RESEARCH PERFORMANCE INDICATORS IN AUSTRALIAN HIGHER EDUCATION

Geoff Harris

November 1994

UNE Working Papers in Economics No. 13

Editor John Pullen

Department of Economics
University of New England
Armidale
New South Wales 2351
Australia

ISBN 1 86389 216 8

Research Performance Indicators in Australian Higher Education

Geoff Harris

1. Introduction

Recent concerns with the measurement of research performance in higher education in Australia need to be understood in the context of the drive for greater efficiency and accountability in the public sector. One of the requirements for greater efficiency are appropriate performance indicators so that progress, or lack of it, can be identified, and appropriate incentives applied.

To begin with, we should note two respects in which universities differ from the public service in general. First, they are remarkably individualistic. As a result of architecture, lecturers spend most of their non-teaching time in their own offices. Much of their research work, especially in the non-sciences, is individual rather than co-operative. By far the greatest input into research, in the non sciences, is individual time and mental application. Teaching allocations, for the most part, give lecturers charge of their own courses, and require little interaction with colleagues. Second, at least until very recently, there has been a strong emphasis on equity as regards teaching loads. These are assigned in roughly equal proportions to all lecturers within a department, irrespective of the differenteffort individuals are allocating to research or 'service'. (As a broad rule of thumb, lecturers are expected to devote a third of their time to each of these three activities.) In a similar vein, pay rates for non-researchers are identical to those of active researchers, although promotion is more likely to be achieved by the latter.

In a broad sense, however, the issues involved in the reform of the public service in general and universities in particular are very similar. As the Secretary of the Department of the Prime Minister and Cabinet has recently noted, under the heading 'making performance count':

> People will not continue to strive for improvement unless they remain convinced that it matters, so we must ensure that our managers show an ongoing interest in the results being obtained, insist on accountability against established indicators, reward good performance as a practical means to make performance count and seek to improve that which is not of sufficiently high standard.
>
> Performance appraisal and pay are practical means to make performance count, providing the opportunity to link individuals' goals and contributions more clearly to those of the agency, to improve communications within organisations and to reward good performance on an individual basis. (Keating 1993.22)

This 'task focused' and individualistic approach is in accordance with the thrust of the recent Public Service Commission report *A Framework for Human Resource Management in the Australian Public Service* (1992), and may be contrasted with a 'high commitment employment system' where much attention is paid to career structures and job design to ensure employee commitment and to thereby enhance efficiency (Weller et al 1993.12-13). The individualistic

nature of lecturers' endeavours, and the 'task-focused' approach to productivity enhancement, are compatible when it comes to understanding and measuring university research performance and also mean that the measurement of individual performance is essential.

After briefly reviewing performance indicators in general, this paper focuses on performance indicators in higher education in Australia and the U.K. Specifically, it seeks to answer four questions:

- which measures should be used?
- which publications should be included?
- who should be included as members of the department?
- which time period should be covered?

2. Performance indicators

It is first important to distinguish between four related but distinct concepts used in evaluating research - quantity, impact, quality, and importance. The first of these is relatively straightforward, and can be measured numerically e.g. the number of articles or pages produced. The impact of a piece of research - its influence on related activities - is generally measurable by the number of citations made to it. Quality and importance, however, cannot be objectively measured: the assessment of quality is highly dependent on value judgements, and the importance of a piece of research may not become clear until time has passed. The confusion of these concepts is both common and crucial. Johnes (1988.56) contends that it is 'highly misleading to assign the term "performance indicators" to [bibliographic] techniques ... unless a very narrow view of "performance" is taken'.

There are many definitions of performance indicators but for our purposes, five aspects are important:

1. They are expressed numerically;
2. They relate inputs to outputs (i.e. they measure efficiency of resource use);
3. They are linked to the overall goals of the organization[1] (i.e. they are concerned with effectiveness in meeting desired outcomes);
4. They allow users to determine how the performance of the individual or organization under study has changed over time and/or how it compares with the performance of other individuals or organizations;
5. They may be used as incentives to influence the activities in socially-desired ways.

---

1 Allen (1988.98-104) lists 73 possible goals for U.K. universities.

PIs may be used at the institutional, departmental or individual level: for convenience we shall discuss their use with respect to departments.[2] Recent reviews of the use of PIs in higher education have been written by Cuenin (1987), Cave *et al* (1988), Pollitt (1990), Johnes and Taylor (1990b), Johnes (1992), Sizer (1992) and Stolte-Heiskanen (1992).

The use of PIs may have a number of objectives. Establishing relevant PIs may be part of the process of clarifying an organization's objectives; they may be used in the difficult task of evaluating final outcomes or impacts resulting from the organization's activities; they may, by relating inputs to outputs, indicate areas of potential cost savings or productivity increases by comparing input/output patterns between organizations; they may be part of a regular review or stocktaking of a department; and they may be used as an input to staff development and/or as an input to staff assessment/reward systems. It is the third of these - the efficiency aspect - which has dominated the development and use of PIs since the early 1980s, particularly with the linking of funding to performance (Pollitt 1990.68).

Several types of PIs may be distinguished (Cave *et al* 1988). These are input PIs, which measure the resources available to the department; process PIs, which measure the conditions (e.g. student:staff ratios) under which a department produces its output; and output PIs, which are concerned with results (e.g. number of publications). As noted, we are concerned to relate a department's output to both the conditions under which it operates and the inputs used to produce its output. These PIs may be compared according to various criteria. Cave *et al* (1988) suggest seven:

1. Type of indicator (input, output or process);
2. Relevance (how does it relate to the department's objectives and mission?);
3. Degree of ambiguity (ease of interpretation);
4. Cheat-proofness (to what extent can it be manipulated by departments?);
5. Cost of collection and availability of comparative data;
6. Level of aggregation (individual, department, discipline);
7. Relation to other PIs (are they unique or multiple indicators of the same attribute?).

Linke (1991, vol. 1.262) adds several others to this list: dependability (validity between departments and institutions) and durability (reliability over time). The list suggests several other points. First, since there are few absolute standards in higher education, the use of PIs usually occurs in a relative framework e.g. how strong is a department's performance relative to other departments (Hattie 1990.251; Linke 1991, vol. 1.xii). Second, inter-disciplinary comparisons are fraught with difficulty, given the very different ethos' concerning research grants, publishing and other aspects of academic activity. Even comparing departments of the

---

2 The AVCC/ACDP report (1988.2) justifies dealing with departmental and institutional PIs, and not individual PIs, on the ground that the latter are encompassed in industrial agreements.

same discipline may not be comparing like with like[3] and institutions with different disciplinary mixes will 'perform' differently as regards costs and other PIs. Third, PIs are not themselves the final word, but an input into the process of deciding on the final word. To quote Linke (1992.8, 13), they are 'an aid to decision-making and potentially a powerful one - but in no sense [are they] a substitute for what is an inherently subjective process [and not] ... a mechanistic or formula-based approach'. Or, as Roe and Moses (1986.2) found in their study of the departmental review process, academics were wary of a mechanistic use of PIs 'which might entirely miss or misrepresent the living reality of a department'. Judgements are needed at a number of points, including which PI to use and how to interpret the results from a number of PIs (i.e. what weights to attach to each). This relates to a point implied several times earlier: underneath it all, we are principally concerned with issues of quality, impact and outcome rather than simply performance. Quality is hard to define but involves judgements as to the excellence of performance and of the value or worth of what has been achieved (Bourke 1986.2, Linke 1991.128-131). The measurement of efficiency and effectiveness, by this argument, is a means to a higher end.

3. Performance indicators in higher education: recent Australian studies and experience

This section reviews a number of major Australian studies, both government and university in origin, which have examined the application of PIs to higher eduction. All are concerned with PIs at the department and university level, but Linke (1984), Roe and Moses (1986) and Roe *et al* (1986) also discuss individual PIs and the first of these considers system level aspects as well. The recommended PIs from five studies are summarised in Table 1.

Australian studies of PIs in higher education began in earnest with the CTEC-initiated report on quality and efficiency in Australian higher education (Linke 1984). This encouraged the development of a more systematic set of PIs but noted two requirements for their effective use:

> [First] that the measurement of educational effectiveness, quality and efficiency is predicated on the assumption of specific educational goas, and to be interpreted consistently and accepted universally requires that these be defined as clearly and precisely as possible ... [Second] that quantitative indicators of educational effectiveness and efficiency are inherently selective and insensitive to local or circumstantial conditions, which requires that they be qualified with an appropriate explanation of the context within which they are used and the limitations of their interpretation. (Linke 1984.iv)

---

3 'Is [the Department of] Medicine at Queensland University truly 'comparable' in terms of its historical and scientific context, for example, to Medicine at Sydney University?' (Grigg and Sheehan 1989.22). See also Moses (1990).

This was followed by Paul Bourke's influential overview *Quality measures in Universities* (1986). The results of these 'broad brush' studies, which had an underlying interest in quality, formed the foundation for the discussion on evaluation in CTEC's *Review of efficiency and effectiveness in higher education* (1986.258-271). The green and white papers on higher education (Dawkins 1987, 1988) laid down firmly the need for improved performance and the measurement of that performance by appropriate PIs, and heralded moves towards linking funding to performance. They made clear, as Hattie (1990) notes, that PIs, which had been seen as the property of individual institutions, would henceforth be much more public.

The AVCC/ACDP (1988) working party on performance indicators followed the green and white papers. Its central emphasis was that

> ... the need for institutions to be accountable and transparent can best be met not by the collection of indicators as such, but by their use in the context of a process of expert review. In that process, performance indicators form part of the necessary raw material of evaluation and assessment. These reviews give rise to judgements and decisions which, though they may make use of information supplied by the collection of data such as performance indicators, are, in the end, made by people or groups of people. It is, therefore, only to this extent ... that the use of performance indicators could be built into the processes of evaluation and management ... [We] are firmly opposed to the use of performance indicators by Government in any purely mechanical fashion, as in formula funding ... (AVCC/ACDP 1988.2)[4]

The next study was an extensive report from the University of Queensland (Grigg and Sheehan 1989). This examined 20 departments in five major universities (Melbourne, Monash, Sydney, New South Wales and Queensland) to determine which PIs could be used to assess institutional and departmental performance and to trial these for an average of the years 1985 to 1987. A range of potential indicators were ranked by academics, and the eight preferred indicators for social scientists are listed in the fourth column of Table 1. 'Refereed journal publications' and 'Chapters in books, editorships' were included in the first eight by 75 per cent or more of respondents.

Two particular PIs were tested in this study: a publications rate index was calculated by assigning weights to different types of publication and the impact of publications was based on the Journal Impact[5] as calculated by the Social Sciences Citation Index (see section 7). That is, an article appearing in a journal which has a higher citation record is given a higher impact score. The results for economics departments have been extracted and are reported in Tables 2 and 3. Table 2 reports the number of journal articles and journal impact. Columns 2 and 3 of that table indicate substantial variation between departments. Queensland for example produced

---

4 For an appraisal of the AVCC/ACDP report, see Teather (1990).

5 Journal impact is the number of citations made to articles published in journal x in period y, divided by the number of articles published in journal x during period y.

the highest number of journal articles per head, but had the lowest 'impact scores', whereas Sydney had the lowest number of journal articles per head but the highest impact scores. Which of the two departments is performing better, and how do they compare with other departments?

