Performance Indicators: Just How Do You Weight Them?

John Hattie and Jim Tognolini

Abstract

The development and use of performance indicators in higher education has become a major topic for research and discussion throughout the world. Research efforts and resources are being channeled into improving their use, These efforts include extending the range (including developing indicators of quality) and objectivity of indicators; improving the procedures for measuring weighting and combining performance indicators; and refining procedures for linking funding and resource allocation to performance indicators. This paper uses a questionnaire and a sample of responses to demonstrate a methodology for making explicit the weightings experts ascribe to individual performance indicators used in the process of comparing research attainment across different university departments.

Performance Indicators: Just how do you weight them?

The question of how performance indicators are weighted receives little attention in the literature. Generally researchers working the area have been more concerned with the number and the nature of performance indicators than with how those indicators might be weighted in forming a composite. Such information, however, can provide valuable insights into exercises where comparisons are to be made among profiles of indicators. For example, in cases where funding is linked to research attainment which is represented by a profile of information containing such measures as the number ~f publications, the number of postgraduate students, etc., or where selection is dependent upon position in a rank order of merit and the performance profile consists of a range of scores on a number of different academic subjects. In this latter case the weighted scores provide valuable information regarding the relative difficulties of the different subjects (see Tognolini and Andrich, 1994).

Whenever comparisons of profiles of performance indicators are made, the individual indicators within the profile have to be weighted. Generally these weightings are implicit and are consequently not used to inform debate about the outcome. However, it is possible to determine how the various indicators are weighted. When weightings of performance indicators have been derived they have proven to be very informative, although often sobering. Cave *et al* (1988), for example, noted that in the United Kingdom the weighting for resource allocation was 61% for teaching-based criteria, 35% for research-based criteria and 4% for special factors.

It would be too easy to assume that the weighting of indicators was linear. If that were so, in the case of research attainment, for example, the implication would be to acquire more bright students, to attract more prestigious staff, to publish anywhere, to ignore standards and the quality of research. It is highly likely, therefore, that the weighting is non-linear and probably follows a Cobb-Douglas model (see Fraser, Walberg, Welch and Hattie, 1987). This model involves non-linear product terms and an important major corollary is that it involves a law of diminishing returns. After initial gains, the increases in performance quickly diminish.

Weighting of Performance Indicators

Some attempt to derive weights has been undertaken in the area of publications. The Engineering Review (Williams, 1988) devised a 'relative publication rate' or 'Basic Publication Unit' (BPU), which weighted, for example, one-sole author chapter in a book as 1.0 BPU, one sole-author book as 10.0 BPU, one joint-author paper (two authors) as 0.50 BPU, and one joint-author paper (four or more authors) as 0.25 BPU. The average BPU per staff member over all university engineering departments in Australia was 0.21. This

means that the average engineer publishes one article in a journal (with three coauthors) each year, or one sole-author paper in a journal every five years.

Meltzer (1949) claimed that there were, on average, 18 chapters per book, and concluded that the weighting of an article to a book should be 18:1. Crane (1965) used a similar procedure for generating a 4:1 index, Carrter (1966) a 6:1 index, and West, Hore and Boon (1980) a 5:1 index. None of these used an analytical procedure to derive these weights.

There has also been debate about how to weight co-authorship. A method that has commonly been used assigns weights proportionally by order: credit = $(1.5^{n-i}) / (\text{Sum } 1.5^{i-1})$, where n is the total number of authors, and i is the particular author's ordinal position. Thus, for example, second authorship in a co-authored paper is given 0.4 credit units, and third authorship in a three-author article is 0.21 (see Howard, Cole and Maxwell, 1987). Not surprisingly, the correlations between the various indexes are very high (see Astin, 1982; Howard, Cole and Maxwell, 1987; Laband, 1985; Smith and Fielder, 1971).

All of these examples involve the authors making prior decisions regarding what they perceive to be appropriate weights. An alternative option for deriving the weights would involve judges ranking or rating the various indicators in some profile on an importance scale. Decision research has found, however, that judges strongly overestimate the importance they place on minor cues and they underestimate their reliance on a few major variables. Judges are often unaware just how few cues affect their judgments. Slovic and Lichtenstein (1971), for example, reported that across a number of studies, three cues usually sufficed to account for more than 80% of the predictable variance in the judge's responses. The most important cue usually accounted for more than 40% of this variance.

In a recent study Hattie, Tognolini, Adams and Curtis (1992) used a different procedure for determining the weights ascribed to individual indicators by various judges. In developing a model for allocating funds across academic departments using selected indicators of research performance, they calculated and made explicit the weightings of individual indicators. This paper describes the quantitative procedure used. While the example concentrates upon research performance indicators, the process is generally applicable to any situation where profiles of information have to be compared.

