SELECTIVITY IN FUNDING: EVALUATION OF RESEARCH IN AUSTRALIA*

J. Jeffrey Franklin

Selectivity and concentration in research funding has become unavoidable in Australia today. Some of the reasons for this are reviewed relative to international economic, scientific/technical and conceptual developments. The resulting need to develop an evaluative culture in and for Australia is discussed. The reasons for undertaking evaluations are outlined, and a working definition of 'research evaluation' that may be suitable within the Australian context is developed. The parameters that may deserve consideration in designing an evaluation are detailed, and a series of conceptual and practical guidelines are forwarded. Several barriers to implementing evaluations that may apply to Australia are addressed. Finally, the implications of the concept 'accountability' for both the recipients of government support and *government itself are briefly raised.*

Keywords: research evaluation, S&T policy-decision-making, Australian research funding policy, R&D management.

INTRODUCTION

The climate for funding and pursuit of science in Australia is in revolution. This revolution is impacting all sectors — academic, industrial and governmental. 'Selectivity', 'concentration', and 'accountability' are the bywords, yet an understanding of their full implications, and of how to operationalise them, is still developing. Consider, for example, the 22 September speech by the new Minister for Employment, Education and Training (DEET), 'The challenge for higher education in Australia', which states:

> For example, not all institutions can be funded for reasearch and a substantial number will not be; equally, no one institution will be funded in future for research across all of its activities . . .
>
> . . . the basis for funding should be shifted away from the current focus on input costs to a more competitive system which emphasises specified outputs and measure of results, and allows institutions to bid for new places on this basis . . .[1]

These statements were foreshadowed by reports from both the Commonwealth Tertiary Education Commission (CTEC) and the

* I am grateful to Ron Johnston for helpful comments on earlier drafts.

Australian Science and Technology Council (ASTEC) earlier in the year.[2,3] Moreover, the formation of the Australian Research Council (ARC) represents a move to: 1) coordinate the national research effort, in part by concentrating and coordinating existing research schemes; and, 2) make more competitive those research funds which are currently provided as general recurrent grants, i.e., increase selectivity and concentration in research.[4]

While these types of changes are fairly new to Australia, they are not new in Europe and North America. Both Britain and the Netherlands have already shifted to conditional, selective financing schemes for higher education research,[5,6] and other countries, including the US, Canada, France and Sweden, have been investigating approaches to selectivity and concentration in funding for some time.[7,8] Government research institutions are no less subject to these demands. The Commonwealth Scientific and Industrial Research Organisation (CSIRO), for instance, has been pressured to hold its budget down and has been encouraged to seek more actively industrial support and collaboration.[9]

Australian industrial corporations are also under pressure, but for somewhat different reasons. Incentive schemes for industrial R&D (IR&D), such as the 150 per cent tax scheme and the Grants for Industrial R&D (GIRD) program, have put 'positive pressure' on Australian firms (since 'negative pressure', such as the threat of budget reduction, does not apply). Statistics suggest that these schemes are working.[10] Offset liabilities have also pressured the Australian subsidiaries of foreign multinationals to perform R&D in Australia. In addition, the Department of Industry, Technology and Commerce (DITAC) has targeted certain 'enabling' topic areas for increased IR&D, facilitating collaborative research between industrial and tertiary education organisations.

The question is: how are these firms, which may never have performed R&D before, deciding on research priorities and directions, and how are government agencies, such as CSIRO, DEET and DITAC, going about making strategic selections? How will Australian universities develop more systematic ways of managing their research effort and of demonstrating the value of that research?

The effect of the changing climate for research is that many increasingly difficult and important decisions must be made systematically and under scrutiny. Now it is necessary to say why certain areas of research were funded and not others. The pressure is to know how well the research achieved the purpose for which it was funded, whether that be advancement of state-of-the-art knowledge or production of a specific technical outcome. It is desirable that this information be used as input to future funding decisions, so that the research is more productive and the use of resources more efficient. Further, the pressure is to make those decisions and choices 'transparent' to the public or the supporting entity, i.e., to make them explicable and defensible. Along with selectivity and concentration, then, comes accountability and the need for research evaluation.

A recent report prepared for the Organisation for Economic Cooperation and Development entitled *Evaluation of Research, A Selection of Current Practices* (hereafter called 'the OECD report') concludes:

> This investigation has revealed that evaluation is taking place throughout the university system in *most* Member countries and that it takes place at a multiplicity of levels. The *tension* lies between the mutual evaluations by scientific peers, which are used to shape the direction and maintain the quality of a discipline, and the wider demands made upon evaluation as an instrument for changing structures, determining the allocation of resources and assessing the performance of an area in contributing to mission objectives. Much of the effort of the use of foreign experts and other indicators and devices all represent ways in which it is sought to demonstrate quality of science to a wider audience which may not be fully convinced by what it sees as the mutual beneficiaries of the peer review system (Emphasis added).[11]

Australia is not yet among 'most' countries, and this 'tension' is far from resolved.[12] The large majority of funding decisions made in Australia are made strictly on the basis of traditional peer review processes, which generally are still considered above accountability or any guidance. Yet, Australia is at a cross-roads of structural change that makes this insufficient in many current situations. This paper argues in support of peer review as the foundation of evaluation, but also for extending and broadening the concept, use, and information base of peer review.