Table 1: Summary of recommended PIs relating to Australian university research

IHME, cited by Bourke (1986) and CTEC (1986)[1]	Roe and Moses (1986)[1]	AVCC/ACDP (1988)[1]	Grigg and Sheehan (1989)[2]	Linke (1991)[1]
Number of publications Citation indices Share of contract research Application for graduate study Number of dissertations accepted Consultances, inventions, patents Invitation to high level scientific conferences Election to membership of learned academies Awards of prizes and distinctions	Overseas and Australian trends in the discipline and position of the department in relation to these 1. Staff research Factors having adverse or favourable influence on research activities and rate of publication Research areas of staff and publications in these. MSs accepted, submitted, in preparation. Conference participation: papers given, attendance Distribution of postgraduate supervision in the department Principles and practices used in allocation of departmental research funds. Research grants received: number and size of grants from sources; number and percentage of staff receiving grants; comparison with similar departments in other universities Special studies programs, sabbatical and professional experience programs Consultation activities of staff 2. Student research Quality of student body at undergraduate and postgraduate intake level.	1. Research grants The number and dollar value of grants per EAS from Commonwealth bodies using peer review, private sector contract research and other funding, including government consultancies 2. Research and scholarly output The number of authored books, edited books, refereed chapters, refereed articles, published conference papers, unrefereed publications, creative works and other significant output, all per EAS 3. Research degrees completed 4. Impact Citations, with weighting according to rating of the journal 5. Professional services e.g. contribution to professional organizations, consultancies/advice to public and private organizations membership of learned societies, editorships of refereed journals, provision of continuing education programs, contributions to the media.	Refereed journal publications Chapters in books, editorships Research Grants Books Keynote addresses Successful postgraduates Editorship of journals Citations	1. Research grants i) average number of research grants for each 10 EAS ii) percentage of individual academic staff receiving a research grant iii) average value of research grants awarded for each 10 EAS 2. Publications Average number of publications per annum over three years, for each 10 EAS under three headings: i) books and monographs, weighted according to whether authored, edited or separate chapters; ii) refereed journal articles; iii) published conference papers. Other original works: i) public broadcasts and recordings; ii) registered patents, inventions and designs; iii) commercial and other published computer software. 3. Consultancies Average consultancy rate, being the number of paid consultancies per year for each 10 EAS 4. Professional Service activity

Table 1 continued

IHME, cited by Bourke (1986) and CTEC (1986)[1]	Roe and Moses (1986)[1]	AVCC/ACDP (1988)[1]	Grigg and Sheehan (1989)[2]	Linke (1991)[1]
	Origin of first degree of postgraduate students. Award and scholarships to postgraduate students Annotated list of completed honours and postgraduate theses. List of current research topics by honours and postgraduate students. Publications by honours and postgraduate students. 3. Others Number of postdoctoral fellows. Distinguished visitors: status, field, affiliation, dates of visit.			

Notes:
1 Not distinguished by discipline.
2 In order of preference, as ranked by social science academics, including economists, from five Australian universities; the first two were included in their top eight by 75 per cent or more of respondents.

Table 2: Measures relating to journal articles averaged for 1985 to 1987

	Average number of journal articles per annum 1985-87	No. of refereed journal articles per head[1]	Journal impact per member of staff	Impact per per paper
Melbourne	20.7	0.72	0.130	0.18
Monash	30.7	0.64	0.125	0.20
N.S.W.	38.7	0.72	0.217	0.30
Queensland	21.3	0.76	0.110	0.14
Sydney	26.0	0.63	0.247	0.39

Source: Grigg and Sheehan (1989.213, 141).

Notes: [1] i.e. of Equivalent Academic Staff

Table 3 introduces a cost perspective, allowing a comparison of the Total Indentifiable Research Expenditure-cost[6] of producing a publication and of producing a 'unit' of impact. The data suggest wide variation in costs: a publication for Sydney and a unit of impact from Melbourne were around four times more costly than those from Monash. We should note that the cost figures used here are not the full cost of research, and exclude the share of recurrent grant attributable to research. Monash University's highly cost-effective performance for 1985-87 results from its relatively low research income from the sources listed in footnote 1 of Table 3. Finally, we can also note wide variation in the number of completed research degrees per EAS, ranging from an average of 0.18 per annum for Sydney to 0.87 per annum for New South Wales (Grigg and Sheehan 1989.149). Such diverse results from the various PIs examined by Grigg and Sheehan - numbers of publications, publication impact, research grant income, cost per publication and unit of impact, and research degrees completed - make it very difficult to conclude which departments are better than others. The results call to mind the statement contained in several reports discussed earlier, that PIs are an input into the process of determining performance and are a complement to judgements by well-informed individuals and groups.

6 See footnote 1, Table 3.

Table 3: Measures relating to research output, averaged for 1985 to 1987

	Total identifiable research expenditure[1]	Publication rate index[2]	Average cost per publication[3]	Average impact[4]	Average cost per 'unit' of impact
Melbourne	233,773	48.3	4633	3.76	59514
Monash	88,852	58.0	1532	5.95	14933
N.S.W.	280,586	74.7	3756	11.69	24002
Queensland	109,094	43.7	2496	3.18	34306
Sydney	329,406	57.0	5779	10.20	32294

Source: Grigg and Sheehan (1989, Appendix L).

Notes: [1] Includes competitive peer-reviewed research grant, contract research, consultancies, donations specifically for research. Excludes the cost of time devoted to research by teaching staff (see Grigg and Sheehan, Appendix A).

[2] Calculated by expressing publications in journal equivalents, where a refereed journal article = 1, a book = 5, etc.

[3] An approximation reached by dividing column 1 by column 2.

[4] i.e. each refereed journal article weighted by the impact factor of its journal.

Between 1985 and 1990, DEET commissioned reviews of four disciplines throughout Australian higher education, viz. Law, Engineering, Teacher Education in Mathematics and Science and Accounting. In general, the reviews studied the work of each higher education institution involved in the discipline with particular emphasis on quality and efficiency. The research performance of each department was studied and two of the reviews, engineering and accounting, developed research PIs. Overall, research performance in these disciplines was unimpressive: a review article concluded that 'there appears to be a minimal amount of research in the disciplines of law, accounting and teacher education' (Connell 1991.38), to which could be added the 'disturbing result' that 25 per cent of teaching and research staff in university engineering schools, and 77 per cent in the CAEs, had not published in a refereed journal between 1982 and 1986 (DEET 1988.volume 1.74).

The *Review of the Discipline of Engineering* (DEET 1988), in its assessment of engineering departments in 24 institutions, used three PIs relating to research: the Relative Publication Rate for each department was estimated in terms of Basic Publication Units[7] per person per annum, 1982-86; the Relative Consultancy Rate for each department was the average number of consultancies per person per annum, 1982-86;[8] and Industrial Patents held or applied for. The

7 Each publication was scored, with one sole author a refereed journal article worth 1.0, a sole authored book worth 10.0 etc.

8 The size of consultancies in terms of time involved and income earned was not measured.

Review considered that these provided 'a reasonable basis for examining systematically the differences in performance between different engineering schools and departmental units ...' (DEET 1988.volume 2.76).

Table 4 presents the three basic measures for eleven universities together with total indentifiable research expenditure, extracted from various parts of the report. Wide variations can be observed between universities in respect of the research PIs and research expenditure. Using the approach of Grigg and Sheehan (1989), we may apply the publication PI to the research expenditure PI and thus estimate the cost per basic publication unit. An enormous range is evident, from a low of $800 for Wollongong to $11,475 for Melbourne.

Table 4: Performance indicators relating to the Australian engineering discipline

	Total Relative Publication (BPU/person /year, 1982-86)	Relative consultancy rate (average number of consultancies per annum, 1982-86)	Percentage of staff with Australian patents/ applications	Total research expenditure ($'000), 1986	Cost per BPU[1]
Sydney	0.85	0.94	3.9	2891	3401
N.S.W.	0.54	0.89	2.0	3592	6651
Newcastle	0.79	1.14	-	1008	1276
Wollongong	0.55	0.63	-	440	800
Melbourne	0.32	0.47	5.6	2672	11475
Monash	0.73	1.65	4.0	1747	2393
Queensland	0.68	0.51	15.3	3270	4809
James Cook	0.29	1.12	-	567	1955
Adelaide	0.36	0.45	13.3	591	1642
W. Australia	0.55	0.77	-	3219	5853
Tasmania	0.22	0.45	-	234	1064

Source: Derived from DEET 1988 (volume 1.71, 72, 80; volume 2.116).

Note: [1] i.e. column 1 divided by column 2.

The companion report on the accounting discipline (DEET 1990) did not develop PIs as such but does present data on several variables which can be used to distinguish performance between departments. These include the number and value of research grants received (both internal and external) and the proportion of staff receiving such grants, 1987-89; days spent in consultancy each year, 1987-89; and Publication Units[9] per staff member averaged over the period, 1981-89.

Finally, a research group chaired by Russell Linke (1991) was set up to develop and test a range of PIs suitable for both system and institutional levels. This built on the AVCC/ACDP (1988) recommendations, and produced what is now the definitive list of PIs for Australian universities, broken into four groupings: PIs relating to institutional context (e.g. student staff ratios); those measuring institutional performance in respect of teaching and learning (e.g. student progression rates); those measuring institutional performance in respect of research and professional services (these are listed in the last column of Table 3); and those relating to participation and social equity. Table 5 presents the results for three unnamed institutions - C, G and L - for four research PIs. Institution G seems to have outperformed C on almost all indicators, but it is much less obvious which of G and L performed better in 1987-88. The point may be reinforced by considering 'Mathematics, Statistics', the AOU for which most data were available (Table 6). Institution C, for example, received fewer research grants than G, but of higher average value. The latter had a stronger publication record than C, but C was clearly superior as regards professional services.

---

9 As with the engineering review, a score was attached to publications of different type. Sole authored book chapters, refereed articles, professional journal articles, research or consulting reports and conference papers all received 1.0. Sole authored books received 10.0 and edited books (single editor) received 4.0 (DEET 1990.volume 2.83).

Table 5: Research performance indicators for 'Economics, Commerce' academic staff in three institutions, 1987-88

	Institution C	G	L
Number of research grants per EAS, 1987[1]	0	0.6	-
Value of research grants per EAS, 1987 ($'000)[1]	0.7	5.0	-
Average publication rate, 1988[2]			
• Books and monographs	0.13	0.56	0.08
• Refereed journals	0.26	0.89	1.00
• Conference proceedings	0.20	0.71	0.00
Percentage of academic staff engaged in specified professional services, 1988[3]			
• Professional organisations	7	11	9
• Editorial services	14	6	11
• Expert bodies	9	11	11
• Examination of theses	5	0	9

1 Linke (1991.95). Zero entries are in some cases due to rounding errors.

2 Linke (1991.100). Based on equivalent sole author publications.

3 Linke (1991.111-113). Units per EAS.

Table 6: Research performance indicators for 'Mathematics, Statistics' academic staff in four institutions, 1987-88

	Institution			
	C	G	I	L
Number of research grants per EAS, 1987[1]	0.3	0.7	-	-
Value of research grants per EAS, 1987 ($'000)[1]	2.9	2.3	-	-
Average publication rate, 1988[2]				
• Book and monographs	0.07	0.00	0.01	0.00
• Refereed journals	0.70	0.86	0.00	1.32
• Conference proceedings	0.05	0.18	0.30	0.74
Percentage of academic staff engaged in specified professional services, 1988[3]				
• Professional organizations	14	3	11	26
• Editorial services	30	9	11	16
• Expert bodies	10	6	11	0
• Examination of theses	4	3	43	0

1 Linke (1991.95). Zero entries are in some cases due to rounding errors.

2 Linke (1991.100). Based on equivalent sole author publications.

3 Linke (1991.111-113). Units per EAS.

These do not exhaust Australian PI studies. Other recent examples include Paul Bourke's (1991) use of a number of citation measures, as part of Linke's (1991) review, to compare the research performance of various science departments in the ANU's Institute for Advanced Studies with that of other Australian universities. The opinions of researchers concerning relevant PIs for their disciplines were examined by Pettit and Low Choy (1992). Their respondents comprised 3990 successful ARC grant applicants from 24 disciplines, including 119 economists. Respondents were asked to rank 20 possible PIs of which seven were strongly supported[10] by economists, namely publications, peer reviewed books, keyonote addresses, conference proceedings, citation impact, chapters in books and competitive grants. The two most important were publications (refereed journal articles) and peer reviewed books.

10 i.e. 2 per cent or less of respondents stated that these PIs were irrelevant (Pettit and Low Choy 1992.18-19).

Paul Bourke of the ANU and Ben Martin of the University of Sussex have collaborated in a major study of British and Australian publications between 1981 and 1990. Based on the SCI, the SSCI and the Arts and Humanities Index, this research aims to assess the strength and weaknesses of publication and citation as tools for university research.[11] The ARC has commissioned several discipline reviews, specifically for history and economics (ARC 1993), to examine whether value for money has been achieved.