The 'a' Index Exercise

Since 1987, the University of Western Australia (UWA) has been conducting an exercise in which all departments within the University submit information on a nominated set of performance indicators of research attainment, to the University's Research Committee. The information is then reviewed by nominated peers and an 'a' index assigned to each department. Research funds are then allocated accordingly. Those departments with the highest 'a' index receive the highest proportion of the funds.

The 'a' index exercise is an example of how both quality and quantity performance indicators can be collapsed onto a single dimension which allows the comparison of research attainment of different departments.

Procedure

In the 'a' index exercise the information from the different departments is reviewed by a panel of peers, established as sub-committees of the Research Committee. In practice the members of these committees would be the ones to participate in the exercise to make explicit the weightings of the indicators. For various reasons a significant number of these members were not available at the time in which this study was being undertaken. As a consequence, a sample of academic staff from across the University was chosen to illustrate the method.

Questionnaires were sent to 61 academics across a variety of disciplines, level of position (professor to lecturer), discipline groupings (Humanities; Social Sciences: Mathematics, Engineering and Physical Sciences; and, Biological and Medical Sciences) ,and productivity (defined as having published three or more works in three years). The questionnaire consisted of profiles of performance indicators for 27

departments (23 actual departments and four fictitious departments) based upon a list of 11 performance indicators for each department (see Appendix 1). Most of the indicators were identified as being applicable to all departments across the university (see Hattie and Tognolini, 1991).

The questionnaire aimed to ascertain each academic's subjective weightings of the various indicators (for more detail on this application of the Lens model see Cooksey and Freebody, 1985; Hammond, McClelland and Mumpower, 1980). It is acknowledged that various academics will have different views how the various indicators could be weighted, and the aim of this study is to make the academic's weighting policy explicit.

A number of modifications were made to reduce the complexity of this difficult task. In deciding the values of the performance indicators, every effort was made to render the task realistic. Actual indicator values from 23 departments were used, although they were all normalised to a department staff size of ten to remove the effects of staff size. The size of ten was used, as a larger number made some of the performance indicators look unrealistic (for example, when a very productive small department of, say five, produces five books, when multiplied by two the value looks unlikely).

38% of the questionnaires were returned, which is a very disappointing rate that probably reflects the inappropriate timing of the exercise (major grant applications were being finalised) and the difficulty of the exercise. A number of staff did not complete the questionnaire because they did not approve of the reasons behind the questionnaire. Others who disapproved still completed the questionnaire because of their beliefs about the importance of this kind of research project. On reflection the response rate could be an indicator of the lack of trust in procedures to assess research in universities, and points to the importance of maximising the collegial involvement in both the procedure and the decisions. This response rate is not so damning when it is appreciated that the power of the methodology is more related to the number of cues. The methodology can be used to inform the weighting scheme of just one person, and thus we have 23 replications.

Results and Discussion

There were sufficient responses to demonstrate the procedure and make some tentative conclusions. The respondents represented a reasonable cross-section of all departments and discipline groupings. The more senior and productive staff tended to be more represented in the respondents. Table 1 presents the standardised regression weights for the 23 respondents, and the R-squared. Note, that each line in the table represents the weightings for each individual; there is no contamination in the weightings based on other academics perceptions as each line is unique. The weights indicate the strength or importance of the various performance measures. The R-squared is a measure of the variance explained by these weights and indicates how adequate these weights (on these particular measures) are for each respondent.

The most impressive feature of these analyses is the number of variables that received minimal weighting. Given that the task requires a consideration of the 11 performance indicators, obviously some are going to be more salient than others. The ones that consistently received low or zero weightings were conferences, keynote addresses, number on editorial boards, and number of Masters students graduated (similar to Grigg and Sheehan, 1989). The most salient overall indicators were journals, PhD students, books, chapters, grants, impact, and percentage publishing.

The average R-squared overall is 0.77, which is remarkably high. This indicates that the variables were sufficient to explain the ratings of the 23 academics. There were some respondents with very low R-squares, and these staff tended to be from the departments that were rated as lowest in the index exercise and/or they weighted dome of the performance indicators significantly negatively. For example, respondent 21 was from a lowly rated department and, while having high positive weights for keynote papers, books and chapters, had negative weights for journals, conferences and grants. Research grants were weighted highly by about one-third of the respondents, and they were usually from engineering and a cross-section of the other departments.

