European Economic Review 35 (1991) 612–618. North-Holland

The use of assessments of British university teaching, and especially research, for the allocation of resources

A personal view

John Flemming

Bank of England, London, UK

The paper is in three parts: *1.* Background; demography of student numbers and age distribution of faculty in the 80s and 90s; the dual support system for academic research by the University Grants (Funding) Committee (Council) and Research Councils, the funding formula; Economic and Social Research Council sanctions on departments too few of whose doctoral students submit on time. *2.* UGC/UFC subject reviews and research rating exercises. Criticisms of correlation of assessed quality with departmental scale; difficulties of mixing input and output indices; bibliometrics and the apparent weighting of prestige journals; the role of consultation with learned societies such as the Royal Economic Society. *3.* Incentive issues and their transmission to individuals. The balance between pure and applied research.

I am not an academic student of this subject, but, as will become apparent in the course of this presentation I am a critical participant in the process in several different capacities.

After rising for several years the size of the 18 year old age group in the U.K. peaked in the mid-eighties and is expected to fall about 15% to the mid 90s before recovering about half the fall. The socio-economic composition of the decline, and economic and others factors affecting staying on at school, mean that at least since 1986 the Government has projected no fall in U.K. student numbers and its successive projections have been raised upwards substantially.

In the early 80s when numbers were expected to fall and when it was thought that faculties were clogged up with dead-wood recruited in the rapid expansion of the early 60s, a scheme was introduced to encourage early retirement. This encountered the usual problems. Typically the wrong people left and many were actually re-recruited at a lower cost to their institutions. In the hope of avoiding such problems were this situation to recur tenure arrangements are subject to new legislation.

Since 1980 the modal age of full time non-clerical university staff has risen

0014-2921/91/$03.50

by about 10 years as one might expect given the much reduced growth of both students and particularly staff numbers. The early retirement scheme kept the numbers reaching age 60 in any year constant at about 1,500, although the modal density is four times that size.

With reduced hiring, particularly of the young, and a generalised squeeze on academic funding, U.K. universities were not very happy for much of the 80s, and emigration, which has been endemic since the war (when British scientists went to the U.S. to work on the atomic bomb), again became a political issue – under the revived title of the 'brain drain'.

I am not going to resolve this issue (or any other). If a relatively prestigious and well paid occupation like university teaching slips down the occupational hierarchy what symptoms should we expect to see other than complaints by incumbents? With students to be taught we are unlikely to see gross vacancies, though one might see temporary stop-gap appointments if the adverse developments were thought to be temporary – which they were not. Thus jobs will be filled and the question is how the background and quality of the new recruits compare with their predecessors; the fact that more come from abroad does not mean that there is no brain drain. Nor, given the rarity of Ph.Ds in British Senior Common Rooms when I joined one without one less than 30 years ago, does the fact that more are in that sense highly qualified.

The brain drain debate is thus essentially about quality and almost entirely about subjective quality about which fruitful debate is difficult – de gustibus non est disputandum. One of the four pieces of relatively objective data is on the proportion of Fellows of the Royal Society resident abroad, which has risen from 16% in 1970 to 21% in 1986. Much of the brain drain occurs when doctoral or post-doctoral students leave intending to return but do not. Enquiries amongst colleagues on the Advisory Board on Research Councils as to the factors likely to determine whether a son doing graduate work in the U.S. comes back, all focussed on a single question: whether he was already 'married' to an English girl!

The support of research in the U.K. is by a dual system. For many years universities were funded by the University Grants Committee which was itself funded directly by the Treasury. The UGC gave block grants to universities and increasingly these were linked to specific development plans. A survey of university teachers' use of their time some years ago suggested that one third of it was devoted to what was called 'research'. From this it was inferred that a significant proportion of funds went on research although in rather unaccountable ways.