4. Performance indicators in higher education: recent UK studies and experience

There are strong parallels between recent UK and Australian experience in the higher education sector which makes examination of UK policy and practice very instructive The following quotation could apply to either country, the only significant difference being that Australia seems to follow the UK after a lag of several years:

> According to the government, the higher education sector could improve its efficiency and effectiveness in several ways:
>
> 1. *Higher education should be more responsive to the needs of the economy.* This will require closer links to be forged between higher education and industry. In addition, it will also be necessary to switch the subject mix away from the arts and humanities towards technical and vocational courses.
> 2. *Higher education depends far too heavily on public funds* and greater efforts are needed to raise private funds through applied research, consultancy and continuing education.
> 3. *Greater selectivity is needed in the allocation of research funding* so that more resources are concentrated in the centres of excellence.
> 4. *The higher education sector needs to be more cost-conscious* and should manage its resources more efficiently and more effectively. This will require the construction and regular publication of a range of performance indicators. These will be used to aid the resource allocation process both within and between institutions. (Johnes and Taylor 1990b.2)

Several studies (e.g. Sizer 1988, Johnes and Taylor 1990b.48-49, Pollitt 1990.62, Cave *et al* 1988.10-14) provide overviews and chronologies of the key events in the move towards the use of performance indicators by British universities. These commenced with major cuts in University Grants Committee (UGC) funding to U.K. universities in 1981; whilst the average reduction was 17 per cent, the cuts ranged from 6 to 44 per cent between universities based on 'selective judgements made according to criteria which were not fully disclosed' (Pollitt 1990.62). There followed a period of relative calm until 1985, when the Jarratt Report (CVCP 1985) allowed vice-chancellors and principals to express their views on the increasingly obvious direction of government policy towards higher education. The Jarrratt Report emphasised greater efficiency and effectiveness and the development of PIs. It was followed

---

11 A preliminary report on this research was published in *Australian Campus Review Weekly*, August 6-12, 1992, 11-12.

by several major government reports (Department of Education and Science 1985, 1987) which made it clear that the relative performance of institutions and departments would henceforth influence funding allocations. Thus far, PIs have only been used to measure research performance - in the two research selectivity exercises of 1985/86 and 1989.

The first of these essentially involved a peer review, informed by performance indicators but in ways which were not made public. In brief, each cost centre (i.e. academic disciplines or related groups of disciplines) in each UK university was ranked as 'outstanding', 'above average', 'average' or 'below average'. Whilst there are many who believed that the UGC 'should be given credit for attempting honestly, dispassionately, and not without some ingenuity, an onerous, complex and pretty thankless task' (Smith 1987.307-308), many criticisms have been levelled at these rankings. Representative of these critiques are Gillett (1987), T. Smith (1987), D. Smith (1988) and Platt (1988). Criticisms may be grouped under three headings: how the UGC went about its task, the validity of the methods used and the use which might be made of the results.

The first groups of criticisms centred on the exercise itself and derive from the enormity of the task of reviewing 37 cost centres across 55 universities given very limited time and resources.[12] The various sub-committees apparently adopted different procedures and awarded different results to different cost centres samplings; by way of example, 26 per cent of geography departments were graded as outstanding, compared with 5.7 per cent of political science departments (Smith 1988.4). In addition, apart from not making public the procedures and weightings used, the UGC was not responsive to appeals and complaints. The second set of criticisms relates to the methods employed, described by one commentator as 'a pretty rough and ready lash-up of techniques' (Smith 1987.309). The most important of these emphasise that the UGC's rankings were inconsistent with other objective indices, particularly numbers of publications per capita and citation rates. From the perspective of psychology, Gillett (1987.46) comments that the

> UGC ratings bear no relation to actual research output of departments in the 'snapshot' period. The UGC ratings are unrelated to either quantity or quality of research as measured by internationally recognised, objective indices. In terms of these indices, the UGC ratings have approximately zero validity. Particularly disturbing is the finding that a number of departments rated as 'below average' actually produce more and better research on average than several departments rated as outstanding.

Similar criticisms came from geographers (Bentham 1987, Smith 1988), political scientists and a number of other disciplines.

---

12 For brief reviews of the 'inner workings' of the UGC exercise, see Smith (1987) and Flemming (1991).

There was criticism of the use of research grant income as an important indicator, when it has no necessary relationship with output in terms of published work. There was also the contention, across a number of disciplines, that the UGC methods favoured large departments. This could occur in two ways: first, there appears to have been an ideological bias in favour of big departments at the time and second, to the extent that the opinions of individual assessors were important, large departments would do better because their aggregate research output would be greater and the department therefore better known.

The third set of criticisms relates, broadly, to the use which was made, or was believed likely to be made, of these rankings. Obviously, cost centres receiving below average rankings were under threat of censure if not closure. Given the inconsistency of some UGC rankings with with other more objective PIs, this was galling to many individuals. More generally, the measures were output PIs only, and did not measure efficiency of resource use. Cost centres could do well in output terms simply because they were well resourced; they might be relatively inefficient, but would still be rewarded by the UGC approach. Another point concerns cost centre comparisons. Whilst it may be valid to compare the same cost centre between universities, the rankings cannot be validly used to compare different cost centres in one university. The UGC rankings cannot tell whether an average history department is worse or better than an above average physics department. This points to the potential for undesirable consequences of such rankings. Given that only research has been subject to performance measurement, it is likely that it will be emphasised ahead of teaching and other activities. Was such an outcome intentional or accidental? More elegantly, Smith (1988.7-8) discusses the possible use of PIs to measure the performance of artists.

> The performance of the painter could be measured by the number of paintings produced, their size, and even by their monetary value in the market, and related to the cost of materials and time taken, but only the most dedicated philistine would argue that these are criteria which can sensibly differentiate among painters on a qualitative scale.

Clearly, the choice of PIs used is crucial, since they send specific signals to the players involved. Without due thought, the signals may result in unintended and undesirable consequences.

The first research selectivity exercise had important results as regards resource allocation. In 1986/87, 35 per cent of government funding allocations to institutions were based on research performance, and 40 per cent of this was determined by peer review, assisted by bibliometric measures (Clayton 1987). By 1995, the UFC (the successor to the UGC) aims to double the proportion of its research block grant which is tied to these indicators (Pollitt 1990.70). The replacement of the UGC by the Universities Funding Council (UFC) in 1987 was much more than a change in title, and gave advance warning of the government's intention to move from a

system of block grants to institutions to allocations based on specific contractual agreements between each institution and the UFC.

A second research selectivity exercise was undertaken by the UFC in 1989,[13] still using the informed peer review approach of the earlier exercise but with more preparation and some significant modifications. In particular, less emphasis was placed on research grant income and more on objective measures of research output. Another feature was the ranking of departments against international standards[14] which, it was hoped, would allow comparisons to be made across disciplines within the same university. The five point scale used was common to all subject areas. The aim of the exercise was to assess research quality, which was not defined, but was measured by publications or equivalent (up to two for each full time member of staff), success in gaining research grants or studentships and research contracts and the professional knowledge and judgement of the advisory groups and panel members, possibly supplemented by advice from outside experts.

The exercise was still roundly criticised, albeit much less so than in 1986. The major criticisms centred on inadequate instructions to departments regarding what to include under various publication headings and continuing concerns over cross-subject comparisons. Some cost centres, unwittingly or with intention to deceive, benefitted from imprecise definitions of research output.[15] A study carried out for the same reference period (1984-88) collected data from 32 UK economics departments, broken into eleven categories.[16] Given that the number of journals featuring in the five year bibliographies was over 800, particular attention was given to publications in major academic journals, these being those indentified by Diamond (1989a) as 'core economics journals' (see section 6.2). Of the ten economics journals in which UK academics published the most, only four were in Diamond's list of core journals. Johnes reports (1990.560) that the weights assigned to different publication types led to different rankings, casting doubt on the usefulness of bibliometric analysis in ranking departments.

The second set of major criticisms related to comparisons between subject areas. One reason for these were the substantial inter-subject differences in mean scores awarded, with its

---

13 The results were published in *The Times Higher Education Supplement*, September 1, 1989.

14 The top rating was awarded to cost centres with 'research quality that equates to attainable levels of international excellence in some sub-areas of activity and to attainable levels of national excellence in virtually all others' (Johnes and Taylor 1990b.157).

15 To quote Johnes and Taylor (1990b.158), 'in some cases, edited books were counted as authored books; book reviews were included as articles; unpublished research reports were included as books; no distinction was made between articles published in non-refereed journals and articles published in refereed journals, and so on ... In some cases, inaccurate publication dates were included in order to gain advantage and publications were included when they should have been attributed to another institution.' Much of this occurred because of the imprecise definitions supplied by the UFC.

16 Viz., papers in academic journals, letters in academic journals, articles in professional journals, articles in popular journals, authored books, edited books, published official reports, contributions to edited works, contributions to conference proceedings, other publications and other media.

implications for shifts in resources away from low-scoring subjects towards high-scoring subjects. Were the differences in research rating between subjects a result of 'real' differences in research quality or to differences in standards between advisory panels? Are differences in research output between subjects largely a result of different resource endowments? If so, the giving of more resources to high performers may worsen efficiency. Using the weighted average UFC rating for each university as their dependent variable, Johnes and Taylor (1990b.165ff) used logit analysis to test the importance of five explanatory variables in explaining inter-university differences. These were student/staff ratio (adjusted for subject mix),[17] 'research only' staff as a proportion of full time academic staff, research expenditure per full time academic staff (adjusted for subject mix), whether or not a university is located on the periphery of the UK (i.e. Wales, Scotland, North Ireland) and whether or not the university is an ex-college or polytechnic. The research only staff and research expenditure variables were themselves strongly correlated. When one was omitted, each of the remaining variables had a significant impact on the universities' research rating and in the expected direction. They explained around 60 per cent of the variation in research ratings between universities. The measurement of research performance, therefore, requires that account be taken of particular inputs and specific characteristics.

17 I.e. given that different subjects typically involve different student/staff ratios, universities with different mixes of subjects would be expected to have different ratios. The variable adjusts for such differences in subject mix.

The foregoing point may be reinforced by considering non-research PIs. The most common PIs in use for UK higher education are unit costs, non-completion rates, degree results, the first destination of graduates and research output. In a series of studies, Johnes and Taylor (1987, 1989a, 1989b) have shown that whilst there were stable and significant differences between universities in respect of each of non-research PIs, 70 per cent of the variation between universities in unit costs could be explained by differences in subject mix; most of the variation in degree results and non-completion rates could be explained by the A level scores of students; and 90 per cent of the variation in the proportion of students in permanent employment six months after graduating can be explained by subject mix and a number of other factors (e.g. regional unemployment rates) outside the control of universities. The use of such indicators, then, as a measure of 'performance' with resource allocation implications, seems to be seriously flawed. Since the type and quality of inputs vary between universities, so outputs will vary between universities. Of course, it is easy to improve performance: closing down high cost disciplines such as engineering, expanding low cost disciplines such as accounting, and inflating student marks would each result in improved 'performance', but to what advantage as far as society is concerned?

The results of the 1992 research selectivity exercise were released in December, 1992.[18] Departments were invited to submit, for each member active in research, their two best articles or books over the past four-and-a-half years. Assessment panels also received data on the total number of articles or books published during the assessment period; the number of postgraduate studentships; future departmental research plans; and outside research grant income. A similar five point scale to the 1989 exercise was employed with five points awarded to departments with 'some research of international excellence and the rest of national excellence'. The institutional rating was calculated by summing the rating of each department multiplied by its number of active research staff and dividing by total 'research active' staff.[19] The mean score for the 60 economics and econometrics departments was 3.73, which ranked equal fourteenth of the 72 subject areas in terms of international standing.

5. Summing up on PIs

From this wide ranging review of PIs, several points stand out. First, since one of the major uses of PIs is to measure performance in an efficiency sense, it is essential to understand whether individual PIs relate to inputs, process or outputs and to relate outputs to inputs. The presentation of output PIs without relevant input data can indicate very little about efficiency.

---

18 *The Times Higher Education Supplement*, no. 1050, December 18, 1992, i-xvi.

