THE ROLE OF EVALUATIVE BIBLIOMETRICS IN IMPROVED RESEARCH PERFORMANCE - THE AUSTRALIAN EXPERIENCE

R. Johnston and J.J. Franklin

The Centre for Technology and Social Change,
The University of Wollongong,
P.O.Box 1144,
Wollongong, NSW 2500
Australia.

INTRODUCTION

The field of scientometrics has grown rapidly, displaying all the characteristic S-curve features since it was first mapped out by Derek de Solla Price in his writings of the early 1960s[1]. The major emphasis has been on the development and refinement of various technical systems. Thus we have witnessed successively the development of publication, citation, co-citation, co-word and patent analyses. The development of each of these has been essentially producer-driven.

Rip[2] has argued that:

> in terms of performance and quality indicators, some stability has been achieved: notions of what they are and how they can be used i.e. 'customer concepts' have been articulated and 'production methods' are available to meet this demand.

He makes the point[3], through a series of historical examples, that the emergence of science indicators has been significantly shaped by concerns about national scientific capabilities. In other words, the very development of science indicators itself provides a case study of a long established field of inquiry within the history and sociology of science, viz. the relative influence of 'internal' logical and social forces versus 'external' socio-cultural and economic forces in shaping the emergence and structure of specialties or disciplines.[4]

The growth in demand for science indicators, however, also serves to highlight the possibility of analysis from a quite different perspective. If one accepts a broad definition of technology as:

> any tool or technique, any product or process, any physical equipment or method of doing or making by which human capability is extended.[5]

then science indicators can be regarded as a newly emerging technology, with steadily increasing capabilities.

One of the more informative concepts in the admittedly immature field of innovation studies has been the technology life cycle. This model has been used variously to describe the life cycle of an industry[6], structural characteristics of innovation[7] and stages of evolution of an enterprise[8]. Adapting somewhat from these, one can develop a model which may be appropriate to the past and future development of science indicators (Table 1)[9].

In the introductory phase of development of a technology, in a highly fluid state, the emphasis is on the performance characteristics of the new product (what can it do, and how reliably) and the products undergo frequent substantial changes. Innovation is stimulated by the recognition of a scientific or technological potential and to a varying extent by potential customer's perceived needs. Production is very flexible, demanding of highly skilled labour, and low in efficiency.

The development of science indicators would appear to accord strongly with this model, with intense competition between different products and companies (ISI, CHI, CRP), a high level of experimentation and of investment in research, a lack of standardisation, with each new application providing a new custom design, and a high level of craft knowledge required to adequately develop and use the products[10].

Table 1

Characteristics of Technology Development

	Fluid pattern	Transitional pattern	Specific pattern
Competitive emphasis on	Functional product performance	Product variation	Cost reduction
Innovation stimulated by	Scientific potential or Information on users' needs	Opportunities created by expanding internal technical capability	Pressure to reduce costs and improve quality
Predominant type of innovation	Frequent major changes in products	Major process changes required by rising volume	Incremental for product and process, with cumulative improvement in productivity and quality
Product Line	Diverse, often including custom designs	includes at least one product design stable enough to have significant production volume	Mostly undifferentiated standard products
Production processes	Flexible and inefficient; major changes easily accommodated	Becoming more rigid, with changes occurring in major steps	Efficient capital-intensive and rigid; cost of change is high
Equipment	General-purpose, requiring highly skilled labour	Some subprocesses automated, creating "islands of automation"	Special-purpose, mostly automatic with labour tasks mainly monitoring and control

With the emergence of stronger demand for science indicators and the application of market discipline, we might expect a shift of the 'industry' more towards the transitional pattern, characterised by the emergence of one or more accepted standard products, and a greater emphasis on improving the production processes to serve the growing market, and on educating users and servicing their needs. Indeed, a new niche market is likely to emerge in the provision of expert systems to enable the development and application of purpose-specific science indicators without dependence on the limited (if not necessarily expensive) supply of specialists.

Finally, it is possible to speculate on a future phase of standardised science indicator products, mass-produced, retailed through the appropriate information systems chainstore, where price and advertising become the major determinants of sales, and of required innovation. Of course, long before this stage, the original innovators will have left, or been forced out, and moved onto new fields of inquiry.

The important point we wish to make is that the growth in interest in and demand for science indicators among science policy and science funding agencies of governments signals a new phase in the institutionalisation of this field. The users will increase the demand for quality and reliability. They will also scrutinise the claims and promises of the proponents (the sales force) of science and technology indicators.

Secondly, to continue the industrial metaphor, with the development of reliable indicator 'products', there occurs segmentation of the market to provide services which fit the product to the characteristics of particular applications. This paper emphasises not product, but the **process** of development and application of science indicators.