More explicit funding of research in the natural and social sciences (but not in the humanities) come from five Research Councils for Science and Engineering, Medicine, Agriculture and Food, Natural Environment and Economic and Social Research. The distribution, but not the scale, of the

funding of these facilities follows advice by an Advisory Board, of which I am a member. These Research Councils differ in several ways. SERC and MRC are much bigger than the others. MRC, AFRC and NERC are often described as 'mission oriented' and spend much of their money on research in their own establishments. SERC also has establishments for nuclear science and astronomy but these are overshadowed by so-called 'responsive mode' funding of applications for support. This is also typical of ESRC. Research Council support is given after 'peer review' of applications and other studies of applicants' track records. The UGC has been criticised for the lack of either prior assessment or university management of the research funded by its grants. In response to this pressure the UGC introduced a formula for determining part of its grant by reference to research-related criteria. In principal under the dual support system the Research Council could assume that an applicant from a university had access to a 'well equipped laboratory' and the Research Council only supplied extras such as research assistance, special equipment, or field trips. With the decline in the real growth of research funding from 10% p.a. between 1960 and 1976 to about $2\frac{1}{2}$% p.a. this assumption became increasingly questionable. Thus the UGC decided to offer an element of matching of RC grants – albeit on a lagged basis.

They also introduced an explicitly judgmental element based on the rating of departments' research output by panels of experts. I have served twice as a member of the economic panel. The 1986 exercise involved collating a lot of data and consideration, but not adoption, of bibliometric techniques. The review was subject to several criticisms. The most vocal, coming from aggrieved departments, was that the invitation to departments to submit lists of publications provided inadequate guidance for the assessors to take account of *scale*.

This complaint was met in the second, 1989, review by requiring a complete list of faculty and some of their publications in the three year review period so that non-contributors had to be declared. There has also been some discussion of the problem created by faculty movement. Is it work done in the review period at X, or by those now at X? In fact mobility is not very great but the fact that the latter criterion would allow universities to 'buy' output by aggressive hiring of recent publishers has discouraged its use.

The criteria are not exclusively bibliometric but include also success in securing RC, foundation or industrial research funding. This is obviously rather tricky. On the one hand given peer review there should be information about the quality of research at X revealed by its success in obtaining grants. On the other hand if a high score is going to lead to more UGC funding we should be concerned with incremental or marginal research output. The key question is whether university X has made good use of its research grants – they are an input and not an output.

It may not surprise you that this issue seems to have been of more concern to economists than to other disciplines. It was addressed in particular in a paper prepared by a committee of the Royal Economic Society (on which I served) when it was consulted, as were many other 'learned societies', by the UGC before mounting their second review.

They accepted little of our advice, probably because we suggested that research was necessary to establish the appropriateness of feasible bibliometric techniques and that considerable care would be necessary to design a system that did not embody perverse incentives (to which I shall come shortly). The UGC/UFC seem to prefer to meet the incentive issue by keeping wide open the question of how the next review is to be conducted and its findings used.

Another issue we addressed was the form in which the results were to be published. Although the UGC/UFC suggested a cardinal reference point for the top of the range ('world class') it was not clear that any other cardinal point was available so that the exercise was essentially a ranking one. Should the full ranking be published or should departments be grouped in e.g. five quintiles?

A lot depends on whether one assumes that the detailed information will become known. If it does, and if the grouping is truly justified by errors in ranking, then a pay-off function which is stepped in term of quintiles may set up perverse incentives. Getting a department up from the bottom of a class into the next one is hopeless. Why not close that department and put resources into one near the top of its class? If there were 50 departments in a subject it would clearly be silly to close the 20th, 30th, 40th in order to put resources into the 21st, 31st, 41st. This could happen if the ranking of universities' departments in different subjects were sufficiently weakly correlated.

Other incentive issues relate to such matters as multiple authorship – within and across universities, the relative weighting of articles and books, length and self or mutual cross citation. In fact I suspect that there have not been any significant distortions of individual behaviour because the universities, which are subject to incentives, have largely failed to transmit the incentives to faculty members by amending their reward systems taking advantage of new flexibility – though some are trying (and I shall come on to a related point shortly). Even In the absence of incentive problems the question how to weight publication of different length in different journals is in principal acute and weighting problems multiply if one starts counting, and weighting, citations as well.

I have mentioned that the Royal Economic Society had urged that research should be undertaken to validate the methodology to be used. We even volunteered to finance such a study should they fail to do so. We obtained somewhat more detailed returns than those required by the UFC

and they were analysed by Geraint Johnes of Lancaster University. The main result of his analysis (published in the Economic Journal) was that the ranking of most departments was extremely sensitive to the weighting scheme used.

One might be tempted to infer from this that the panel of assessors would find agreement difficult to achieve. This was not in fact the case in either exercise in which I was involved. Under English Law revealing what happens in the Jury Room is a serious offence, but I have never attended a meeting. In the recent exercise two were scheduled. I could not make the first but sent in my ranking, others tabled theirs, and agreement was sufficiently close for the chairman to feel (correctly) that he could elicit a consensus without a second meeting.