19 I.e. if an institution had two departments of 10 and 20 staff, which received ratings of 4 and 3 respectively, its score would be (10 x 4) + (20 x 3) = 100:30 - 3.3.

Second, comparisons of the relative effectiveness of departments and institutions are very difficult to make, notwithstanding the use of PIs. Given the government's intention to reward, from 1994 onwards, those institutions which use their resources most effectively (Baldwin 1991.3-4), this is an important matter. As an illustration, consider rankings publications in refereed journals for three institutions, as reported in Table 7, derived from data presented in Linke (1991.100). Whilst saying nothing concerning absolute performance, the rankings for six AOU groups indicate that each of the three institutions ranked first in two. From these data it is impossible to determine which is the best performing institution, a point also discussed in section 3. The compilation of more than one PI requires that their respective importance be estimated i.e. that weights be attached to each. Different weights are likely to lead to different results.

Table 7: Rankings of three institutions according average publication rate in refereed journals, 1988

Institution	First ranking	Second ranking	Third ranking
G	English/Communication Studies Economics/Commerce	Psychology/ Behavioural Sciences Mathematics, Statistics, Civil/Structural Engineering	Physical/ Materials Sciences
C	Physical/Materials Sciences Civil/Structural Engineering	English/ Communication Studies	Psychology/ Behavioural Sciences Economics/ Commerce Mathematics/ Statistics
L[1]	Psychology/ Behavioural Sciences Mathematics/ Statistics	Physical/Materials Sciences Economics/Commerce	English/ Communication Studies

Note: [1] Engineering was not reported for Institution L.

The use of more output PIs may make the answer more difficult rather than easier. In terms of the value of research grants earned per EAS in 1987, institution C ranked first in three AOUs and institution G in the other three. The point of both these examples is that numerical PIs need to be interpreted by informed judgement; PIs are indicators which make judgement more informed rather than the final measure themselves. As a recent commentator puts it:

> Numerical measures should act as a control for peer review, but should be grounds for reconsideration only where there is a gross mismatch. Numerical measures cannot substitute for fine decisions. It is only because the number system is so finely divided that this is not obvious. (Collins 1992.15)

The same considerations apply, with even more force, when considering whether the desired final outcomes of an activity are being met.

Third, PIs themselves will influence ouput. In our research example, if numbers of publications are the main indicator, then we may expect many smaller articles (in size and significance) rather than fewer larger ones, but in what sense will research performance have improved? If citations are used, there will be an inflation of citations, but in what sense will research performance have improved? If peer review is used, resources may be allocated to enhance a department's image, but will its research performance improve as a result? Those responsible for establishing performance indicators need to be mindful of their influence on the type of research carried out and the form in which it is published. More broadly, the use of PIs only to measure research, as by the UGC/UFC in the UK, and the tying of results to funding allocations, may well result in lesser efforts as regards teaching.

Fourth, there remain very considerable difficulties in comparing performance across disciplines.[20]

6. Previous studies of research performance

More than 30 studies are summarised in Table 8. These include six Australian studies, which examine a range of disciplines, and 25 studies of the economics discipline published from 1975 onwards, very largely for U.S. universities. Notes, comments and replies are

---

20 This issue has been studied by Moses (1990).

Table 8: Summary of studies of research output in university economics departments

Coverage	Principal objective(s)	Principal measure(s)	Principal results	Reference
Australia, 1959-77	To quantify publication output within one department of agricultural economics over time; to test the influence of average age, number of courses, qualifications, number of research assistants, number of tutors and departmental size, all on an annual basis, as explanations for variation in research output over time.	Number of pages contributed to selected journals, conference papers, non-refereed journals, books (whole or chapter), monographs, bulletins and miscellaneous publications per lecturer per annum.	The department was found to be maintaining its publications output over time. Of the explanatory variables, two were significant. Aging had a negative impact on output but this was balanced by the department becoming better qualified.	Anderson (1978)
Australia, 1974-78	To measure the research performance of Monash University across all disciplines for the five year period.	A publication rate index, derived by assigning points to different publications.	Publication rate indices were calculated for different ranks, and the distribution of PRI was examined for one faculty.	West et al (1980)
Australia, 1972-80	To compare a small number of universities across a range of disciplines in terms of ARGC grants.	Value of grants, and number of grants per staff member.	Results were presented for the seven largest universities and five disciplines, not including economics.	Brown and Nunn (1981)
Australia, 1978-82	To compare 19 universities across three disciplinary groupings in terms of ARGC grants received per staff member.	Value of grants received, and number of grants, per staff member.	Rankings for 19 universities suggest that small universities performed relatively well and that comparisons confined to the largest universities are therefore misleading.	Bourke and Simondson (1982)
Australia, 1981-82	To compare universities according to ARGC and NH & MRC grants	Value of grants and number of grants received per staff member	Nineteen universities were ranked.	Hancock (1983)

Australia, 1975-82	To compare the research performance of Australian university physics departments.	Number of publications (excluding books, literature reviews and conference papers), number of citations received per paper, and research grants received from selected funding bodies, each in total and per capita terms.	Eighteen universities were ranked. Small departments performed relatively well.	Campbell and Campbell (1984)
United States, with particular reference to southern US, 1970-74	To rank US economics departments, and separately, the top southern US departments.	Total pages published and number of publications, in 24 economics departments, In one variant, double weight was given to publications in six 'top' journals.	The top 100 US and the top 20 southern departments were ranked. The double weighting given to publications in the other six journals made little difference to rankings.	Niemi (1975a)
Southern United States, 1974-76	To rank southern US economics departments by citations	Total citations and mean number of citations per department member, excluding self citations.	A ranking of southern departments is presented.	Gerrity and McKenzie (1978)
United States, 1972-74	To rank economics departments according to the publication performance of their members, by each of full professors, associate professors and assistant professors.	Number of pages per faculty member in 45 major economics journals.	The top 42 universities were ranked for each of the three professorial levels. Measurements were also made of pages per faculty member in the top 10, 20 and 30 journals.	House and Yeager (1978)
North America, 1970-74	To rank economics departments according to the publication performance of their members; to compare this with other ways of ranking; and to examine the effects of tenure on individual productivity.	Number of full-length articles in 20 major economic journals, by department and by individual. (The affiliation was the author's most recent department).	The top 82 universities were ranked by articles per faculty member but also by total articles produced and percentage of the faculty who published; peer reviews were found not to correlate highly with articles per faculty member; and tenure was unimportant in explaining variation in publication per head between individuals.	Bell and Seater (1978)

Canada, 1975, 1977	To evaluate scholarly productivity in 31 Canadian universities with particular emphasis on psychology departments	Total number of publications (university in total) and total and average citations (psychology departments)	Six universities account for 51 per cent of total scholarly publications; productive universities overall were found to be productive across a range of disciplines and were relatively large and wealthy; total university publications were closely related to citations earned by psychologists.	Rushton and Meltzer (1979)
World-wide, 1950-79	To rank institutional economics departments according ot the publishing performance of their members.	Number of pages in four major US economics journals, in total.	The top 25 institutions (22 academic and 3 non-academic) were ranked; the institutional distribution of contributors became more equal between 1950 and 1979.	Shim (1982)
United States, 1974-78	To rank economics departments according to the publishing performance of their members; to shed light on publication constraints and incentives.	Number of *American Economic Review* - equivalent size pages in 24 major economic journals, 1974-78, in total and per staff member.	The top 240 schools were ranked on the basis of total pages and per faculty member; rankings were compared with results of previous studies; pages per faculty member were regressed against some explanatory variables, with professorial salaries being the most important incentive and teaching load the most important constraint.	Graves et al. (1982)
Worldwide, 1978-83	To update the rankings of Graves et al.; to include 40 top departments outside the US; and to compare the 1974-78 and 1978-83 periods.	Number of pages in *America Economic Review*-equivalent size pages in 24 major economics journals, in total and per staff member.	243 US departments, and the 40 top non-US departments, were ranked. Whilst there was some movement by individual universities over the two periods, aggregate concentration ratios changes very little.	Hirsch et al. (1984)

United States, 1960-72; 1970-79	To rank US economics departments according to the publication performance of their members.	Number of *American Economic Review* - equivalent pages published in four top journals, with credit given to the department only if the author was listed as a member of the department in specified lists (as opposed to the affiliation listed on the publication).	The top 50 departments were listed. Significant differences were found between the total pages by current affiliation compared with listed affiliation.	Hogan (1984)
United States, 1970-79	To examine the relative publishing performance of Ph.D. programs.	Number of *American Economic Review* - equivalent pages published in three top journals, in total and per capita terms.	Authorship was highly concentrated among graduates of a small number of programs, with four programs accounting for 43 per cent of total pages.	Hogan (1986)
United States, 1978-81	To rank Ph.D. - granting departments by the number of citations attributed to the members.	Total citations and citations per staff member. Adjustments were made for the age of staff members (because older staff, whose cited work may have been published years before, will have an advantage) and citation dispersion (because high dispersion can reflect a lack of 'depth').	The top 122 departments were ranked by total and per capita citations, and the top 40 also controlled by age and dispersion. The top 20 departments measured by citations are very similar to those measured by publication-based ranking. After the top 20, there were many changes.	Davis and Papanek (1984)
United States, 1971-83	To examine the quantitative and qualitative publishing performance of the top 50 economics departments, and the performance of their graduates with respect to publishing and their placement in the top 50 schools.	Number of *American Economic Review* - equivalent pages, in total and per capita terms, in 27 journals; total citations, citations per faculty member, citations per article. Each of these five measures were applied to both current members of staff, and to graduates according to their department of origin. A composite ranking was developed on the basis of all measures.	The top 50 departments, as identified by Graves et al. (1982), were re-ranked, according to staff performance. On the quantitative measures, very similar results were recorded to previous studies. There was also strong consistency between these and the first two citation measures, but not with citations per article. For graduates, the total measures gave similar results but on per graduate measures, many 'name' departments fell in the rankings. Changes over time between 1971-76 and 1977-83, were also reported.	Laband (1985a)

Canada, 1975-84	To rank Canadian economics departments by the research productivity of their graduates.	Number of articles per graduate, number of *American Economic Review* - equivalent pages per graduate. Number of citations per article; number of citations per graduate.	Seven Canadian departments were ranked.	Laband (1985b)
Worldwide, 1974-76	To measure the extent to which editors of journals favour authors from their own institutions over outsiders.	Average length of articles published by authors from the same institution as the journal editor and those from different institutions.	Regression analysis carried out on 5,880 articles in 39 journals found strong evidence of greater average length of articles from authors from the same institution as the editor.	Laband (1985c)
United States, 1975-84	To rank US economic departments by the research productivity of their Ph.D. graduates.	Number of articles published, number of *American Economic Review* - equivalent sized pages published and number of citations, each on a total and per graduate basis, in 24 major economic journals; number of citations per article.	Eighty departments had graduates who had published in the sample journals. There were notable differences between these rankings and those based on the research productivity of staff.	Laband (1986a)
North America, 1980-84	To rank agricultural economics departments by citations received.	Total citations to works of current staff; citations per staff member, citations adjusted for years since receipt of highest degree; and citations per member of staff adjusted for the time available for research.	The top 25 US departments were ranked. Different citation performances were noted for three types of US universities, and Canadian universities. On average economics staff had three times the citations of agricultural economics staff (19.8 vs. 6.4)	Beilock et al. (1986)
United States, 1978-83	To rank economics departments by *Journal of Economic Literature* abstracts per capita.	Number of abstracts of journal articles published in the *Journal of Economic Literature*, per capita.	The top 50 US departments were ranked. In general, support was found for previous rankings, but some prestigious departments did not perform well, and some small departments performed strongly.	Golden et al. (1986)

