PUBLICATION RATES AND RESEARCH PRODUCTIVITY

Leo H T West;
Terry Hore;
Pak K Boon

Higher Education Advisory and Research Unit,
Monash University

Introduction

The higher education system in Australia is under considerable financial pressure, and this pressure seems likely to continue for some time. The universities, as part of the higher education system, have begun to emphasise their unique role of research as an area that should receive special consideration. Those of us inside the universities see the strong case that can be made for this. Unfortunately this situation does not apply universally outside the universities (for example, some sections of the press have concentrated their attacks specifically on the universities).

Given these forces, it seems very likely that the research activity of universities will come under closer scrutiny; that universities will need to demonstrate the effectiveness of their research. This will not be simply a systemic matter. As Klein[1] has pointed out, one effect of cost cutting in the universities will be increased competition at all levels of the system, between universities, between faculties, between departments, between individuals. Each will be called upon to demonstrate the effectiveness of their research activities.

It is no surprise that evaluations of research are beginning to appear in the international literature. The same financial pressures that are occurring in the Australian higher education system have already occurred in the USA and UK. The Ladd-Lipset[2] survey of 4400 faculty members in 161 colleges reported output of university staff as a body. Various studies have investigated the research activities of specific disciplines, for example, schools of education in the USA (Guba and Clark[3]), departments of engineering in the USA (Liu[4]) and departments of psychology in Canada (Schaeffer and Sulyma[5]).

A substantial part of this movement towards evaluating research activity has involved **quantitative measures of research.** Will the same thing happen in Australia? If so, how valid are such measures?

We have approached this question from two directions. We have conducted a theoretical analysis of quantitative measures of research activity within a framework of measurement in social science in general (including a brief review of literature specific to measuring research activity). In addition we have developed and used a quantitative measure of research activity — indeed we have tried to develop the best measure given the information available and our perception of the validity problems.

In this paper we describe the latter.

Thus, as an exercise we set ourselves an imaginary brief. We imagined that we had been asked to develop a quantitative measure that could be used in our own institution to evaluate the research activity of individuals, of departments, and of faculties.

We want to emphasise that our intention was not to conduct an evaluation of research *per se* but to examine the validity of such evaluations.

The development of a quantitative measure of research activity

What is measurable?

The construct "research activity" is not itself measurable. Our first step was to examine the range of measurable quantities that were available and that were applicable as indices of research activities. The available measurables considered were:

- Number of papers, books, etc., published.
- Number of papers presented at professional meetings.
- Number of citations to published work.
- Research funding received — amount and number.
- Peer evaluation of research and publication.
- Number of Ph.D. and Masters dissertations supervised.
- Number of invited papers, guest lectures.
- Offices held in professional associations.

In considering each of these, three matters were important:

(a) what is their relationship with the overall construct "research activity",
(b) what data source is available, and
(c) does the measurable quantity apply to all academic staff.

Summary comments on these matters are included in Table 1.

Table 1

Measurable	Relationship with the construct "research activity"‡	Availability of data sources	Applicable to all Academic Staff?
Papers, books published	Measure of productivity, overlayed with some elements of quality	Monash *Research Report** gives full listing for each year	Yes
Papers presented professional meeting	Measure of productivity, fewer elements of quality (usually not refereed)	No compiled list, would need to contact all staff	Yes
Citations	Measure of contribution of the researcher to his discipline	Science citations, social science citations indices etc.	Does not cover all disciplines (at least not evenly)
Research funding received	Measure of quality and productivity of research compounded with priorities of funding body, discipline area and grantmanship	Available from administration records	Yes
Number of Ph.D., Masters dissertations supervised	Measure of eminence and leadership compounded between disciplines by different post-graduate participation rates	Available from administration records	Yes
Invited papers, guest lectures	Measure of eminence and exposure possibly compounded by excellence as a lecturer	No compiled list, would need to contact all staff	Yes
Offices held on professional associations	Measure of eminence, compounded by administrative ability	No compiled list, would need to contact all staff	Yes
Peer evaluation (in same discipline)	Measure of quality of research with problems of "cronyism"	No formal data source. A formal procedure for collecting would be needed	Yes

* **Research Report** is an annual publication of Monash research publications and activities.
‡ A much more detailed examination of this relationship was undertaken in a separate analysis (see page 2).

It was clear that any measure that was not applicable to all staff was unacceptable. For this exercise, it was not possible to collect data specifically. These two criteria reduced the potential measurable quantities to publications, research funding and dissertations supervised.

