has written:

An exercise which I occasionally inflict on my students is to classify a set of places on the basis of "quality of life." The test is to choose defensible criteria, find sensible indicators, and amalgamate them so as to assign places to such categories as "outstanding," "above average," "average," and "below average." Students who do not reveal their data and method fail, as I suspect them of skimping the exercise or of producing judgments which they are unable or unwilling to justify. On 27 May 1986, in a similar exercise, the UGC failed.¹

The proposals to create a three tier university system² may make the UGC's selection exercise both irrelevant and more relevant. Its results may be supplanted, but the methods used to grade departments will become of greater importance because similar methods will have to be used to grade the universities. The policy of putting research eggs into fewer baskets will be disastrous if the wrong baskets are chosen.

UGC little room for manoeuvre

Despite the clamour that surrounded the UGC's selection exercise, it had only limited room to change the grant that it gives to a university.² Over 60% of the grant is for teaching, and although there are plans to be selective in awarding this part of the grant this has not yet happened: measuring teaching quality is even more fraught than measuring research quality. A further 15% or so is related to student numbers (as a proxy for staff numbers and research student numbers) and about 6% is for overheads on grants won from research councils or charities or for contract work, which leaves only 14% with which the UGC can be selective (as another 4% is for "special factors").

The universities still receive their money from the UGC as a block grant (although this too may eventually change³), and they are not bound to give more money to departments rated highly. It would obviously be difficult, however, not to do so as the departments know how they have been graded. Vice chancellors are fond of what they call "top slicing," a process whereby they keep some of the block grant aside to spend as they will—perhaps to bring on an

advancing department or to bring back from the brink one that is fading. A vice chancellor might also choose to hasten the end of a badly graded department.

Mechanisms of evaluation and immediate reaction

Although the UGC did not define exactly how it determined its selections and did not conduct the process publicly, we do know roughly what it did—although the mechanisms probably differed from subject to subject.⁴ In 1985 the UGC wrote to all universities asking them to produce a 500-600 word profile of "cost centres" together with the five best research publications of the centre. A "cost centre" was larger than a department. The UGC's committees for the various specialties, each with about 12-15 members, then divided into subcommittees that looked at the profiles and the publications. The subcommittees also counted the number of grants and research studentships gained from the research councils over the past five years; the amount of support from the large medical charities; the numbers of fellows of the Royal Society in the centres; the amount of money gained for "new blood" posts over the past three years; and the amount of money gained from industry. In addition, research councils gave their private opinions, although some were sceptical about the whole exercise. The subcommittees (some of which reportedly contained as few as two members⁵) put all this information together and graded cost centres and sometimes departments from one to five. The final results were given out with departments graded average, below average, or above average, with some starred to show international excellence.

Many departments thought that they had been misjudged and that an establishment cabal had handed out predictable judgments—people were quick to notice that the south generally had done well while the Celtic fringes had done badly.⁴ (Professor J Lamb from St Andrews University has shown, using an expert system, that for his subject—physiology—the UGC rankings were significantly influenced by the distance of the university from Oxbridge and London. This has not been tested for other subjects (J Lamb, personal communication).) Other critics were concerned that smaller departments had done badly because of being lumped into the larger cost centres, and academics wrote to *The Times* saying that the assessment was "wholly lacking in intellectual credibility" and its methods "questionable," "crude," and "seriously defective."⁶ One of the most serious criticisms was that this doubtful cross sectional snapshot of quality would condemn those departments that were poorly graded; they had little chance to come back as there was no appeal mechanism and the assessment was not to be repeated for four or five years. The judgments may thus become self fulfilling prophecies.

Evaluating the evaluations

Since the gradings were made some more academic studies of the selection exercise have been published. Lamb argued that "there should surely have been a formal model for assessment, with

the various factors stated explicitly and weighted properly," and then conducted his own analysis of the quality of physiology departments.⁶ He used four measures: total research grant income, total grant income from the research councils, total number of PhD students, and the numbers of PhD students allocated to departments under the research council's quota system. He then graded the departments into four groups and compared his grading with the UGC grading. He found that the two were significantly correlated and that only about 40% of the UGC grading was accounted for by his measures. He also found that 16% of the UGC grading was accounted for by the size of the department—the bigger the department the higher the grade.