Table 1 : Standardised Regression Weights and R-Squares for the 23 Respondents

No	PhD	Mst	Jrl	Key	Conf	Books	Chaps	EdBk	Imp	3pub	Gt	R^2
1	.24	-.35	.49	.20	.18	.46	.10	-.01	-.19	.21	.24	.62
2	-.30	.10	.72	-.09	.07	.10	.23	.20	-.22	.43	.58	.80
3	.16	.23	-.07	-.17	.16	.63	-.31	.20	-.20	.09	-.17	.85
4	.05	.41	-.07	-.20	.15	-.22	.06	-.11	.42	-.09	-.11	.63
5	.22	-.18	.40	-.13	-.15	.71	.12	.17	-.19	.53	.04	.82
6	.11	-.18	.44	-.08	.21	.52	.06	.22	.33	.05	.40	.93
7	.35	.22	.51	-.09	-.17	.21	.13	.20	-.28	.03	.13	.65
8	.53	.27	.45	.01	.14	.16	.07	-.02	-.05	.75	.00	.94
9	.31	.04	.34	.00	.18	-.14	.25	.03	.13	.00	.32	.70
10	.31	.08	.29	.14	.04	.08	.07	.09	.40	.05	.40	.83
11	.31	.01	.36	-.04	.08	-.07	.32	.20	.33	.29	.43	.90
12	.51	.15	.22	-.07	.00	.09	.29	-.15	.24	.21	.34	.65
13	.53	.27	.02	-.17	-.03	-.04	-.10	.17	.38	.07	.13	.78
14	.56	.07	.25	.12	.05	-.21	.07	.22	.00	.12	.22	.79
15	.73	.13	.21	.14	-.06	.17	.09	.05	.37	.16	.00	.69
16	.34	-.01	.27	-.08	.14	.48	.22	.02	.28	.13	.37	.71
17	.15	.37	.30	.08	.26	-.30	.10	-.06	.59	.09	.07	.82
18	.30	.27	.40	.35	-.04	-.12	.33	-.09	.23	-.04	.17	.82
19	.34	.26	.16	-.23	-.24	-.17	-.03	.22	-.17	.21	.02	.62
20	.33	.24	.69	.11	.09	.06	.10	.01	.11	.22	.17	.87
21	.29	-.08	-.22	.33	-.43	.30	.35	-.28	.27	-.12	-.29	.58
22	.49	-.02	.36	.09	.06	.03	.22	.10	.16	.10	.44	.89
23	.13	.03	.85	.06	.07	.61	.16	-.15	.13	.01	.08	.92
Avg	*.33*	*.10*	*.33*	*.01*	*.03*	*.15*	*.13*	*.05*	*.13*	*.15*	*.17*	*.77*

A hierarchical linked cluster was calculated across these 23 respondents and there were three distinct clusters. The first and largest cluster of respondents (n=11) were from across all but the humanities departments and from the more productive departments at UWA. These respondents placed the highest weightings on PhD students and journals, with medium weightings on impact, and grants. The second group (n=5) were from less productive departments, across all departments and subcommittee groupings, and they weighted PhD students and journals only. The third group (n=6) were from the humanities and they weighted journals and books, with medium weights for number of staff with three publications in three years.

Across the academics there were similar weightings for PhD, Masters, journals, keynote conferences, conferences and chapters. The respondents from the humanities had a higher weighting for books and lower weighting for chapters. The humanities and social sciences weighted editorial boards positively whereas the respondents from the other two areas weighted them negatively; the converse pattern occurred for impact. The other related variable (three publications in three years) was weighted positively by social sciences and humanities and zero by the other two, and grants were weighted positively by social sciences and the medical respondents and zero by the others.

Conclusions

A key issue associated with using performance indicators in decision making concerns the explicit weightings of the indicators. Generally this difficult issue is avoided by leaving the weighting to the tacit scheme of the various experts making the decisions. However, the unwillingness or inability to articulate the weightings has the potential to become a major obstacle to the use of performance indicators in decision making. A more overt approach in which that tacit weightings of the experts are made public would seem to be more beneficial for the acceptance and use of performance indicators in the long term. A procedure for determining the weights, using the judgments of experts, has been developed in this paper.

It is suggested that whenever panels of experts are asked to make comparisons of profiles of indicators they should provide information how they implicitly weighted the indicators in arriving at their judgement. The exercise could either be completed when they decided upon their ratings, or as part of the

process to decide their weightings. For example, the panels either individually or collectively, could decide upon their ratings and then determine the regression weights. This could inform the members how they, as individuals, were weighting the indicators and might lead them to modify the weightings or, at least, confirm that the weightings were as desired. The weightings could also be determined at the end of the exercise and this would be informative to participants as to the basis of the decisions.