Research funding appears to be used in some non-formal evaluations as a measure of research activity (grants received are invariably quoted in *curriculum vitae*, in institution publications, and so on). There are at least two threats to construct validity in the use of research funding as an index: the part played by what the Americans call "grantmanship", and the different patterns of grant giving in different fields.

Most professional academic organisations in the USA provide training courses in grant-getting techniques. This indicates that there are skills that can be learned that aid in gaining grants, skills that are not necessarily research skills. An important aspect in gaining grants is the field of research. Grant giving bodies have their own priorities, so that some fields attract more grants for reasons such as perceived social worth, etc., that are not related to the research itself. Liebert[6] calculated the average number of grants per scholar in twenty-seven fields in the USA. Agriculture received 1.636 grants per scholar, chemistry 1.289 and medicine 1.117 while, at the other end of the scale, foreign languages received 0.155, physical education 0.133 and nursing 0.120. Such large differences could hardly be due to differences in research productivity or quality between the disciplines.

Size of grant is also a function of field. Much of the research in medicine requires expensive equipment, while research of similar quality in history may require only a few dollars for photocopying. Hence a "dollars count" can bias the fundings towards some fields.

Research funding was rejected as a measurable for these reasons.

Numbers of dissertations supervised have serious threats to validity, too. There are substantial differences between the post-graduate participation rates in different disciplines that are not related to research quality or ability of the researcher or department. In part these rates are determined by the job market, in part by conventions. For example, in Australia Ph.D.'s in medicine are rare in comparison to chemistry. For a cross-disciplinary measure, these differences would be very significant, and this measure was rejected for this reason.

Number of publications also has problems as a cross disciplinary measure, but was chosen on the most appropriate and most available measurable quantity. The index developed called **publication rate index** (PRI) is a measure of research productivity, or perhaps visibility. However, it is influenced by a range of factors other than research productivity, including

personality differences — for example some individuals "rush into print", others delay until all the "loose ends" have been tied up.

The PRI is not necessarily a measure of research quality, although there are elements of this due to refereeing of journals and books. Nor is it necessarily a measure of contribution to the discipline. A single paper may have a greater contribution than one hundred other papers.

These matters need to be kept in mind when the results for an individual, a department, etc., are considered.

An important additional limitation to the validity of the index as a measure of research activity is the data base. The Monash **Research Report** is a self selected source. Staff provide details of their publications through their departments. The compilers set down guide-lines as to what is acceptable, but these are interpreted individually. Overlying the compilation are motives of visibility, of status, etc., and these are not consistent across the university. The most casual inspection of the **Research Report** will reveal many variations in type of publication interpreted as acceptable.

The formularisation adopted

Despite the increasing number of quantitative indices of publication rate that are appearing, there is no consensus about the weightings to be used in the operational formula. Issues that needed to be resolved were: the relative weights to be given to books, edited books, articles, etc.; whether differential weightings should be given to multiple author publications; whether journal prestige should be included.

On the relative weightings to be given to books and articles, Meltzer[7] gave a score of one to an article and a score of 18 to a book for single authors. For multiple author books, but not articles, the score was divided by the number of authors. Crane[8] used a book to article relativity of four to one, Cartter[9] weighted theoretical or research books as equivalent to six articles, text books as three articles and an edited collection as two articles. We decided to allocate a 5 : 3 : 1 ratio for authored books, edited books and articles. This decision, while based on our reading and some informal discussions with colleagues, is completely arbitrary.

Journal prestige is a difficult question. It is "common knowledge" that some journals have higher quality than others. Yet where this common knowledge has been investigated, consensus is not as high as one would expect. In psychology and education it has been shown that prestige rankings of journals vary considerably as a function of work position and interest area (Koulack and Kesselman[10]; Luce and Johnson[11]). In this study, no account was made for differences in the publication source. An entry in the **Research Report** was defined as a "publication".

Author number and order was accounted for in the formula. If a publication is worth a score of one, then it would be counted twice if it has two authors — indeed unless a count is taken of author number, a five-author article would have the same value as a single-author book! Differentiating author order has its problems (discussed below); nonetheless, it was included in the formula. The actual formularisation is now described.