This is the Matthew effect—to those that have shall be given while to those that have not shall be taken away. The same effect was found in an analysis of the grading of geography departments,⁷ while a study that tried to relate the UGC ratings of geography departments with measures of publications, citations, and research grants found only weak correlations and several discrepancies.⁸ Peacock, a physicist, looked too at the relation between staff and publications in university physics departments and poured scorn on the "critical mass" theory beloved of research managers: some small departments produced vigorous "chain reactions" (in terms of publications), whereas some departments that should have "gone critical" produced only sluggish chain reactions.⁹

Evaluation in psychology

One group of academics who know about evaluation are psychologists, and one psychologist with an interest in evaluation, Raphael Gillett, applied his skills to looking at the UGC's ratings when his own department was poorly rated.^{5 10 11} In his comprehensive study of the UGC grading of psychology departments he found that between one quarter and one third of the UGC grading was accounted for by departmental size.⁵ And this, he argued, was wholly unsurprising because the method of asking cost centres to select their five best publications would bias the results towards the larger departments: even if the mean quality of two department's publications are the same the best five of the larger department will inevitably be better than the best five of the smaller department. This, he argued, is why the UGC should have looked at total research output (publications) over the chosen time period. (At a recent meeting at the CIBA Foundation on the UGC rankings Professor John Sizer, representing the committee, said that the five publications were used mainly to identify cost centres that were so weak they could not produce even five publications.¹²)

Gillett makes many other criticisms of the UGC method in a closely argued paper, including the facts that the UGC did not tell the cost centres how they would judge the five papers; five outstanding publications might come from "a solitary genius . . . in a department of dullards"; most of the measures used by the UGC have never been validated in a controlled independent study; number of professional distinctions is a measure of past rather than present performance; using the number of research grants ignores the quality and quantity of research they generate and the fact that they artificially inflate the publication rate of departments, and again bias the results against smaller departments; and, finally, psychological evidence shows that groups—such as the research councils—who must choose a few people from among many strong candidates to receive a grant (which gives the recipient a great advantage) greatly overestimate the validity of their own judgments.^{5 13}

After making these criticisms Gillett then conducted his own assessments of psychology departments and related them to the UGC gradings. His conclusion was that "the UGC ratings have approximately zero validity." He reached this conclusion by counting the number of publications in learned journals and books in 1984-5 from 37 of the 40 psychology departments in Britain and dividing the total by the staff of the department. He also assessed the impact of the articles by awarding each article the citation impact of the journal in which it was published and again dividing this by the number of staff in the department.

Gillett defines the citation impact as "the number of citations of articles from a particular journal to be found in the scientific literature in a given period, divided by the number of citations of articles published by that journal in the same period."⁵ This is a move of doubtful validity because the number of times that individual articles within journals are cited varies enormously—for instance, up to half of the articles published in some disciplines are never ever cited, while others are cited dozens of times over a few years. This might not matter if Gillett was dealing with hundreds of papers from each department, but he does not give overall numbers in his paper. He gives only averages, which is a fault. (He told me that the size of psychology departments varies from six to 29 if only lecturers are counted and from six to 55 if full time research workers are included. As the average number of publications for each lecturer and full time research worker ranges from 0.7 to 3.0 annually this means that some numbers of publications might be small.)

Gillett then related his scores to the UGC classification, and the results are shown in the table. On both of these measures

Average productivity (number of publications divided by number of lecturers and researchers) and average citation impact (aggregate citation impact divided by number of lecturers and researchers) for British psychology departments for 1984-5 related to UGC grading of the departments. (Reproduced with permission of R Gillett and the Bulletin of the British Psychological Society⁵)

UGC grading	Productivity	Mean productivity	Citation impact	Mean citation impact
Outstanding	2.2, 0.6, 1.0, 1.7, 2.0, 1.1, 1.5, 1.7	1.5	0.48, 0.33, 0.17, 1.84 1.52, 0.51, 2.08, 0.03	0.87
Above average	1.8, 1.1, 2.5	1.8	0.61, 0.36, 1.39	0.79
Average	2.4, 1.0, 1.8, 2.9, 1.2, 3.0, 0.9, 1.6, 1.0, 1.5, 1.5, 0.9, 2.0, 1.9, 0.8	1.6	2.44, 0.58, 2.00, 1.64, 0.55, 0.80, 0.28, 0.23 0.27, 0.90, 0.66, 0.12, 1.71, 0.65, 0.17	0.87
Below average	1.6, 1.8, 1.8, 1.0, 2.3, 2.6, 0.7, 0.9, 1.3, 1.1, 1.3	1.5	0.18, 2.36, 0.85, 0.91, 1.11, 0.97, 0.35, 0.11 0.73, 0.65, 0.62	0.80