United States, 1971-83	To compare the relationships between the various measures by Laband (1985)	Number of *American Economic Review* - equivalent pages, total and per capita in 27 journals; total citations, citations per faculty member, citations per article.	A strong degree of consistency in rankings occurred across a wide range of measures.	Brar et al (1987)
North America, 1980-84	To rank departments of agricultural economics by citation analysis, but including five additional regional journals, second and subsequent authors, and dealing differently with self-citations.	Total citations from the SSCI, the SSCI plus regional journals, and a data base, termed AGEC, comprising the *American Journal of Agricultural Economics* plus five regional journals. Each of these three were analysed with joint authors allowed for and self citations omitted.	The top 73 North America departments were ranked. The inclusion of regional journals with the SSCI, the inclusion of joint-authors and exclusion of self-citations made little difference to rankings. The AGEC data base led to substantial changes in rankings.	Beilock and Polopolus (1988)
United States, 1980-86	To provide information on the publishing records of Ph.D. granting departments in order to determine the most popular fields, by department.	Total and per capita *American Economic Review* - equivalent pages in the 45 most-cited journals in 19 fields of economics.	In general, there was a close relationship between performance in microeconomic and macroeconomic theory and overall publishing performance, but some important variations in performance occurred across fields.	Tremblay et al (1988)
United Kingdom, 1980-84	To rank British economics departments by research performance and to explain difference between departments.	Number of articles, and number of pages, in 20 top journals.	The then 40 universities were ranked and a number of explanatory variables were examined.	Johnes (1988)
United States, 1988	To rank departments based on representation on the editorial boards of 25 top journals.	Number of editorial board memberships, in total and relative to department size.	The top 22 departments were ranked. A strong correlation was found between editorial board memberships and more conventional measures of departmental productivity.	Gibbons and Fish (1991)

United States, 1971-85	To rank the 'research quality' of individual economists	Total number of citations received and the total adjusted for years since receipt of Ph.D.	The top 150 economists were ranked. Their distribution between universities was very uneven, with 40 per cent coming from five departments (Harvard, MIT, Chicago, Yale and Princeton.)	Medoff (1989)
United States, 1975-85	To rank economics departments according to measures derived from the 108 most cited journals.	Quality-adjusted standardised pages, i.e. *American Economic Review* - equivalent pages multiplied by quality weights derived from Liebowitz and Palmer's (1984) ranking of journals. Credit was attributed to the department where the staff member was employed in 1984, rather than to the affiliation mentioned in the article.	The top 152 departments were ranked. There was strong consistency, especially among the top 20, between this and earlier studies. Departments were also ranked by subject areas within the discipline.	Tschirhart (1989)
United States, 1981	To review the assessments of economics departments, and to compare the rankings achieved using different methods.	A wide range of citation measures and publication al and per capita terms.	Important differences occurred between rankings using citation measures as opposed to publication measures. Citation measures were suggested as superior.	Liebowitz and Palmer (1988)

excluded.[21] Each study is summarised with respect to its geographic and time period coverage, and its principal objectives, measures and results. The table leads us to consider a number of issues in the measurement of research productivity which need to be answered before any study can be tackled.

6.1 Which measure(s) should be used?

The earlier studies the economics discipline used counts of articles or pages contributed by members of a department as their basic measure. This evolved to standardised pages per member, usually based on *American Economic Review* - equivalent pages. The advent of compendia of citations, principally the annual *Social Sciences Citation Index*, added a range of citation measures. Versions of these two measures have been used by the majority of studies, but others have also been employed: the opinion of a sample of economists as to the research strength of different departments (i.e. peer rankings); the number of abstracts published by department members in a major abstracting source (e.g. the *Journal of Economic Literature*); the extent to which members of departments are members of the editorial boards of selected journals; and research grants received from outside the university. Such measures can be reported in aggregate terms or in per capita terms and they may be weighted by some indicator of quality or worth. Table 9 summarises the range of measures used, some 23 in all, broken into Publication Measures, Citation Measures and Other Measures. Under each, 'total' figures cover aggregate departmental output and provide a measure of overall department strength. Since big departments will probably be big producers of research simply because they have more researchers, small departments are likely to be disadvantaged by such a measure. In order to reflect average productivity, per capita measures of research output are used by most studies.

Given that its nature and implications are less intuitively obvious than some other measures, citation analysis is examined in detail in section 7. Of the 'other' measures, each has a general usefulness but also suffers from important limitations. Peer rankings are likely to be influenced by size and are therefore more useful as an indication of total research output rather than average productivity. More seriously, peer rankings may well be impaired by the 'halo effect', whereby past greatness continues to have an effect well after the objective basis for such a reputation has diminished, and by the composition of the reviewing team.

---

21 Many related disciplines have also been evaluated in respect of research productivity. Examples include regional science (Kau and Johnson 1983), U.S. business administration (e.g. Williams 1987; Niemi 1988), non-U.S. accounting (Reeve and Hutchinson 1988), U.K. accounting (Gray *et al* 1987; Gee and Gray 1989), wordwide statistical theory/econometrics (Hall 1987; Phillips *et al* 1988) and U.S. economic history (Niemi 1975b).

Table 9: Measures of research productivity

	Measure	Studies
Publication measures		
1.	Number of publications - total	Bell and Seater (1978); Laband (1986a); Rushton and Meltzer (1979)
2.	Number of publications (per capita)	Campbell and Campbell (1984); Laband (1985b), (1986a)
3., 4	Number of publications, total or per capita, with weights attached to different publications	West et al. (1980); Niemi (1975a)
5.	Number of pages - total	Shim (1982); Hogan (1984), (1986)
6.	Number of pages - per capita	House and Yeager (1978); Hogan (1986)
7, 8.	Number of pages, total or per capita, with weights attached to different publications	Anderson (1978); Tschirhart (1989)
9.	Number of standardized pages - total	Niemi (1975a); Graves et al. (1982); Hirsch et al. (1984); Laband (1985a, 1985b); Brar et al. (1987)
10.	Number of standardized pages - per capita	Graves et al. (1982); Hirsch et al. (1984); Laband (1985a, 1985b, 1986a); Brar et al. (1987)
11, 12	Number of standardized pages, total or per capita, with weights attached to different publications	Niemi (1975a)
Citation measures		
13.	Number of citations - total	Gerrity and McKenzie (1978); Davis and Papanek (1984); Laband (1985a, 1986a); Beilock et al. (1986); Beilock and Polopolus (1988); Brar et al. (1987)
14.	Number of citations - per capita	Gerrity and McKenzie (1978); Davis and Papanek (1984); Laband (1985a, 1985b, 1986a); Beilock et al. (1986); Brar et al. (1987)
15.	Number of citations - per article	Campbell and Campbell (1984); Laband (1985a, 1985b, 1986a)
16.	Number of citations, total or per capita, adjusted for example, for age of staff	Davis and Papanek (1984); Beilock et al. (1986); Medoff (1989)
17.	Proportion of staff cited	Beilock et al. (1986)
Other measures		
18.	Peer rankings	Bell and Seater (1978); Meador et al. (1992)
19, 20	Number of abstracts, total or per capita	Golden et al. (1986)
21, 22	Membership of editorial boards, total or per capita	Gibbons and Fish (1991)
23, 24	Research grants received from external sources, total or per capita	Brown and Nunn (1981); Bourke and Simondson (1982); Hancock (1983); Campbell and Campbell (1984)
25.	Percentage of own graduates placed in top departments	Laband (1985a)

Gillett (1989) distinguished between three types of peer review - journal peer review, grant-giver peer review and impressionistic peer review. He found that only the first - measured by number of publications per capita or cost per publication - provides a true performance indicator, in that it links outputs to inputs. He reported fifteen serious defects in grant-giver peer reviews and nine major flaws in impressionistic peer reviews.

The *Times Higher Education Supplement* has carried out peer reviews of subject areas since 1984, based on replies received from heads of departments to nine questions:

1. Which in your view are the five best departments in British higher education institutions in your subject bearing in mind mainly the output and quality of their research?
2. Which in your view are the five best departments in British higher education in your subject bearing in mind mainly the quality of their teaching of undergraduate students?
3. How much external funding in terms of research grants and industrial support did your department attract over the past three years?
4. In the last five years how many books and articles have been published by members of staff in your department?
5. How many members of academic staff does your department have at present?
6. How many did it have in 1979?
7. How many of the staff are in permanent posts?
8. Can you please give some indication of the average UCCA score of undergraduate entrants to your department? What is the highest and what is the lowest score obtained by a successful candidate in the past three years?
9. Excluding your present department, in which department would you like most to hold an academic post in Britain? And in the rest of the world?

UK psychology departments were reviewed in 1985 in the THES and Gillett (1989) compares the results of this with other performance measures. He found very small correlations between peer ranking and number of publications per capita ($r = 0.07$) and average citation impact ($r = 0.31$). By contrast, the correlation with size of department was 0.68, suggesting that size acts as a distorting influence on peer rankings. There was, incidentally, a very small correlation ($r = 0.11$) between size and publications per capita (see section 4.2.1). The relationship between departmental size and peer review ranking has been recognised in many studies (Cave *et al* 1988.85).

The foregoing is not to suggest that peer rankings are not obtained from careful research methods. A recent study (Meador *et al* 1992) uses a peer review conducted by the Conference Board of Associated Research Councils (1982). Each of 1254 departments across 23 academic disciplines, including 73 economics departments, were rated according to the 'scholarly competence and achievements' of their staff. The rankings, on a six point scale, provided by

not less than 20 per cent of staff in the disciplines being evaluated; they were not asked to consider their own department. There was a positive correlation ($r_s = 0.49$, $p < 0.05$) between the mean rating achieved by this method and the number of articles per staff member per annum over two or three years; for economics, the correlation was higher ($r = 0.64$, $p < 0.05$).

Research grants per capita are commonly used as a measure of performance, the assumption being that success in winning grants is a measure of individual or departmental quality. There are a number of problems. In considering peer review, it will be recalled, Gillett (1989) examined reviews by grant giving agencies. He found that there was 'an overwhelming case against the use of any of the three measures [per capita research income, number of research grants per capita and the directly voiced opinion of grant giving agencies] based on grant giver peer review' (Gillett 1989.32). His main criticism of the use of grant giver peer reviews, it should be emphasised, is that they provide 'a pure output measure' (1989.22). More seriously, however, the use of research grants as a measure appears to ignore the fact that funding is much more an input to research than an output. Obviously, well-funded departments can afford to hire research assistants and the like which should result in enahnced research output. The use of research grant income as a measure of output this involves double counting (i.e. for both the research income and its resulting output) and cannot be seriously justified. Perhaps surprisingly, there may be very weak relationships between research grant income and rates of publication and citation (e.g. Bentham (1987) for UK geography departments; Gillett (1989) for UK psychology departments; Cave *et al* 1988.96-99 for UK physics and psychology departments; and Tables 3 and 4). Most research in some disciplines is carried out without the need for funds. This is well illustrated in Singer's (1992) report of a chat between Bertrand Russell and his modern day head of department, concerning why Russell hasn't submitted an ARC grant application. Russell objects that he doesn't need money to fund his current research but the head isn't satisfied:

> HEAD: If you get a grant, it isn't just the *grant* money that the department gets. We get *extra* money for every grant we get; and it's money we can spend on whatever we like: promotions, a junior clerical assistant to help our overworked secretary, even conference travel. It's a kind of reward for being a productive department.
> RUSSELL: But I am productive! You just congratulated me on *Principles of Mathematics*. And I published *The Foundations of Geometry* only a few years before that. That's not so bad for a man of thirty, is it?
> HEAD: I know, Bertie, you're a hard-working fellow, and clever, too. But in Canberra they don't really care about publications. Grants, that's what they count. If you are getting large ARC grants you are productive, and if you are not, you aren't. It's as simple as that. (Singer 1992.15)

Finally, research funds are provided for certain research fields and not to others, according to the benefits perceived likely to result by the funding agency. These may be short term, private benefits and not long term social benefits.