THE PROCESS OF SCIENCE INDICATOR USE

The reasons for the growing interest in and reliance on science indicators in science policy have been widely canvassed. In particular, Ziman[11] has argued that under the new conditions of 'steady state' the science system has to respond to demands for greater accountability and demonstrable economic return and, as a consequence, resource allocation decisions, and the criteria on which they are based must be both more explicit and more transparent.

These trends, as Frank Press has argued,[12] are not a consequence of the failings of science. Indeed, to a significant extent they are a consequence of its successes. For with the internationalisation of markets and the rise to dominance of non-price factors and, in particular, of quality and performance based on technological innovation, science has become too important a resource to be under-utilised.

The decision-making processes of science may have been adequate in the past, where their relationship to international competitiveness and economic survival were much more distant. But with science as a key ingredient, the needs for more urgency and more focus have become paramount.

In this environment, science indicators are not only a tool to assist in the making of more rational decisions which can be publicly scrutinised. They also usher in a process leading to quite different structures of decision-making.

As with decision-making by bureaucracies in other areas, there has been an increase, identified by Franklin[13], in:

> i) centralisation; ii)planning; iii) reliance on expert advice; and iv) the application of 'intellectual technologies', which Daniel Bell defined as the substitution of problem-solving rules and algorithims for intuitive judgment. This type of process relies on reduced and codified information i.e. indicators.

However, while there has been great progress in the production of reliable science indicators, and in the perception of their value, the process of their application is still in an extremely primitive form. In essence, it rests on the naive but very traditional assumption that the provision of extra information automatically leads to improved decision-making. This assumption is quite inappropriate in terms of the model of decision-making and its apparently apolitical claims of neutrality.

i) The Model of Decision-Making

There are a range of distinct literatures which criticise this assumption from different perspectives. Crick[14], in his seminal 'In Defence of Politics", argues strongly for the value of the conflict and compromise approach to decision-making in contexts where different values are held and points out the limitations and dangers of confusing decision-making with information processing.

Nelkin[15], through her analysis of a range of public controversies in which scientists have become involved, has shown that the contribution of scientific expertise rarely if ever leads to improved decision-making. Indeed, it was more common that the consequence of the provision of scientific information was to increase the polarisation between contestants and their views.

From another perspective Lindblom[16], and more recently Collingridge[17], have made fundamental criticisms, on both theoretical and empirical grounds, of the model of 'synoptic rationality' as applied to political or bureaucratic decision-making. The many constraints upon this decision-making leads to a much greater emphasis on incremental changes (what has been inaccurately, and somewhat unfairly, referred to a 'muddling through'). In addition--as individuals or single institutions almost never make a decision in isolation, but are forced to interact and compete with a wide range of 'others' with different objectives, values and agenda--the process of decision-making is a highly dynamic one which can be characterised by the adoption of a variety of strategies of 'mutual partisan adjustment'.

ii) The Claims of Apolitical Neutrality

In addition to its inadequate model of decision-making, the rationale for the development of science indicators has rested on an implicit technocratic claim of the ability to provide technical information to decision-making systems in a form which carries no assumptions or values and offers no challenge to established political authority.

It is, of course, possible to criticise and reject this approach from the standpoint of a general critique of technocratic assumptions[18]. However there is also a more specific weakness in the assumption of an apolitical status in the application of science indicators, and that emerges from the particular social system of science.

For the whole system of 'management' of the scientific enterprise has rested on judgements of quality with regard to people (through references), their past performance (through citation, etc) and their plans (through refereeing of proposals). This peer review system is the primary basis for the allocation of resources.

The application of science indicators within this process, which provides decision-makers (who would not necessarily qualify as peers) with a semi-objective basis for resource allocation decisions, represents a clear and direct challenge to the authority of scientists to make such decisions. Indeed, some scientists would see this as a challenge to the intellectual authority of science itself.

The consequence of this challenge were clearly illustrated by the response to the University Grants Committee's attempt in 1984 to establish a five scale rating (ranging from outstanding to below average) for all academic units in British universities:

> The result was an immediate and vehement protest from the academic community. The ratings were criticised as simplistic, unsystematic, inconsistent in criteria, unsupported by sufficient data, and biased in favour of the larger and more well established institutions. Studies were mounted to show that the UGC ratings did not correlate as well as might be expected with rankings

based on performance indicators (the number of publications per staff member).[19]

The behaviour of the peer community summoning evidence and arguments from all available sources, including other performance indicators, reflects the pattern of dispute and contest common when institutional authority is threatened.

The conclusion to be drawn is that there is a need for those engaged in the development and application of science indicators to show a much greater awareness of their social context, and to adopt a reflexive stance which mirrors precisely those significant advances made in the sociology of scientific knowledge over the past two decades.