A single-author article was allocated a value of 1. Books were given a value of 5 and edited and translated books 3. In many disciplines there is an author order convention, with the first named author being the one making the major contribution to the research. This convention often varies within a discipline depending on the journal's editorial policy. Some journals use an alphabetical listing of authors, rather than a "contribution" order. To account for these differences, a distinction was made between alphabetical and non-alphabetical joint authorships. For example, in a two-author alphabetical article the authors were allocated 0.5 each; where the order was non-alphabetical the first named author was allocated 0.6 and the second 0.4. This system fails to distinguish the accidental alphabetical order, which is a decreasing problem as the number of authors increased. Errors due to this were considered to be minimal. For three-author non-alphabetical articles it was initially intended to award values of .5, .3, .2 respectively and for the four-author case .4, .25, .175, .175. However, multiple authorships up to 12 were found. As the weights were arbitrary, an algorithm was developed which enabled the calculation to be programmed into a calculator. The algorithm used approximated very closely the intended allocation indicated above. The actual form of the algorithm is given below in the specification of the value allocation:

For journal articles

1. Single-author: allocate 1
2. Multiple-author alphabetic: allocate to each author 1/n where n = no. of authors
3. Multiple-author non-alphabetic
   (a) 2 authors allocate 0.6 to first, 0.4 to second
   (b) 3 or more: allocate $0.6\left(\frac{n+5}{4n-2}\right)$ to first
   $0.4\left(\frac{n+5}{4n-2}\right)$ to second
   $\frac{1}{n-2}\left(1-\frac{n+5}{4n-2}\right)$ to all others

For books the above scores were multiplied by 5; for edited books by 3.

The values allocated for 3, 4, 5 and 6 authors were illustrated below.

No. of authors	Value allocated to each author					
	1st	2nd	3rd	4th	5th	6th
3	.48	.32	.2			
4	.39	.26	.18	.18		
5	.33	.22	.15	.15	.15	
6	.3	.2	.12	.12	.12	.12

For any individual this formula was applied to all his publications quoted in **Research Report** and the resulting scores were summed for each year. The summed values will be called the **Publication Rate Index** of the staff member for that year.

Compiling Publication Rate Indices

Our imaginary brief included developing an index to be applied to individuals and departments and faculties. The compilation of the **Publication Rate Index** (PRI) therefore involved two phases (a) the compilation of the PRI for Monash staff members and (b) the collection of these together in departments.

The first phase used the **Research Report.** Under each departmental listing, names of all authors were listed and the PRI for each name was computed. This apparently simple (although tedious) task was made quite difficult at times by errors found in entries. A number of error types were identified. The most serious of these were (a) incorrect names, usually changes in initials which make it difficult to be sure whether it referred to one person or two; (b) the same publication being listed more than once with a change in authors (in multiple-author publications); (c) co-authors changing position in multiple listing of a publication; and (d) variations in title of the same publication. All of these errors were detected because the publications were listed in more than one department (and so discrepancies were found that could be checked against the original source), but their presence raises serious questions about the accuracy of single entries.

The first phase procedure also signposted potential problems with departmental listings, for in many cases the same publications and names appeared in more than one department. It also became very clear after an initial attempt at this listing that the **Research Report** contained names (under a department's publication) of people who were not members of that department and/or the university — and that the use of the asterisk to indicate non-university membership was used quite inconsistently.

To deal with these complications the following strategy was adopted

- For a particular year, the **University Calendar** was used to develop a list of staff members of each department within a faculty.
- The list of names with PRI scores from a particular department as listed in **Research Report** was compared with the faculty lists and was then divided into three groups (a) a member of the department (as defined by the **University Calendar**), (b) a member of another department and (c) not a member of any department.
- When all departments of a faculty had been treated in this way, group (b) names were added to the appropriate department; the resulting department lists thus contained members of the department (as defined by the **University Calendar**) and others who were not members of any department in that faculty in that year.
- Where the same publication appeared in two departments in the **Research Report** and the author was not a member of any department, the score was divided between the departments.

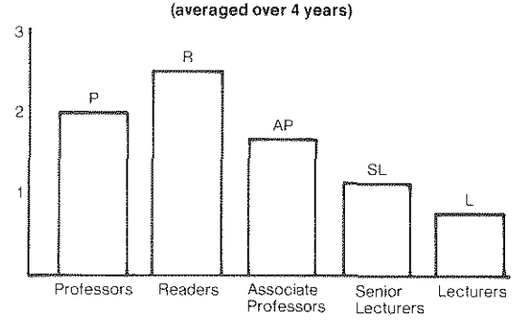
These operational rules were guided by the principle that a publication that derives from research in a department should contribute to the PRI of that department, even though the author may have left, be an outsider, a research student, etc. We are aware that the rules used do not take account of research done across faculties, but the likely incidence of this was thought to be too small to justify the very considerable additional work involved.