of "quantity" and "quality" the means of the four groups of departments are very similar, but many departments graded as "below average" are doing better than those graded "outstanding," and on each measure the department scoring the lowest was rated outstanding by the UGC (and in each case it was a different department). For both measures Gillett then proceeded to correlate his score with that of the UGC and found correlations of 0.01 and 0.03—that is, no correlation. Next he correlated UGC ratings with departmental staffing (once for lecturing staff and once for lecturing and research staff) and found correlations of $r=0.48$ and $r=0.60$, which led him to conclude that between one quarter and one third of the variance in UGC ratings was accounted for by the size of the departments.

He also correlated research productivity by his two measures with his two measures of staffing and found that r varied between -0.09 and 0.11 . Thus he concluded "the currently fashionable view that 'big departments are best' is not supported." This I find an interesting sideline to his paper because among the medical researchers I spoke to it is taken as almost axiomatic that big departments are best, and the research councils are certainly being pushed along that line.

Philip Levy, one of the UGC's two psychology advisers, has criticised Gillett's paper by suggesting that "he buries key issues in his limited discussion of the validity of two simplistic, yet somehow overelaborated, indicators."¹⁴ In his diffuse paper Levy primarily attacks the whole notion of using factors like numbers of publications and citations: he prefers "a wide range of peer judgments based upon a wide range of evidence about 'research quality' taken from a broader range of years," and talks mockingly about "judging the relative beauty of cathedrals by dividing their size by the number of labourers." Levy is thus defending the status quo, which sounds very like a mechanism of established academics making

decisions over claret behind closed doors and refusing to justify themselves. The more "objective" measures do have defects but at least they can be seen, validated, and debated.

Levy also raises more technical questions about Gillett's paper—on correcting the outputs of departments by their size and on using data from only one year—but Gillett dismisses these quickly in his reply.¹⁵ Gillett's study does, however, have its problems (as he admits) and his central point is not that his assessments are better than the UGC's but rather that they are measures that are open: "The measures and indices to be employed need to be discussed, and their underlying assumptions need to be articulated, debated, and where possible, tested."

The UGC's main arguments against this are likely to be, firstly, that the judgments are made not by outsiders but by academics active in the subjects they are assessing, and, secondly, that no system can be perfect and that it is better to get on and do some productive research rather than argue over the tedious process of allocating funds. It also argues that only small changes in funding have been produced by its selection exercise. These arguments are not satisfactory because they overestimate the cohesiveness of the academic community (which is in fact notorious for its dissensions) and underestimates the importance of what the UGC is doing. Although it is true that only small changes have so far been made in funding, changes in morale and reputation have been much more severe; and the same sorts of methods may soon be used to effect much bigger changes in funding.

Biomedical views on the validity of the UGC's selection exercise

Perhaps because they do not know how or because they do not have the time, medical academics have not disputed the UGC's judgments as have the psychologists and geographers. (One dental academic has, however, done so: Lloyd has made various counts of publications from dental schools and then compared these measures of research performance with the UGC grading.¹⁶ "Differences," he concludes, "are numerous and in some places glaring.") Almost all of the medical researchers whom I spoke to thought that the UGC had got it about right with its selection exercise; and this included the dean of a medical school whose preclinical departments had received very poor ratings. Many were concerned, however, that the UGC selection might condemn departments and make it very difficult for them to "rise from the dead." Several mentioned the department of biochemistry in Dundee, which since the arrival of a new professor has risen rapidly to international eminence. Departments do tend to rise and fall in quality (as indeed do journals and most institutions), and if after the UGC's exercise all the escalators go down and none up then that will eventually prove very serious for British research.

Response of the UGC

Although it is to be replaced by the University Funding Council in April 1989, the UGC is preparing for its next selection exercise. Professor John Sizer of the UGC told the recent meeting at the CIBA Foundation that the committee is to consult widely on reactions to its first round of selection before beginning the second round.¹² The UGC accepts that the assessments can be improved but points out that there have been no detailed criticisms of its judgments in most disciplines. (It would be rash of the UGC to consider therefore that its judgments were right: as far as I can see, in every case where there has been detailed examination of the UGC's rankings they have been found badly wanting. Perhaps somebody should look in detail at the medical assessments.)