Applied Research which may be undertaken on commission from industrial or public sector bodies rarely produces publications that appear in the most respected journals. I had hoped that the money itself together with any matching UGC/UFC funds and the weight given to industrial funding as such in the research evaluation exercise would mean that this legitimate activity would be appropriately rewarded and its extent safeguarded. This does not appear (on admittedly a small sample) to be the case. Rather the weight on prestige publications was overwhelming in the research weighting and industrial firms are said to be anxious to place their contacts with highly rated departments. Thus a sound department that has concentrated on such work in the past not only finds its efforts unrecognised but its market threatened by its failure to obtain a high rating.

The search for performance indications in Higher Education has been at least as intense in the U.K. on the teaching side as for research. Many of the indicators used are in fact simply disaggregated cost data. The more interesting indices relate to:

completion rates,
degree results,
employment experience.

All of these are problematic for several reasons; is a dropout a failure? On the one hand the student may have learned much of use to him and society, on the other hand inducing him to stay may involve wasting resources on him. Degree results are obviously manipulable unless examination were to become an external and centralised function. Employment patterns differ markedly by subject and by type of course, e.g. those involving work experience often have success in placing their students – but this is not clear evidence of a superior mode of instruction as such 'sandwich' courses are expensive to organise.

British univerities have different histories and present pattern of organisation as well as subject mixes which enable them to be grouped, e.g. Oxbridge,

Scottish, ExCAT (College of Advanced Technology), etc. Dummy variables representing these categories remain important in regressions 'explaining' teaching-related performance indicators after all other individual institution-related variables have been accounted for – which is a disturbing challenge.

While completions rates have featured as a Performance Indicator at the undergraduate level their counterpart *submission* rates have been taken even more seriously by some research councils – notably the ESRC which found a few years ago that much less than half of the students it supported for doctoral work had submitted within four years. It proposed to apply sanctions – essentially removed from the list of institutions to which ESRC-supported Ph.D. students could go – against institutions which failed to meet a norm which was itself set to rise over time.

This procedure suffered in several ways. The sanctions applied at Departmental level although a student's progress depends on himself and his supervisor. In order to get statistically significant data on submissions in smaller universities all social science had to be aggregated which meant that economists could be betrayed by sociologists or vice versa.

In any case the number of ESRC awards was falling so that departments were having to look elsewhere, e.g. to foreign students, for their market anyway. Moreover the falling value of ESRC awards encouraged graduate students to seek employment at which point completion may add less value and certainly takes longer.

Finally submission rates are themselves manipulable. Most universities have an option to 'refer back' a thesis which is essentially unready. One university department I know of is contemplating effectively requiring four year submissions, whether the student is ready or not, and expecting to make greater use of 'reference back'.

Despite all these serious weaknesses I believe that the ESRC campaign has focussed attention on an important issue and raised the standards of graduate student supervision. Whether this would remain true if it were spread more widely and institutionalised is more questionable.

Having resolved no issues, conclusions may not be in order so I will finish with some more general reflections. The first is that it seemed easy to spend 10% more each year in real terms; when that era ended it was much more necessary to justify costs. This has led to an enormous proportional increase in the resources devoted to the management of science at all levels: the ABRC receives more funds for science policy studies; the UGC/UFC rating exercises make demands costed in the millions; universities are expected to set up committee structures, to develop research strategies and to manage research resources; when research councils offer what looks like extra money for which univerisites must compete one sees rent seeking at work with a vengeance.

Much of this may be healthy, but where performance indicators are

concerned nearly all that we have are at best proxies. This should remind us of two rules: of Goodhart's Law (formulated originally in the monetary field) that any observed regularity will break down when it is used for policy purposes; and the danger of mistaking proxies for the real thing as exemplified by some econometrics.

More generally, while my criticism of the piecemeal reforms introduced in Britain may be influenced by an unconscious nostalgia, it is not based on the premise that the previous arrangements were optimal. It is right that the balance and structure of a system as idiosyncratic as ours should be subject to scrutiny and pressure. What has evolved even more than what has been designed requires challenges if the process is to take it in the right direction – but, perhaps as in nature, the official challenges in this case show little more consistency and coherance of design than do the universities themselves. Nevertheless some of the participants in this system believe that its efficiency has been raised perceptibly.