The number of abstracts in a source such as the *Journal of Economic Literature* (Golden *et al* 1986) suffers from the subjectivity involved in the decision whether to abstract a particular article; in fact only about 20 per cent of the articles published in the journals included in the *Journal of Economic Literature* abstracts are abstracted. Membership of editorial boards (Gibbons and Fish 1991) probably suffers from the old-boy network which may not reflect current research productivity. Laband (1985c) found evidence that editors favoured articles, in terms of average length, from authors from their own institution, and McDowell and Amacher (1986) found that an in-house editor resulted in more publications from his/her department in that journal.[22]

6.2 Which publications should be included?

Table 8 indicates a wide variation in what is measured as research output. Typically, the Publication Measures include articles (which may or may not include 'notes', 'comments' and 'replies') in the n top journals, with n varying between four (Shim 1982; Hogan 1984) and 45 (House and Yeager 1978). Such an approach has several weaknesses. First, the selection of top journals may be subjective. Several studies (Graves *et al* 1982; Davis and Papanek 1984; Hirsch *et al* 1984) followed Niemi's (1975a) use of 24 top journals. Niemi (1975a.98) comments that 'a review of journal rankings in ... [four pre-1975] studies showed that the twenty-four journals ... were consistently rated among the top 30-35 economics periodicals', by assorted panels of economists. Somewhat more objective was Liebowitz and Palmer's (1984) ranking of journals based on citations in 1980, which included 21 of Niemi's 24 journals in their first 36.[23] Publications may be refereed articles in 'regional journals', the inclusion of which is a feature of Beilock and Polopolus' (1988) study of agricultural economics departments, or journals of other disciplines; they may be books or monographs which are important in some branches of the discipline, particularly economic history and the history of economic thought; and they may include chapters in books or monographs, papers in refereed conference proceedings and a miscellany of other vehicles in which research is 'published'.

---

22 The foregoing not exhaust the possibilities. For example, Davis and Astin (1987) use seven indices of reputation (admittedly not productivity) including 'self-reported visibility, total number of honours and awards received, whether or not the scholar was listed in *American Men and Women of Science*, ... total number of citations of the self-reported most important piece ..., and the total number of citations of the three most-cited pieces.' (1987.264)

23 It may be noted here that Liebowitz and Palmer chose their journals from those listed in the *Journal of Economic Literature* which includes, for reasons which are not spelled out, a number of non-economic journals. As a result, the *Yale Law Journal*, the *Journal of The Royal Society B*, the *Michigan Law Review* and *Demography* are ranked in Liebowitz and Palmer's most cited 20 journals.

Defining a refereed article is itself a matter of judgement. Miller and Punsalan (1988.ix) define a refereed journal as one having

> a structured reviewing system in which at least two reviewers, excluding in-house editors, evaluate each unsolicited manuscript and advise the editor as to acceptance or rejection.

On this basis, articles in a large number of highly reputed economics journals would be termed non-refereed. For example, 16 of Diamond's (1989a) list of 27 core economics journals are refereed and five are non-refereed (see Table 10): six were not included in Miller and Punsulan's listing of 242 economics journals indexed in the *Journal of Economic Literature* in 1986, for reasons of non-response to their survey.

Table 10: Diamond's 27 core journals in economics classified by Miller and Punsulan's criterion

Journals using two or more external reviewers	11
Journals using one or more external reviewers and one or more board reviewers	5
Non refereed journals[1]	5
Not included[2]	6

Notes: [1] *Journal of Financial Economics, Journal of Labour Economics, Journal of Mathematical Economics, Journal of Monetary Economics, Journal of Political Economy.*

[2] *Brookings Papers on Economic Activity, Economics Letters, Journal of Economic Literature, Journal of Economic Theory, Journal of International Economics, Journal of Law and Economics.*

The question is less relevant to studies using citation measures. Using compendia such as the SSCI, the total citations to the work of an individual in any year can be determined. In the case of the SSCI, citations are counted if they appear in the bibliographies/reference lists or footnotes of articles in selected source journals published in that year. The number of citations are not confined to articles published in these journals but to any item, including books, unpublished papers and theses, which appear in the source journals. We could note here that the figures of 5200 source journals given by Davis and Papanek (1983.225) and Beilock *et al* (1986.595) and 4300 by Liebowitz and Palmer (1984.79, 1988.83) are misleading.[24] In 1991, for example, there were 1432 'fully covered source journals', including 29 added in 1991, plus 1271 'selectively covered source journals'. The latter are journals from the SCI from which

24 There were in fact 1478 fully covered and 2936 selectively covered source journals in 1980, a total of 4414.

selected articles are included in the SSCI. Of the fully-covered journals, 144 are economics journals. Some 850 are of US origin and 18 derive from Australia. In 1991, 61,828 articles and 60,820 other sources (principally book reviews) gave 1,873,031 citations. This represented 1.5 citations per item cited and 4.4 citations per cited author. On average, articles had 17.9 references in their bibliography.

6.3 Who should be included as members of the department?

This question arises particularly with respect to aggregate research performance of the department. If we use departmental lists of publications, it is possible that the publications of people of varying degrees of association with the department are included e.g. postgraduate students, visiting lecturers, honorary research affiliates, etc. In one sense, of course, these are indeed part of the department's total research output. Given, however, that departments vary as to their endowment of such people, comparisons between departments may require a definition of staff who are to be counted. This may come from faculty lists (e.g. Bell and Seater 1978) or to a publication such as the *Guide to Graduate Study in Economics, Agricultural Economics and Related Fields* (Hogan 1984).

A related question is the basis used to attribute credit for a publication to a department, given the mobility of academic staff. Typically, attributions have been made on the basis of the institution listed on the article itself. This is presumably where the author was located at the time the article was accepted for publication, but may not be the place where the work was carried out or, indeed, the author's present location.

As Liebowitz and Palmer (1988.105) remark, 'When one is asked about the quality of an economics department, one usually reacts by asking in turn', "Who do they have there now?" not "What did they publish there last year?"[25] Consistent with this view that it is the present staff complement of a department which determines its prestige, some of the studies summarised in Table 3.8 (e.g. Bell and Seater 1978; Hogan 1984; Tschirhart 1989) use the author's most recent affiliation. Another problem with relying on the affiliation listed on the article is that the author may not be a member of the department under scrutiny, but of a related department (e.g. econometrics, economic history) or of a research centre. This has led to the Australian National University scoring particularly strongly in publication counts, given the existence of at least three research centres in economics, in addition to its teaching department.

Finally, it should be noted that some studies (e.g. Laband 1985a, 1985b; Hogan 1986) were specifically interested in the productivity of Ph.D. graduates from different departments.

---

25 This is echoed by Laband's (1986b) remark that what is said may be less important than who says it.

6.4 Which time period should be covered?

Typically, Publication Measures have covered a five year period although several, based on a small number of journals (e.g. Shim 1982; Hogan 1984), have covered substantially longer periods. With the benefit of their data being available in existing computerised data bases, citation analyses have typically covered around ten year periods. Selection of years will depend upon the precise research question being asked. Depending upon the mobility of academics, data referring to periods more than ten years old may be of little use in determining the current quality of an institution's research environment but may be useful if the research question emphasises the direction a department has been moving in as regards research productivity. The latter aspect was of particular concern to Laband (1985a).

The question is again less important for the citation measures. Of the studies listed in Table 8, the standard approach is to count all citations to the work of an individual made in any year (e.g. Gerrity and McKenzie 1978; Laband 1985a, 1985b, 1986). Some articles (e.g. Davis and Papanek 1984; Beilock *et al* 1986; Medoff 1989) have confined cited publications to those published in a given number of years, thus avoiding the bias involved by including citations to works published many years ago. This may be important where the average age of academics varies significantly between departments.

7. Citation analysis

The nature and implications of citation analysis is less intuitively obvious than some of other performance measures and we shall therefore discuss the technique in some detail at this point.

The use of citation analysis began with the establishment of the Institute for Scientific Information by Eugene Garfield in the early 1960s and in particular its publication of the Science Citation Index (SCI) and the Social Sciences Citation Index (SSCI). In this study, we concentrate on the latter. Created in 1966, the SSCI provide information on the number of times a publication is cited in any year, based on the citations contained in articles published 1432 (in 1991) fully covered source journals. The number of citations made to any individual author, in respect of any of his/her publications, can be counted for any year, and added with others to calculate the number of citations made to the members of a department in any year. Whilst the citing articles are limited to those published in source journals, the cited publications can be from any journal and any discipline and may also be books, theses, discussion papers and manuscripts in typescript.

The SSCI has also been used to evaluate journals according to their relative influence, as measured by the number of citations made to articles published in particular journals. For economics, this has been carried out by Liebowitz and Palmer (1984), who ranked some 107 journals according to the citations articles published in each journal received in 1980, and more recently by Diamond (1989a), who identified 27 'core economics journals' using SSCI citations in 1986.[26] This second use is of value to this study because it aids the decision as it what constitute the 'core journals' in economics. The top 122 economics journals ranked by impact factor are presented in the *Social Science Citation Index*, 1988, volume 6. Less relevant to our interests are studies of 'article popularity'. Laband (1986b) examined 5880 articles published between 1974 and 1976 in 40 major economics journals. Between 1977 and 1982, 84 per cent were cited between zero and nine times[27] and 0.3 per cent (17 articles) were cited more than 100 times. Citations received by an article were found to positively and significantly influenced by the author's reputation (based on previous citations), the length of the article and the relative ranking of the journal (based on average citations received per article by each journal). As a general remark, as many as 90 per cent of journal articles are never cited and 'probably no more than 50 per cent are ever read by more than their author and the referees' (Collins 1992.15).

The SSCI lists authors of source articles by institution. It might be thought that these could simply be added to provide a measure of total departmental output, but several SSCI procedures, some unavoidable, make this hazardous. First, citing authors are allocated to their departments only if this is recorded in the article; otherwise they are allocated to their university. In order to avoid problems of including non-departmental members, the SSCI lists would have to be checked against lists of staff members. Second, the SSCI lists only first authors and ignores any joint authors.

A number of criticisms have been levelled at the use of citations as a measure of research productivity.

1. How can the impact of an individual with a few citations to each of many publications be compared to that of an individual with the same number of total citations, but comprising many citations to a few publications? There is a presumption that the latter has more impact in that the fewer articles are of greater significance, and some writers e.g. Cole and Cole (1967) have suggested using the author's three most cited publications as a measure of his/her impact. The presumption is by no means undisputed, and both total citations and citations per article are commonly counted. Cole and Cole (1971.26)

---

26 A related study (Christenson and Sigelman 1985) compared peer rankings of sociolgoy and political science journals, in terms of the average importance of their contributions to the field' to the SSCI 'impact factor' scores i.e. the average number of citations received by articles in a journal, adjusted for the number of articles published in each journal per annum. Finding correlations of 0.53 to 0.57, the authors conclude (1985.967) that a journal's prestige may endure in spite of what it merits.

27 Unfortunately, Laband does not report the number of zero-cited articles separately.

examine whether the sheer number of papers results in more citations. They concluded that they do not,[28] and that therefore total citations were a good indication of significance. This conclusion is supported by later work (e.g. Garfield 1981, Cole and Zuckerman 1984).

2. How can the citation performance of individuals be compared when the citations to one are concentrated within the last few years, whereas those to another refer to publications written many years ago. It could be argued, as do Moed *et al* (1985), that the latter person's work has proved its durability and its contribution to basic knowledge. On the other hand, citations to recent work indicate that the researcher is still active.
3. Citations may be commendatory, critical, trivial or neutral, but are not distinguished as such by the SSCI. The presumption is that a commendatory citation is more valuable than a critical one (assuming, presumably, that the criticism is valid). However, an article containing error or omission still needs to be worth criticising, rather than disdainfully ignored and is valuable to that extent. There is also the potential benefit of the 'fruitful error' (Cole and Cole 1971.25) which leads on to insightful further research. We should not overemphasise the importance of critical citations. Stigler and Friedland's (1975.488-489) examination of 5581 citations[29] to 700 economics articles found that 648 were favourable and 566 were unfavourable with the vast majority (78 per cent) being neutral.
4. Citations may be given as a matter of course to literature reviews which of themselves add little to knowledge although they may perform a valuable role in delineating the state of knowledge. On the other hand, citations may not be given at all when an innovator's work becomes common knowledge. Liebowitz and Palmer (1988.94) doubt that any more than a handful of publications are treated in this way.
5. The treatment of all citations as of equal worth has been questioned, and the suggestion made that greater weight be given to citations made by eminent people (e.g. Cole and Zuckerman 1984.242). Cole and Cole (1971.25), however, suggest that similar results would occur from such a method as from the present practice of counting all citations as being of equal worth.
6. Fields within a discipline vary in size as measured by the number of publications, number of researchers involved, the number of journal outlets and the coverage of relevant journals in the SSCI. Within disciplines and within fields, there will be shifts over time, with some parts expanding and others contracting. As a result, some individuals will be much less cited than others which may be no reflection of the quantity and quality of their research, and only a partical reflection of its impact (Moed *et al* 1985). A related aspect is

---

28 They did, however, find that highly cited scientists with a modest total number of publications ('perfectionists') tended to have received higher recognition than those with the same number of citations but more publications ('prolific scientists').