A PROCESS-ORIENTED SCIENCE INDICATOR EXPERIMENT

In 1988, the Centre for Technology and Social Change was contracted by the University of Wollongong in Australia to develop a set of indicators of research performance. This decision was stimulated by the view of the senior management of the University that the Commonwealth Minister of Employment, education and Training, John Dawkins, had signalled in a Green tabel 2 **WOLLONGONG EXTERNAL INDICATORS**

INDICATOR TYPE	Type of Comparison	Indicator Name	Indicator Description	Indicator Calculation (per department/discipline* per year & total)
Research Activity:	a. department (or University) vs. own past performance	1. Raw Research Personnel (RRP)	no. of equivalent full-time (EFT) research personnel	Σ fractional-years spent on research by all personnel
		2. Raw Internal Funding (RIF)	total direct, internal financial support for research	Σ $'s directly allocated internally
		3. Raw External Funding (REF)	total financial support for research other than RIF	Σ $ from all external grants and contracts
		4. Research Personnel Intensity Ratio	EFT researchers as % of all academic staff	RRP / total no. academic staff
		5. Funding Distribution Ratio	distribution of direct research $'s among research staff	RIF / RRP
		6. External Support Ratio	% of research funding that comes from external sources	REF / (REF + RIF)
Research Output:	a. department (or University) vs. own past performance	7. Raw Research Output (RRO)	research output as measured by no. publications per year	Σ papers from academic staff in the SCI/SSCI in year n
		8. Personnel Productivity Ratio (PPR)	no. of publications per EFT researcher	RRO / RRP
		9. Funding Productivity Ratio (FPR)	no. of publications per internal $ of support	RRO / RIF
	b. department vs. department	10. Internal Effort Index	publications in disciplinary area as % of all Uni.W'gong publications	RRO for discipline or department D1 / (Σ RRO's, D1-Dn)
Research Impact:	a. department (or University) vs. own past performance	11. Raw Research Impact (RRI)	impact of research as measured by no. citations to papers	Σ citations received by year-n papers in years n, n+1, and n+2 **
		12. Raw Journal Influence (RJI)	expected research impact as measured by appearances of papers in highly influential journals***	Σ papers in highly cited international journals
	b. department (or University) vs. international standards	13. Impact-Output Ratio (I-OR)	citations per paper	RRI / RRO
		14. Funding Impact Ratio (FIR)	citations per $ of direct financial research support	RRI / RIF
	c. department vs. department	15. Journal Influence Ratio	no. of papers in influential journals as % of total papers	RJI / RRO
Other Academic Activities:	a. department (or University) vs. own past performance	16. Raw Teaching Intensity (RTI)	total no. of teaching hours for all academic personnel	Σ hours teaching per year for all academic personnel
		17. Raw Graduate Intensity (RGI)	total no. of enrolled post graduate students	Σ enrolled post-graduate students
		18. Raw Dissertation Intensity (RDI)	total no. of dissertations accepted (or completed)	Σ dissertations accepted
		19. Graduate Intensity Ratio	no. of registered post-grad. students per EFT researcher	RGI / RRP
		20. Dissertation Intensity Ratio	no. post-grad. dissertations accepted (or completed) per EFT researcher	Σ RDI / RRP
	b. department vs. department	21. Teaching Load Ratio (TLR)	measure of teaching load	Σ all students / RRP

Notes:

* Papers will be counted by department and/or discipline, depending on how well the departments' publications fall within standardized disciplinary boundaries.

** A 3 year citation counting period is used, because it has been noted that most papers receive the maximum number of citations in the 3rd year, and because 3 years seems like the shortest reasonable period. If the decision is made not to gather data retrospectively, then the impact of present-year papers cannot be measured until 1989.

*** Journal citation rates will be used to select the most influential journals in each discipline. A more sophisticated alternative to this indicator would be to weight each individual paper (or citation) by the citation rate of the publishing (or citing paper's) journal.

Paper[20] his intention to require the establishment of such procedures in the near future. It also recognised the availability of expertise in indicator development within the University.

On the basis of a series of prior projects by TASC[21], a process-oriented approach was adopted, which emphasised:

- the active participation of those whose performance was to be measured by the indicators;
- the use of the indicators for self-assessment with the objective of performance improvement, as much as for performance evaluation.

This required extensive consultation with academic staff, the preparation of informative material and the organisation of workshops in order to develop a consensus about appropriate indicators and the process by which they should be developed and applied.

Two types of indicators were constructed. 'External' indicators were designed to provide comparability across institutions (though not, we should insist, across departments), and were broken up into output and impact measures[22]. Standard input indicators were also collected, primarily for the purposes of normalisation. These indicators are listed in Table 2.

The second set of indicators were 'internal' and were designed by the individuals, departments and faculties themselves, using the normal decision-making structures. Each department was asked to nominate up to ten indicators which represented a quantifiable measure of a significant professional activity. The most frequently nominated indicators are listed in Table 3.