This procedure was applied to all faculties for the years 1974 to 1978, producing lists for each department for each year containing department members and others (as defined above). In this form they could be used to obtain an individual's mean PRI over the five-year period, a department's mean PRI or mean PRI per academic staff or establishment positions, etc., or similar measures for faculties.

Some Results

Figure 1 contains the PRI averaged over four years for academic ranks. The index can be interpreted (approximately) as "article equivalents (single author) published per year". These results provide some face validity to the measure. One would expect the publication rate order to fall with academic rank, except for readers whose post is a research-oriented one.

Figure 1
Mean PRI for different academic ranks



In Table 2, the frequency distribution of the same index is shown for a single faculty. For this faculty the mean PRI is 1.28, the median 0.6 and the mode 0. This shows that some academics publish a great deal, others publish rarely. This finding is consistent with similar surveys in U.S.A. and U.K. Lofthouse[12] reviewed surveys of publication rates and concluded that:

> *In the U.S. probably as many academics publish as do not. Of those that do, few are prolific. The British system functions in a similar fashion to the prestige U.S. section. The view that all academics are prolific publishers is wrong.*

Table 2
Frequency distribution of PRI averaged over four years for one faculty.

Mean PRI Range	Frequency (Percentage)
0	34.0
0.1– 1.0	32.1
1.1– 2.0	14.2
2.1– 3.0	7.6
3.1– 4.0	4.7
4.1– 5.0	1.9
5.1– 6.0	2.8
6.1– 7.0	1.9
7.1– 8.0	0.0
8.1– 9.0	0.0
9.1–10.0	0.0
10.1–11.0	0.0
11.1–12.0	0.9

The Validity of Quantitative Measures of Research

Construct validity
In this attempt to measure research activity, we used publication rate as the most useful measurable quantity for the evaluation. This was based more on pragmatic grounds than anything else. The availability and applicability of other measurables is extremely limited on a system-wide basis. Publication rate, of course, is only partly related to the construct "research activity" as discussed earlier, so that we must already have considerable concern about construct validity and indeed, in our title have distorted the construct to "research productivity". Publication rate is also compounded by other factors, such as personality.

This construct validity problem does not seem to cause too much concern in other areas where quantitative measures are used. For example, television ratings are collected at particular times that seem to be well known to the television networks. Nobody really believes the ratings during a rating period are representative of other times. Yet those ratings are used very strictly in the costing of advertising. A more obvious example is the use of blood-alcohol concentration (BAC) as a measure of drink-driving. It is well established that there is a far-from-perfect relationship between driving behaviour and BAC. Yet the measure is widely used (the construct validity problem in this case has been resolved by legal definition. The offence is now driving with a BAC greater than 0.05 or 0.08 — not drink-driving).

Other threats to validity
The experience of attempting to develop and use an index of publication rates has revealed that (1) the data bases are inadequate in that they contain errors, duplications, etc., and there is no simple way of estimating their extent; (2) the conventions that apply in one discipline vary so much that a universal formularisation is difficult to justify; and (3) the formularisation itself is arbitrary.

Conclusions about the validity of quantitative measures for evaluating research
In the light of the threats to validity discussed briefly above, our initial conclusion is that quantitative measures for evaluating the research of universities, faculties, departments and/or individual academics are very questionable.

It should be noted, however, that the same arguments could also be applied to subjective methods. Apart from some aspects of the arbitrariness of the formularisations, all of the other sources of invalidity are shared by non-quantitative evaluations. So attempting this quantitative measure has highlighted a real problem with non-quantitative evaluations of research. Such evaluations can (and do) occur without the need to examine problems like: what to do about multiple-authorship and order? What are the relative worth of books and journal articles? Is a publication in an international journal better than one in an Australian journal (in all sub-disciplines)? How accurate are the data bases? What differences in convention exist between disciplines? Is publication a valid measure of research? and so on. All of these issues arose in an attempt to produce a quantitative measure, yet they all exist in every attempt to evaluate the research of an individual, a department, a faculty, or a university. The subjectively based methods that form the bases of many appointment and promotion decisions, department reviews, etc., have apparently been able to proceed without the need to examine these problems. Perhaps wise decision-makers do take these questions into account, but this would not seem to accord with the common folklore that research is easy to evaluate but teaching is not.