29 Stigler and Friedland's (1975) method, it should be noted, allows for multiple citations. That is, they included each citation to the cited publication in a citing publication, up to a maximum of one per paragraph.

that citation practices vary significantly between fields (Liebowitz and Palmer 1984; Kroc 1984): some emphasise journal articles whilst others (e.g. economic history) attach more weight to books. Within any field, theoretical and methodological works are likely to be cited most frequently (Garfield 1979).

7. The significance of some research is not always recognised by contemporaries and may only occur after a long lag. Alternatively, some undeserving research might be received many favourable citations for other reasons (see criticism 8 below). By and large, Cole and Cole (1971.24-25) argue, there are probably few significant publications which go unrecognised and, by implication, few undeserving articles receive much praise.
8. Self citations can inflate individual and departmental citation scores and many studies omit them. It can be contended, however, that many self citations truly represent the influence that the previous work of an author has had on his present work (Liebowitz and Palmer 1988.94). The study of agricultural economics departments by Beilock and Polopoulos (1988.406) found that 20 per cent of citations were self citations and that the difference made to departmental ratings by their inclusion or exclusion was 'modest'. It should be noted that because the SSCI lists articles by their first authors only, their figure is an underestimate i.e. the article may cite second and subsequent authors. Liebowiz and Palmer (1984) discuss 'gratuitous citations' (e.g. to the journal to which the article is submitted, or to its editor) but conclude (1988) that such practices exert little influence on rankings.
9. To the extent that citations are used for promotion or for the allocation of funds to individuals and departments, there will be a tendency for 'citation inflation' to occur (Liebowitz and Palmer 1984).
10. Cited articles have not always been read by the citing authors (MacRoberts and MacRoberts 1986).
11. Authors outside North America are not likely to be widely cited, principally because their publications are concentrated in journals which are not widely cited in the SSCI source journals i.e. their articles, wherever published, are not likely to be cited by articles in the source journals. The background and orientation of non-North American authors will also influence citation results in ways which could mislead an unwary user. Those who have studied and researched in the United States are more likely to have begun and to continue publishing in journals which tend to be more frequently cited in source journals than those whose experience has been concentrated elsewhere. A similar argument could be developed concerning the topics upon which different individuals research: some are much more amenable than others to publication and subsequent citation in the United States.

    Despite the above, it has recently been forcefully asserted (Bairam 1990) that Australian and New Zealand economics departments should be assessed using citation analysis. There is some validity in this argument, provided it is remembered precisely what

citations based on the SSCI are measuring i.e. the referencing of an article in (very largely) North American source journals. It is quite conceivable that an article published in an Australian journal might have a significant impact of some other kind (e.g. on economic policy formation) but never be cited in the SSCI. Citations, then, measure only one kind of impact.

12. Some of the procedures used to generate the SSCI operate to limit its usefulness. We have mentioned that the SSCI lists articles by the first author only; this will affect individual rankings but will not affect departmental rankings so much if most joint authorships occur, as Leibowitz and Palmer (1988.95) assert, within departments. The largest conceivable adjustments as a result of co-author bias will cause little change in rankings (1988.97). Another aspect of this is that it is sometimes difficult to accurately identify first authors, given that first names and initials are used. Second, Moed *et al* (1985) document the extent of errors found in the companion Science Citation Index, suggesting that around 10 per cent of total citations are missed (1985.139-140). Whilst they agree that no systematic error is likely, they note that the omission of a few highly cited publications can be serious for departments. Third, not all relevant journals are used as source journals by the SSCI and only journals are used as sources. Some citations go unrecorded as a result. Fourth, whilst the SSCI does list source journal articles by the institutional affiliation of the author, this is not sufficiently precise to allow inter-university comparisons of the performance of, for example, departments of economics: authors may come from many departments other than economics and may not be 'members' of a department.

Earlier studies using citations, it may be noted, referred to their ability to measure the 'quality' of a researcher's work. Later studies have been careful to distinguish quality, which it may be 'virtually impossible to operationalize' (Moed *et al* 1985.134), from impact. Moed *et al* (1985.133ff) describe the 'research front' of a particular field as consisting of the publications in that field over a period of time. Citations, then, are an indication of an article's impact on the research front. They may occur in the first few years after publication or they may be more durable and make a more permanent contribution to basic knowledge in the field.

Despite this litany of criticisms and limitations, citation analysis is widely accepted as a legitimate and major tool in the measurement of research productivity with particular value as a measure of the impact of individual publications, researchers and departments. Citation counts are reasonably strongly correlated with other measures, typically between 0.60 and 0.75 (Cole and Zuckerman 1984.231). A U.S. study of 148 full professors of economics found that an additional citation added more to their salaries than an additional article or book (Hamermesh *et al* 1982). Quandt's (1976.741) assertation that citations 'permit tentative predictions as to who future [Nobel] prizewinners will be' has been confirmed by Sauer (1988.858). Of Quandt's 26

most cited economists in 1970, eight have subsequently been awarded the Nobel Prize, along with three who had received it prior to 1970. Ten of the studies listed in Table 4.1 use citation analysis either solely or in conjunction with other measures. Liebowitz and Palmer (1988.94) are typical of those who, whilst accepting at least some of the criticisms levelled at citation analysis, accept its validity as a proxy measure of research output and influence.

> *We do not suggest that citations can or should be used as the sole measure of a paper's quality independent of a serious reading.* We would suggest, however, that it often would be illuminating for those who consider a paper with few citations to be of high calibre to ask why it has not generated more citations ... Publication tabulations may be correlated with reputation, they may be useful indicators of the direction in which the reputation is going, and they may be useful for other reasons, but they do not provide as close a measure as citation tabulations do of reputation or impact on the profession.
> (Leibowitz and Palmer 1988.94, 105, emphasis as in the original.)

8. Measuring research productivity: an overview

We can sum up the methodologies employed, and results obtained from this set of studies as follows, bearing in mind that U.S. institutions are much more at home with bibliometric indicators, given that central government funding allocations are not relevant.

First, whereas most of the earlier studies used Publication Measures, more recent studies have used a range of Publication and Citation Measures. The latter measures are regarded by some authors as a better indicator of the quality of an economics department.

> It is bigger news to the profession when someone moves from one employer to another than when that person has an article accepted or rejected for publication. The reason is that the perceived quality of an economics department is based on the reputations of the members of that department, and these reputations are based, in turn, on the cumulative impact that the individual member's writings have had on the profession. Publication tabulations may be correlated with reputation, they may be useful indicators of the direction in which the reputation is going, and they may be useful for other reasons, but they do not provide as close a measure as citation tabulations do of reputation or impact on the profession. Liebowitz and Palmer (1988.105)

This is based on the view, it should be noted, that it is who is in the department at any time, rather than its actual output, which determines its research standing.

Second, irrespective of the measures used, there is a strong similarity of rankings of the top say, twenty, U.S. departments (Tschirhart 1989; Gibbons and Fish 1991). Laband (1985a) used twelve separate measures to produce a composite index to rank 50 departments for publications between 1971 and 1983. He found a 'basically similar' ranking to those of Graves *et al* (1982), but did point out some significant movements within the rankings. These he

attributed to his measures providing 'independently useful information vis-à-vis the relative quality of faculty teaching and research across departments' (Laband 1985a.237). A re-study of Laband's data by Brar *et al* (1987), with the aim of determining whether different measures yielded significantly different rankings, concluded that they did not. In their view, simple measures were adequate. This is a view not entirely shared by Liebowitz and Palmer's (1988) article entitled 'Assessing assessments of economics departments' which concludes that use of citations leads to different results. My assessment is that they overstate their case: the rank correlations they report between measures (1988.97, 101) are very high.

Third, there is a very high degree of variation in research output within and between departments, which several examples will make clear. In Davis and Papanek's (1984) study, the top 20 departments had almost seven times the total number of citations than did the other 102 departments. For citations per member of staff, the ratio was a little over six to one. In Graves *et al* (1982), the top 20 departments had more than twice the number of AER-equivalent pages in the top 24 journals than the remaining 220 departments. In per capita terms, the ratio was 8.3 to one. Medoff's (1989) study of individual economists shows that, on the basis of citation measures, 24 of the top 25 economists are at 12 universities and that seven of the top ten economists are at five universities. Several studies have calculated concentration ratios, being the ratio of the number of pages published by the top n institutions to the total number of pages published. Hirsch *et al* (1984) found, for example, that the top ten of their 273 institutions produced 30.8 per cent of total pages and that the top 25 produced 55.4 per cent.

REFERENCES

Allen, M., 1988, *The goals of universities*, Milton Keynes, Society for Research into Higher Education/Open University Press.

Australian Research Council (ARC), 1993, *Reviews of grant outcomes No. 1. Economics 1986-90,* Canberra, AGPS.

Australian Vice Chancellors Committee (AVCC) / Australian Council of Directors and Principals (ACDP), 1988, *Report of the AVCC/ACDP working party on performance indicators*, Canberra, AGPS.

Bairam, E., 1990, 'Publication rates and research productivity in New Zealand university economics departments, 1984-88', University of Otago, *Economics Discussion Papers*, No. 9009.

Baldwin, P., 1991, *Higher education: quality and diversity in the 1990s*, Canberra, AGPS.

Beilock, R.P., Polopolus, L.C. and Correal, M., 1986, 'Rankings of agricultural economics departments by citations', *American Journal of Agricultural Economics*, 68(3), 595-604.

Beilock, R.P. and Polopolus, L.C., 1988, 'Ranking of agricultural economics departments: influence of regional journals, joint authorship and self-citations', *American Journal of Agricultural Economics*, 70(4), 403-409.

Bell, J.G. and Seater, J.J., 1978, 'Publishing performance: departmental and individual', *Economic Inquiry*, 16(4), 599-615.

Bentham, G., 1987, 'An evaluation of the UGC's ratings of the research of British university geography departments, *Area*, 19(2), 147-154.

Bourke, P., 1986, *Quality measures in higher eduction*, Canberra, Commonwealth Tertiary Education Commission.

Bourke, P., 1991, 'A bibliometric profile of research in the Institute of Advanced Studies, ANU, 1976-88', in Linke (1991, vol. 2.86-124).

Brar, J.S., Nazemzadeh, A., and Chow, P., (1987), 'Ranking top economics departments', *Atlantic Economic Journal* XV(1), 126.

Cave, M., Hanney, S., Kogan, M., and Trevett, G., 1988, *The use of performance indicators in higher education: a critical analysis of developing practice*, London, Jessica Kingsley.

Cave, M., Kogan, M., and Hanney, S., 1989, 'Performance measurement in higher education', *Public Money and Management*, 9(1), 11-16.

Christenson, J. and Sigelman, L., 1985, 'Accrediting knowledge: journal stature and citation impact in social science', *Social Science Quarterly*, 66(4), 32-49.

Clayton, K., 1987, *The measurement of research expenditure in higher education*, School of environmental sciences, University of East Anglia.

Colander, D., 1989, 'Research and the economics profession', *Journal of Economic Perspectives*, 3(4), 137-148.

Cole, S. and Cole, J.R., 1967, 'Scientific output and recognition: a study in the operation of the reward system in science', *American Sociological Review*, 32, 377-390.

Cole, J.R. and Zuckerman, H., 1984, 'The productivity puzzle: persistence and change in patterns of publication of men and women scientists', in Steinkamp and Maehr (1984.217-258).

Collins, H., 1992, 'The good, bad and unquoted', *The Times Higher Education Supplement*, January 10, 15.

Committee of Vice-Chancellors and Principals (CVCP), 1985, *Report of the steering committee for efficiency studies in universities*, (the Jarratt Report), London, CVCP.

