PUBLICUATION RATES AND RESEARCH PRODUCTIVITY

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 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 Klein1 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 sector have already occurred in the USA and UK. The Ladd-Lipsit2 survey of 4,400 faculty members in 161 colleges reported output of university staff as a body. Various authors have investigated the research activities of specific disciplines, for example, schools of education in the USA (Quahe and Clark4), departments of engineering in the USA (Li5) and departments of psychology in Canada (Schaifler and Sulyn6).

A substantial part of this movement towards evaluating research activity has involved quantitative measures of research activity within a framework of measurement in social science in Leo H West; Terry Hore; Pak K Boon
Higher Education Advisory and Research Unit, Monash University
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 measurable 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.
- Peer evaluation of student research (in same discipline).

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 giving 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 certain priorities, so that some fields attract more grants for reasons such as perceived social worth, etc., that are not related to the research itself. Lefebre calculated the average number of grants per scholar in 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 only require 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, therefore differences would lead to index invalidity, 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 "inhibit" 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 referring to publications and books. Nor is it necessarily a measure of contribution to the discipline. A single person may have a greater contribution than one hundred other persons.

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 date base. The Monash 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 not interpreted individually. Quality of publications, editorial policy and compilation are motives of visibility, 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 formulation 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 need to be resolved were: the relative weights to be given to books, edited books, articles, etc., whether differential weighting 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's guideline that a book to article ratio of 3:1 is appropriate for multiple authors, single author books, but not articles, the score was divided by the number of authors. In contrast, Gail's approach was to use a formula for the number of authors. (The publication rate index was actually designed to enable the calculation of publication rates for individuals and departments of the same form as the original source, the Research Report. Three departments (a) a Research student, (b) Mr. Smith, and (c) Mr. Jones were all members of the same university department. A score was calculated for each, and divided by the number of authors. The result was then multiplied by the number of years in which the score was calculated.

Differences in visibility, translation, citation, etc. made it impossible to formulate an index that was valid for all disciplines.
In Table 2, the frequency distribution of the same index is shown for a single faculty. For this faculty the mean PRI is 1.26, 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 reviewed surveys of publication rates and concluded that in the U.S.A. probably as many academics publish as do O'Halloran, and in the U.K. little as well.

The British system functions in a similar fashion to the prestige U.S. system. The view that all academics are prolific publishers is wrong.

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

<table>
<thead>
<tr>
<th>Mean PRI</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-1.0</td>
<td>34.0</td>
</tr>
<tr>
<td></td>
<td>1.1-2.0</td>
<td>31.1</td>
</tr>
<tr>
<td></td>
<td>2.1-3.0</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>3.1-4.0</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>4.1-5.0</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>5.1-6.0</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>6.1-7.0</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>7.1-8.0</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>8.1-9.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>9.1-10.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>10.1-11.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>11.1-12.0</td>
<td>0.9</td>
</tr>
</tbody>
</table>

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 this evaluation. This was based more on pragmatic grounds than anything else. The availability and applicability of other measurable quantities is extremely limited on a research end base. 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 loci to be found in the areas where quantitive measures are used. For example, television ratings are collected at particular times that seem to be well known to the televiewing networks. Nobody really believes the ratings during a rating period are representative of other times. Yet those ratings are used very effectively in the cost 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 relation 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 defined as 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 databases 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 formulaisation is difficult to justify; and (3) the formulaisation 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 formulaisations, 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.

The offence is now defined as BAC greater than 0.05 or 0.08 (not drink-driving).

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 aren't entirely fair (i.e. not validly related to the construct) they seem to be. 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 Ph.D. students, etc. If his failure is because of 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 to the advantage 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 such as "A survey of drug taking in Aboriginal and migrant women in inner urban Australia." Or, for that matter, in sub-disciplines that are not funded?)

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 articulated more openly than they are at the present time.

Other issues concerning the use of quantitative evidence for decision making

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 aren't entirely fair (i.e. not validly related to the construct) they seem to be. 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 Ph.D. students, etc. If his failure is because of 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 to the advantage 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 such as "A survey of drug taking in Aboriginal and migrant women in inner urban Australia." Or, for that matter, in sub-disciplines that are not funded?)

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 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