References

Astin, A W (1982) "Why not try some new ways of measuring quality?" *Educational Record.* 63, 10-15.

Carrter, A M (1966) "An assessment of quality in graduate education." *American Council of Education,* Washington DC, 1-38, 58-70.

Cave, M, Hanney, B, Kogan, M and Trevett, G (1988) *The Use of Performance Indicators in Higher Education: A Critical Analysis of Developing Practice.* London: Kingsley.

Cooksey, R W and Freebody, P (1985) "Generalised multivariate lens model analysis for complex human interference tasks." *Organisational Behaviour and Human Decision Processes,* 35, 46-72.

Crane, D (1965) "Scientists at major and minor universities: A study of productivity and recognition." *American Sociological Review,* 30, 699-714.

Fraser, B, Walberg, H, Welch, W, and Hattie, J A (1987) "Synthesis of educational productivity research." Monograph in *International Journal of Educational Research Series.* Oxford: Pergamon.

Grigg, L and Sheehan, P (1989) *Evaluating research: The role of performance indicators.* The University of Queensland.

Hammond, K R, McClelland, G H, and Mumpower,J (1980) *Human Judgement and Decision Making: Theories, Methods and Procedures.* New York: Praeger.

Hattie, Tognolini, Adams and Curtis (1991) *An evaluation of a model for allocating research funds across departments within a university using selected indicators of performance.* Report of an Evaluations and Investigations Program conducted for DEET.

Howard, G 5, Cole, D A and Maxwell, S E (1987) "Research productivity in psychology based publication in the Journals of the American Psychological Association." *American Psychologists,* 42, 975-986.

Laband, D N (1985) "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, 216-240.

Meltzer, B.N. (1949) "The productivity of social scientists." *American Journal of Sociology,* 55, 25-29.

Slovic, P and Lichtenstein, S (1971) "Comparison of Bayesian and regression approaches to the study of information processing in judgement" *Organisational Behaviour and Human Performance,* 6,649-744.

Smith, R and Fiedler, F G (1971) "Measurement of scholarly work in academic institutions: A review of the literature." *Educational Record,* 52, 225-232.

Tognolini, J, and Andrich, D (1994) "Accounting for different units when comparing diverse profiles in the context of small groups." Under review.

West, L H 1, Hore, T and Boom, P (1980) "Publication rates and research productivity." *Vestes,* 23, 32-37.

APPENDIX 1

Performance Indicators In Research Weighting Indicators For Measuring Research Performance

The aim of this questionnaire is to ascertain how academics would weight various indicators of RESEARCH attainment. All information is confidential and at no stage will respondents be identified. The questionnaire should take between 20-30 minutes to complete.

The task is to consider a profile of indicators for various departments and then to form an overall view of the research attainment in that particular department using a six-point scale:

1 = Low research attainment
2 = Moderate research attainment
3 = Moderate to High research attainment
4 = High research attainment
5 = Very High research attainment
6 = Excellent research attainment

The several indicators of research performance are:

Abbreviation	Indicator
Area	Area of the Department: Band M = Biological and Medicine; PhSc = Physical Sciences; Hum = Humanities; SSc = Social Sciences
No. PhD	Number of graduates awarded PhD or DSc
No. Mst	Number of graduates with Masters awarded research
No. Jrl	Number of articles in refereed journals
Key Conf	Number of keynote addresses delivered at national or international conferences
No. Conf	Number of published/refereed published conference papers
No. Books	Number of books, edited books and monographs (excluding textbooks)
No. Chaps	Number of chapters
No. Ed Bds	Number of staff on editorial boards of refereed journals
Jrl Imp	Journal impact
3 pubs/3 yr	Percentage of academic staff publishing at least 3 articles/books in the 3 years of the review
$ Grts ('000)	Expenditure of refereed research grants for the Department

1. These indicators are not exclusive nor necessarily the best to reflect the research in each department. The departments also do not necessarily bear any resemblance to Departments in the University of Western Australia.
2. There are no correct answers.
3. The indicators in this questionnaire are adjusted to represent a department of ten staff members. Thus, the size of the department is not a variable in this exercise.
4. The data for each indicator are based on a three year period 1986-1988.
5. The number of publications takes into account multiple authoring and reflects only the number of publications directly attributable to persons in that department (eg., if only one of
6. the two authors of a co-authored paper is a member of the department, then the credit to the department is five).
7. Where an indicator is not available or not able to be calculated, n/a appears:

You are asked to consider each profile of indicators and decide what the overall research attainment would be for the department. Please tick the box which best represents your view of the department's research attainment.

Please ensure that **all** departments are rated.

Thank you for completing this questionnaire, particularly as we know that it is not an easy task.