Professor Sizer told the CIBA meeting that no matter how many data are available for the next assessment the decisions will still be a matter of judgment. The Science Policy Support Group, which organised the meeting, was keen that better bibliometric assessments should be made before the next round, but the UGC does not have the money to pay for them.

Selectivity within universities

The UGC is anxious that selection between universities should be complemented by selection within universities, which it made clear in its 1984 document, *A Strategy for Higher Education Into the 1990s*.¹⁷ Existing administrative mechanisms are thought to be inadequate, and new research committees are advocated. In addition, the UGC wants each university to know how research funds it receives are being spent, which is obviously a prerequisite for proper management. Universities are also going to be required to produce academic plans, and the UGC has waved a stick at the universities on this by saying that requests for special funding will not be forthcoming unless plans are produced.

The recent report from the Advisory Board for the Research Councils and the UGC describes how Glasgow University has responded to these calls from the UGC for selectivity on research within the university.² It has set up a research committee that has the function of advising "the university on the selective distribution of that part of its resources devoted to research." The committee has started by preparing an inventory of all research going on within the university—both internally and externally funded—and all the resources devoted to the projects. It has also gathered simple output measures (including publications and patents) and has asked each head of department to prepare a "matrix" showing how much time staff spend on research. "Performance indicators" are also being studied, but "the only qualitative information is the UGC rating of research strength."

This sort of exercise is going to have to begin in all British universities, although several universities are unhappy that the universities' double functions of teaching and research should be separated: some told the Advisory Board for the Research Councils-UGC working party that "to separate considerations of research from those of teaching is to misunderstand the intimate relationship between those activities in universities." But universities that resist the trend to management and selection may be the first to go to the wall. Again, however, these selection exercises are being done behind closed doors, and much is to be lost by backing the wrong horses.

Planning jointly with the research councils

One of the worries of increasing selectivity between and within universities is that the universities and the research councils may support different groups²—despite the research councils being consulted when the UGC makes its ratings and the funding from research councils being one of the factors that determines the UGC rating. The Advisory Board for the Research Councils is uncertain what to do about this.² It seems to want to tighten arrangements between research councils and universities, but it is also worried that "no grant giving agency can be sure that it will consistently and reliably 'pick the winners,' and there is the risk of leaving good projects unfunded."

Conclusion: the heart of the problem

In its statement above the advisory board seems to me to have identified the heart of the problem—and then, indeed, to have backed off from it. For wanting to manage research much more to ensure that it produces more "value for money" and more that will be of public benefit is all very well, but can it actually be done? The techniques of measuring the quality of research are crude and unvalidated, and the techniques for predicting which research projects will prove "profitable" are even weaker. Instead we have committees of establishment figures "backing their hunches," and those hunches may well be wrong. Britain's flagging research effort would then be stifled still further rather than resuscitated.

References

- 1 Smith DM. UGC research ratings: pass or fail? *Area* 1986;18:247-50.
- 2 Advisory Board for the Research Councils. *A strategy for the science base*. London: ABRC, 1987.

- 3 Secretary of State for Education and Science, Secretary of State for Wales, Secretary of State for Northern Ireland, and Secretary of State for Scotland. *Higher education: meeting the challenge*. London: HMSO, 1987. (Cmnd 114.)
- 4 Anderson A. Research gradings stir emotions. *Nature* 1986;322:299.
- 5 Gillett R. Serious anomalies in the UGC comparative evaluation of the research performance of psychology departments. *Bulletin of the British Psychological Society* 1987;40:42-9.
- 6 Lamb JF. Grading of universities. *Nature* 1986;323:576.
- 7 Gleave MB, Harrison C, Moss RP. UGC research ratings: the bigger the better? *Area* 1987;19:163-6.
- 8 Benthall G. An evaluation of the UGC's ratings of the research of British university geography departments. *Area* 1987;19:147-59.
- 9 Peacock G. Critical mess. *New Scientist* 1987 August 13:67.
- 10 Gillett R. No way to assess research. *New Scientist* 1987 July 30:59-60.
- 11 Gillett R, Aitkenhead M. Rank injustice in academic research. *Nature* 1987;327:381-2.
- 12 Turney J. Rethink on research ranks. *Times Higher Education Supplement* 1987 December 4:1,3.
- 13 Einhorn HJ, Hogarth RM. Confidence on judgement: persistence of the illusion of validity. *Psychological Review* 1978;85:395-416.
- 14 Levy P. On indices of research quality: a response to Raphael Gillett. *Bulletin of the British Psychological Society* 1987;40:211-3.
- 15 Gillett R. The UGC evaluation: a rejoinder to Philip Levy. *Bulletin of the British Psychological Society* (in press).
- 16 Lloyd CH. The research productivity of UK dental schools in the years 1980-1985. *Medical Science Research* 1987;15:349-53.
- 17 University Grants Committee. *A strategy for higher education into the 1990s*. London: UGC, 1984.