The project is still on-going and in part is caught up with changing requirements emanating from the Government. However, in terms of the value of the Wollongong experiment, Franklin[23] has been able to report that:

. It has contributed to more open communication about academic goals and the role of the university, serving as a vehicle for interaction between administration and staff;

. it has directly involved those being evaluated in the process of determining some of the performance criteria;

. it has raised awareness of the importance of improving performance and of demonstrating performance to external audience;

. it has raised awareness of the special strengths of the University which set it apart form other HEI's;

. it has supplied data which demonstrate some of those strengths and which can be used to defend the University;

. it has contributed to the University's ability to influence the national agenda and terms of reference for performance indicator development.

CONCLUSION

These Proceedings reflect the very considerable breadth and scale of advance in science indicators generally and in evaluative bibliometrics in particular. However insights are available from the sociology of scientific knowledge, science policy, and innovation theory which make it possible to predict that the field may be moving from its stage of enthusiastic growth and infancy to one of greater maturity where different considerations and pressures will be felt. The challenge is to use these insights to engage in active 'constructivism' of the future shape of discourse and inquiry.

References

1 Derek de Sollo Price, *Science Since Babylon*, Yale University Press, New Haven, 1961; *Little Science, Big Science*, Columbia, New York, 1965; 'The Scientific Foundations of Science Policy' *Native*, 233, 17 April, 1965; 'Networks of Scientific Papers', *Science*, 149, 510-515, 30 July 1965; see also Derek Price Memorial Plenary published in *Science, Technology and Human Values*, Vol.13, No 3/4, pp 349-384, 1988.

2 A. Rip, 'Mapping of Science: Possibilities and Limitations', A F J Van Raan, *Handbook of Quantitative Studies of Science and Technology*, Elsevier, Amsterdam, 1988, p 244;

3 A Rip, 'Science and Technology Indicators: Supply and Demand', mimeo , Department of Science Dynamics, Amsterdam, 1987.

4 For a review of this issue, see R Johnston, 'Contextual Knowledge: a Model for the Overthrow of the Internal/External Dichotomy in Science',, *Australian and New Zealand Journal of Sociology*, Vol 12, pp 193-203, 1976.

5 D. F. Schon, *Technology and Change: the New Heraclitus*, Pergamon Press, 1967.

6 J J van Duijn, *The Long Wave in Economic Life*, Van Gorsum, Assen, 1983.

7 R. Rothwell and W Zegveld, *Re-industralisation and Technology*, Longman, Harlow, 1985.

8 W J Abernathy and J M Utterback, 'Patterns of Industrial Innovation', *Technology Review*, Vol 80, 1978.

9 Adapted from Abernathy & Utterback, *op cit*, reference 8.

10 One interesting example is retailed in L Leydersdorff 'Problems with the Measurement of National Scientific Performance', *Science and Public Policy*, Vol 15, 149-152, 1988; J Anderson et al, 'On-line Approaches to Measuring National Scientific Input: a Cautionary Tale', *Science and Public Policy*, Vol 15, pp 153-161, 1988.

11 J Ziman, *Science in a 'Steady State': The Research System in Transition'* Science Policy Support Group, London, 1987.

---

12 F Press, 'The Dilemma of the Golden Age', Address to Members of the National Academy of Science, Washington, 1988.

13 J J Franklin, 'Performance Indicators in Higher Education: The Participative Approach', in T Wyatt and A Ruby,eds., *Indicators in Education*, Australian Conference of Director-General of Education, Sydney, 1988.

14 B R Crick, *In Defence of Politics*, Penguin, Harmondsworth, 1974.

15 D Nelkin, *Controversy: Politics of Technical Decision*, Sage, Beverly Hills, 1984.

16 C E Lindblom, *The Policy Making Process*, 2nd edition, Prentice Hall, New Jersey, 1978.

17 D Collingridge, *The Social Control of Technology*, Frances Pinter, London, 1980.

18 For example, see J Meynard, *Technology*, Free Press, new York, 1969.

19J J Franklin, *op cit*, reference 13.

20 *Higher Education: A Policy Discussion Paper;* Australian Government Printing Service, 1987.

21 J J Franklin and K Bryant, *Experience with the 1983-84 Co-Citation Bibliometric Model of Australian Science*, Department of Industry, Technology and Commerce, Canberra, 1988; J J Franklin, F Collyer and R Johnston, *Research Evaluation, Impact and Strategy*, TASC Report, 1988.

22 H F Moed, W J M Burger, J G Frankfort and A F J Van Raan, 'The Use of Bibliometric Data for the Measurement of University Research Performance', *Research Policy*, Vol 14, pp 131-149, 1985.

23 J J Franklin, *op cit*, reference 13.