In this light, perhaps our dismissal of quantitative measures has been too hasty. Because qualitative measures have many of the same problems, perhaps we should examine other advantages and disadvantages of quantitative and qualitative evaluations of research.

Other issues concerning the use of quantitative evaluations of research
In other contexts at least, administrators seem to favour quantitative measures. We have already mentioned television ratings and BAC as measures that contain many of the same problems. We could add the Consumer Price Index as a measure of inflation and many others. These measures appear to be acceptable to members of the industries in which they operate. So why not universities?

Such commitment to quantitative measures is easy to understand. They are open. Even if they are not entirely fair (i.e. not validly related to the construct) they **seem** to be so. The rules are known to everyone. This is not necessarily true of subjective measures. A staff member who fails to be promoted in a subjective evaluation method because of his research activity may never know whether he should have published more, sought larger research grants, supervised more graduate students, etc. If he fails because his **Publication Rate Index** (PRI) is not high enough, he can see how to maximise it next time. In that apparent advantage of objective measures lies also the source of their major disadvantage for universities. If an individual wanted to increase his PRI, he could do so without the increase reflecting any increase in the construct itself. One good example of this (in a different context) was the Federal government's change in Medibank so that it decreased the CPI (due to the formula) but did not reduce at all the cost of medical care to the individual, and therefore did not reduce the real cost of living. Such behaviour may be acceptable, even admirable politically, but it is undesirable in university research. Do we really want university research to degenerate into game playing; into an elaborate tokenism that manipulates the formula at the expense of genuine research, or at research that leads to quick publications, for example, in "pop" areas? (We suggest that we could get many publications from a project like "A survey of drug taking in Aboriginal and migrant women in inner urban areas".)

The introduction of quantitative indices of research for use in management decisions would, in our opinion, lead to just this situation. On these grounds alone their use should be rejected. However, if subjective evaluation is used, we contend that it should be required to meet the same standards of validity that would be required of objective measures. Further, we believe that the "rules of the game" should be articulated more openly than they are at the present time.

Nonetheless the decisions will still need to be made. Despite our conclusions we may need to continue our search for a quantitative index. If we were to do so, our next candidate would be a citations to publications ratio.

If anything positive has emerged from this research, it is a challenge to a common folklore. Whenever the issue of promotion through teaching is raised, the major argument against its use is that it is so hard to measure that we are stuck with giving great credence to research activity which is "so much easier to measure". Teaching **is** hard to evaluate, but, we submit, it is no harder to evaluate than research.

REFERENCES

1. Klein, R. "Universities in the market place" *New Universities Quarterly*, Summer, 1979, pp. 306–320.
2. Ladd-Lipset Survey *The Chronicle of Higher Education* XV, 12, pp. 12; 13, pp. 2, 1977.
3. Guba, E. G. & Clark, D. L. "Levels of R & D productivity in school of education" *Educational Researcher*, Vol. 7, No. 5, May 1978, pp. 3–9.
4. Liu, H. C. "Faculty citation and quality of graduate engineering departments" *Engineering Education*, American Society for Engineering Education, April, 1978, pp. 739–741.
5. Schaeffer, D. L. & Sulyma, I. M. "Citation rates and the quality of Canadian psychology departments", *Canadian Psychological Review*, Vol. 20, No. 1, January, 1979, pp. 22–37.
6. Liebert, R. J. "Research-grant getting and productivity among scholars: recent national patterns of competition and favor", *Journal of Higher Education*, XLVIII, (2), 1977, pp. 164–192.
7. Meltzer, B. N. "The productivity of social scientists", *American Journal of Sociology*, 55, July, 1949, pp. 25–29.
8. Crane, D. "Scientists at major and minor universities: a study of productivity and recognition", *American Sociological Review*, 30, October, 1965, pp. 699–714.
9. Cartter, A. M. *An Assessment of Quality of Graduate Education*, Washington, American Council on Education, 1966.
10. Koulack, D. & Kesselman, H. J. "Ratings of psychology journals by members of the American Psychological Association", *American Psychologist*, 30, 1975, pp. 1049–1053.
11. Luce, T., & Johnson, D. M. "Rating of educational and psychological journals", *Educational Researcher*, Vol. 7, No. 10, November, 1978, pp. 8–10.
12. Lofthouse, D. Thoughts on "Publish or Perish", *Higher Education*, 1974, 3, pp. 59–80.