Conference Report

Fiddling while tobacco burns: sixth world conference on smoking and health, Tokyo, 9-12 November 1987

SIMON CHAPMAN

Every four years since 1968 sections of the world's community interested in the control of smoking have gathered for a world conference. This year it was the turn of Japan to be the host country. Previous host countries (the United States, the United Kingdom, Sweden, and Canada) all had important strategies and tactics for controlling smoking or important reports to tell the world. The Swedish and Canadian conferences in particular saw plenary speakers throw gauntlet after gauntlet down to the tobacco industry, calling for political action and using terms like "war" and "enemy."

Japan was quite a different story. With a staggering two thirds of Japanese men smoking, a government with a monopoly on tobacco and a growing enthusiasm for export to the vast markets of South East Asia, and the recent start of vigorous marketing targeted at Japanese women, who have a low prevalence of smoking, it seemed clear that the local organising committee was looking to the conference as a means of strengthening the hitherto shaky hand of its domestic health agencies.

Rapidly rising lung cancer in Japan

The future will be the judge of the wisdom of that intention, and the outstanding opening plenary lecture by Japan's Professor Takeshi Hirayama augured well. Professor Hirayama spoke on the effects of smoking on health with information drawn from a massive Japanese data bank (the second largest study ever on the subject). For years the international tobacco industry has publicly gloated over the apparent anomaly in the high prevalence of smoking in men in Japan and the comparatively low prevalence of lung cancer. In a health educator's nightmare of slides crammed with graphs Professor Hirayama showed that, as with most matters of science, the tobacco industry is now drastically ill informed on that score. The recent incidence of lung cancer in Japanese men is accelerating at a faster rate than that in any other country for which data are available.

Asked from the floor to explain this increase, Professor Hirayama speculated about the cohort lag effect arising from the large increase

in smoking during and after the Korean war and the possible role of the vast number of additives increasingly being used by the industry in efforts to modify tar and nicotine yields, combustion temperature, and flavour. Professor Hirayama was the first to present evidence on the role of passive smoking on lung cancer in non-smokers, his paper having proved seminal in the recent history of control of smoking.¹ At the conclusion of his address he received an adulatory vote of thanks from Professor Stanton Glantz, who named him as the person who had made the most important recent contribution to the control of smoking. Despite huge applause Professor Ernst Wynder, chairing the session, asked that further questions be about "scientific content" and not "politics."

Conference organisers are often faced with conflicting agenda over the need to break even financially and the desire to make some tangible progress. Financial considerations often dictate that papers of little merit or importance are accepted because of the need to swell the attendance. In Tokyo some 199 papers were crammed into two days, with only a fraction having more time than the 12 minutes (discussion included) allowed for presentations in special and free paper sessions. The result was both chaotic and superficial. The principal victims were certainly any development or obvious strengthening of global strategy and the abject neglect, beyond the now mandatory lip service, of any important focus on tobacco and the Third World.

Exploiting the Third World

Only 22 of the papers presented were delivered by delegates from the Third World, and of these, eight were small, local survey reports and a further two fundamental research reinventing the wheel and of questionable relevance in countries with a virtual absence of even basic measures of control. By the World Bank's classification the projected combined populations by the year 2000 of low income countries (for example, China, India, and Bangladesh), lower middle income countries (for example, Egypt, the Philippines, and Nigeria), and upper middle income countries (for example, Malaysia, Brazil, and Greece) will constitute 80% of the world's population (table).² Furthermore, the percentage aggregate increase in the populations of these countries from 1982 figures will be 41% compared with a growth in the population of industrial market and east European economies of only 8%.

The head of the World Health Organisation's Tobacco or Health