Commonwealth Tertiary Education Commission (CTEC), 1986, *Review of efficiency and effectiveness in higher education*, Canberra, AGPS.

Connell, W.F., 1991, 'Reviewing the quality of higher education', *The Australian Universities' Review*, 34(2), 36-41.

Cuenin, S., 1987, 'The use of performance indicators: an international survey', *International Journal of Institutional Management in Higher Education*, 11(2), 117-139.

Davis, D.E. and Astin, H.S., 1987, 'Reputational standing in academe', *Journal of Higher Education*, 58(3), 261-275.

Davis, P. and Papanek, G.F., 1984, 'Faculty ratings of major economics departments by citations', *American Economic Review*, 74, 225-230.

Dawkins, J.S., 1987, *Higher education: a policy discussion paper*, Canberra, AGPS.

Dawkins, J.S., 1988, *Higher education: a policy statement*, Canberra, AGPS.

Department of Employment, Education and Training (DEET), 1988, *Review of the discipline of engineering*, 3 volumes, Canberra, AGPS.

DEET, 1990, *Accounting in higher education. Report of the review of the accounting discipline in higher education*, 3 volumes, Canberra, AGPS.

Department of Education and Science (DES), 1985, *The development of higher education into the 1990s*, Cmnd 9524, London, HMSO.

Department of Education and Science (DES), 1987, *Higher education: meeting the challenge*, Cnd 14, London, HMSO.

Diamond, A.M., 1989a, 'The core journals of economics', *Current Contents*, 21(1), 4-11.

Diamond, A.M., 1989b, 'Most-cited economics papers and current research fronts', *Current Contents*, 21(2), 3-8.

Flemming, J., 1991, 'The use of assessments of university teaching, and especially research, for the allocation of resources: a personal view', *European Economic Review*, 35(4), 612-618.

Garfield, E., 1981, 'The 1,000 contemporary scientists most cited 1965-78, Part I: the basic list and introductions', *Current Contents*, 13, 5-14.

Gee, K.P. and Gray, R.H., 1989, 'Consistency and stability of UK academic publication output criteria in accounting', *The British Accounting Review*, 21(1), 43-54.

Gerrity, D.M. and McKenzie, R.B., 1978, 'The ranking of southern economics departments: new criterion and further evidence', *Southern Economic Journal*, 45(2), 608-614.

Gibbons, J.D. and Fish, M., 1991, 'Rankings of economics faculties and representation on editorial boards of top journals, *Journal of Economic Education*, 22(4), 361-371.

Gillett, R., 1987, 'Serious anomalies in the UGC comparative evaluation of the research performance of psychology departments', *Bulletin of the British Psychological Society*, 40(2), 42-49.

Gillett, R., 1989, 'Research performance indicators based on peer review: a critical analysis', *Higher Education Quarterly*, 43(1), 20-38.

Golden, J., Carstensen, F.V., Weiner, P., and Kane, S., 1986, 'Publication performance of fifty top economics departments: a per capita analysis', *Economics of Education Review*, 5(1), 83-86.

Graves, P.E., Marchand, J.R. and Thompson, R., 1982, 'Economics departmental rankings: research incentives, constraints and efficiency', *American Economic Review*, 72(5), 1131-1141.

Gray, R.H., Haslam, J. and Prodham, B.K., 1987, 'Academic departments of accounting in the UK: a note on publication output', *British Accounting Review*, 19(1), 53-71.

Grigg, L. and Sheehan, P., 1989, *Evaluating research: the role of performance indicators*, Office of the Academic Director of Research, University of Queensland.

Hall, A.D., 1987, 'Worldwide rankings of research activity in econometrics: 1980-1985', *Econometric Theory*, 3(2), 171-194.

Hamermesh, D.S., Johnson, G.E. and Weisbrod, B.A., 1982, 'Scholarship, citations and salaries: economic rewards in economics', *Southern Economic Journal*, 49(3), 472-481.

Hattie, J., 1990, 'Performance indicators in education', *Australian Journal of Education*, 34(3), 249-276.

Harris, G.T., 1993, *Research performance in Australian university economics departments*, Ph.D. thesis, University of New England.

Hirsch, B.T., Austin, R., Brooks, J. and Moore, J.B., 1984, 'Economics departmental rankings: comment', *American Economic Review*, 74(4), 822-826.

Hogan, T.D., 1984, 'Economics departmental rankings: comment', *American Economic Review*, 74(4), 827-833.

Hogan, T.D., 1986, 'The publishing performance of U.S. Ph.D. programs in economics during the 1970s, *Journal of Human Resources*, 21(2), 216-229.

House, D.R. and Yeager, J.H., 1978, 'The distribution of publication success within and among top economics departments: a disaggregative view of recent evidence', *Economic Inquiry*, XVI(3), 593-598.

Johnes, G., 1988, 'Determinants of research output in economics departments in British universities', *Research Policy*, 17(2), 171-178.

Johnes, G., 1990, 'Measures of research output: university departments of economics in the UK, 1984-8', *Economic Journal*, 100(2), 556-560.

Johnes, G., 1992, 'Performance indicators in higher education: a survey of recent work', *Oxford Review of Economic Policy*, 8(2), 19-34.

Johnes, J. and Taylor, J., 1987, 'Degree quality: an investigation into differences between UK universities', *Higher Education*, 16(4), 581-602.

Johnes, J. and Taylor, J., 1989a, 'Undergraduate non-completion rates: differences between U.K. universities', *Higher Education*, 18(3), 209-225.

Johnes, J. and Taylor, J., 1989b, 'The first destination of new graduates: comparisons between universities', *Applied Economics*, 21(7), 357-373.

Johnes, J. and Taylor, J., 1990a, 'The research performance of UK universities: a statistical analysis of the results of the 1989 research exercise', University of Lancaster, Department of Economics, *Discussion Paper EC 25/90.*

Johnes, J. and Taylor, J., 1990b, *Performance indicators in higher education: UK universities*, Buckingham, Society for Research into Higher Education, Open University Press.

Kau, J.B. and Johnson, L.L., 1983, 'Regional science programs: a ranking based on publication performance', *Journal of Regional Science*, 23(2), 177-186.

Keating, M., 1993, 'View from the top', *Directions in Government* 7(11), 22-23.

Kroc, R.J., 1984, 'Using citation analysis to assess scholarly activity', *Educational Researcher*, June/July, 17-22.

Laband, D.N., 1985a, 'An evaluation of 50 "ranked" economics departments - by quantity and quality of faculty publications and graduate student placement and research success', *Southern Economic Journal* 52(2), 216-240.

Laband, D.N., 1985b, 'A ranking of the top Canadian economics departments by research productivity of graduates', *Canadian Journal of Economics*, XVIII(4), 904-907.

Laband, D.N., 1985c, 'Publishing favoritism: a critique of department rankings based on quantitative publishing performance', *Southern Economic Journal*, 52(3), 510-515.

Laband, D.N., 1986a, 'A ranking of the top U.S. economics departments by research productivity of graduates', *Journal of Economic Education*, 17(1), 70-76.

Laband, D.N., 1986b, 'Article popularity', *Economic Inquiry*, XXIV(1), 173-180.

Liebowitz, S.J. and Palmer, J.P., 1984, 'Assessing the relative impacts of economics journals', *Journal of Economic Literature*, XXII(1), 77-88.

Liebowitz, S.J. and Palmer, J.P., 1988, 'Assessing assessment of economic departments', *Quarterly Review of Economics and Business*, 28(2), 88-113.

Linke, R., 1984, *Report of a study group on the measurement of quality and efficiency in Australian higher education*, Canberra, Commonwealth Tertiary Education Commission.

Linke, R., 1991, *Performance indicators in higher education: report of a trial evaluation study commissioned by the Commonwealth Department of Employment, Education and Training*, 2 volumes, Canberra, AGPS.

MacRoberts, M.H. and MacRoberts, B.R., 1986, 'Quantitative measures of communication in science: a study of the formal level', *Social Studies of Science*, 18, 151-172.

McDowell, J.M. and Amacher, R.C., 1986, 'Economic value of an in-house editorship', *Public Choice*, 48(2), 101-112.

Meador, M., Walters, S.J.K. and Jordan, J.M., 1992, 'Academic research productivity: reply, still further results', *Economics of Education Review*, 11(2), 161-167.

Medoff, M.H., 1989, 'The ranking of economists', *Journal of Economic Education*, 20(4), 405-415.

Miller, A.C. and Punsalan, V.J., 1988, *Refereed and non-refereed economics journals: a guide to publishing opportunities*, New York, Greenwood Press.

Moed, H.F., Burger, W.J.M., Frankfort, J.G. and Van Laan, A.F.J., 1985, 'The use of bibliometric data for the measurement of university research performance', *Research Policy*, 14(2), 131-149.

Moses, I., 1990, 'Teaching research and scholarship in different disciplines', *Higher Education*, 19(4), 351-375.

Niemi, A.W., 1975b, 'Journal publication performance of economic history programs: 1960-1969 and 1970-1974', *Journal of Economic History*, 35(3), 635-641.

Niemi, A.W., 1988, 'Research productivity of American business schools, 1975-1985', *Review of Business and Economic Research*, XXIII(2), 1-17.

Pettit, A.N. and Low Choy, S., 1992, *ARC performance indicators survey*, Statistical Consulting Unit, Queensland University of Technology.

Phillips, P.C.B., Choi, I. and Schochet, P.Z., 1988, 'Worldwide institutional and individual rankings in statistical theory by journal publications over the period 1980-1986', *Econometric Theory*, 4(1), 1-34.

Platt, J., 1988, 'Research policy in British higher education and its sociological assumptions', *Sociology*, 22(4), 513-529.

Pollitt, C., 1990, 'Measuring university performance: never mind the quality, never mind the width?', *Higher Education Quarterly*, 44(1), 60-81.

Quandt, R.E., 1976, 'Some quantitative aspects of the economic journal literature', *Journal of Political Economy*, 84(4), 741-755.

Reeve, R.C. and Hutchinson, P.J., 1988, 'The contribution of non-U.S. institutions to academic accounting journals', *Abacus*, 24(1), 90-94.

Roe, E. and Moses, I., 1986, *Departmental reviews in higher education institutions*, Canberra, CTEC/AGPS.

Roe, E., McDonald, R., and Moses, I., 1986, *Reviewing academic performance: approaches to the evaluation of departments and individuals*, University of Queensland, St Lucia.

Shim, K.R., 1982, 'Institutional performance in the publishing of economic papers during 1975-1979', *Economics of Education Review*, 2(3), 295-300.

Singer, P., 1992, 'Not what you produce, but how much you spend', *Modern Times*, March, 15-16.

Sizer, J., 1988, 'In search of excellence: performance assessment in the United Kingdom', *Higher Education Quarterly*, 42(2), 152-161.

Sizer, J., 1992, 'Performance indicators in government - higher education institutions relationships: lessons for government', *Higher Education Management*, 4(2), 156-163.

Smith, T., 1987, 'The UGC's research rankings exercise', *Higher Education Quarterly*, 41(4), 303-316.

Smith, D., 1988, 'On academic performance', *Area*, 20(1), 3-13.

Stigler, G.J. and Friedland, C., 1975, 'The citation practices of doctorates in economics', *Journal of Political Economy*, 83(3), 477-507.

Stolte-Heiskanen, V., 1992, 'Research performance evaluation in the higher education sector: a grass-roots perspective', *Higher Education Management*, 4(2), 179-193.

Teather, D.C.B., 1990, 'Performance indicators in Australian higher education: the context and an appraisal of the 1988 report', in F.J.R.C. Douchy *et al*, *Management information and performance indicators in higher education*, Van Gorcum, Assen/Maastricht, 103-118.

Tschirhart, J., 1989, 'Ranking economics departments in areas of expertise', *Journal of Economics Education*, 20(2), 199-222.

Weller, P., Gardner, M., Ryan, N., and Stevens, B., 1993, 'The role of the public sector: implications for the Australian public service', *Canberra Bulletin of Public Administration* No. 72, 1-23.

Williams, W.W., 1987, 'Institutional propensities to publish in academic journals of business administration', *Quarterly Review of Economics and Business*, 27(2), 12-17.