Progress is being made. For example, the National Health and Medical Research Council (NH&MRC) is now actively investigating information sources to supplement peer review. Many universities are reportedly in the process of developing research management strategies (voluntarily or under government pressure), and at least one has already established a program for research performance evaluation that will use both quantitative and qualitative indicators.[13] These activities represent a beginning.

However, a recent study commissioned by the Commonwealth Tertiary Education Commission concludes: "we lack a developed culture of evaluation."[14] It seems that now is the appropriate time for us to develop that culture. Some guidelines have already been supplied by the wealth of recent research and experience overseas. We do not need to reinvent the wheel, only to modify it for Australian conditions.

This paper aims to further the process. An attempt is made to establish some conceptual and practical guidelines for an approach to research evaluation that may be appropriate within the Australian context, drawing on the international experience and literature. As an introduction to this, the changes and conditions that have made the concepts of selectivity, concentration and accountability prominent today are reviewed briefly.

SOME REASONS FOR THE GROWING PRESSURE TO EVALUATE

In the past decade, parallel developments in three areas have lead to the increased need to plan and evaluate research more systematically:[15]

1. changes 'external' to science;
2. changes 'internal' to science and technology systems; and,
3. conceptual changes.

In addition, a fourth development — expanded understanding about evaluation methods and techniques and the availability of large, international databases — has increased the practice of evaluation.

'External' and 'Internal' Developments

The 'external' changes are primarily economic and social. In short, world markets have internationalised, and nations and large corporations compete alike for new knowledge and new markets. Technological and scientific innovation is widely recognised as critically important to national economic competitiveness. Competitive advantage based largely on natural resources and a protected national market is no longer tenable. The story is familiar by now.

Parallel 'internal' changes include: 1) the rate of turnover in new scientific and technical knowledge appears to be increasing rapidly, particularly in certain high-tech fields;[16] 2) the distance, and even distinction, between what traditionally has been called 'basic' and 'applied' research (in the Physical/Biological sciences) has diminished;[17] and, 3) the scale of intrinsic competition within science seems to have increased, both in geographic and cognitive terms.

It has been observed that competition for intellectual leadership and resources in science has intensified, shifting from an individual to an institutional and national scale, as levels of scientific and technological activities have soared.[18] Under these circumstances, the perception among some researchers that public support for research has been declining is reasonable, but generally mistaken.[19] The most current statistics available show an overall increase in Australian government support for higher education research from 1976 to 1986 (though the percentage of total government expenditure allocated to research may have declined).[20,21,22] The dilemma, then, appears to be this: there is more science worthy of support than there are dollars. Also, the number of academic researchers has increased more rapidly than the rise in support, so that the money is spread ever more thinly. The problem, then, is not that the money is less, but that the science is more.

The science is more in more ways than one. The cost of performing leading-edge research in the Physical and Biological sciences appears to be increasing with reliance on costly machinery, instruments and large staffs of technicians.[22] This has motivated international collaboration

and cost sharing in 'resource-intensive' science, such as High Energy Physics (e.g., CERN), and led to consideration of national selectivity policies at the disciplinary level.[24] Not even the United States can afford to maintain world leadership in all disciplines now. Selectivity is unavoidable.

Conceptual Developments

The thinking about science and technology and the funding of it also has changed, primarily in two areas. First, understanding of the innovation process has advanced significantly. The 'linear model' of innovation, which assumes a one-directional flow of development from fundamental science to applied technology (or the opposite), has been rejected in favour of a much more complex and reciprocal model of the relationship between science and technology.[25] Neither the 'science push' nor 'demand pull' theories are adequate. As a result, it is much more difficult to make an argument for public support of fundamental science strictly on the grounds that it is a prerequisite to technical innovation, because the opposite may be the case as well.

Further, the 'internal'/'external' distinction being used in this paper is now recognised as largely artificial. The traditional boundaries between Science, the institution, and the rest of society have eroded in both real and conceptual terms. Research in the sociology and history of science has demonstrated the extent to which scientific 'fact' is 'socially constructed'.[26] It is argued that researchers have always translated 'external' social, political, and economic constraints and incentives into their problem selection and priority setting decisions, and then translated the results from the laboratory back to that 'external' context. Therefore, advances in understanding about how science works, more than economic necessity, have thrown serious doubt on the notion that science cannot and should not be steered toward social and economic goals. However, these arguments were developed around the Physical and Biological sciences and may apply in a very different way to the Arts and Humanities. (The difficult and unresolved questions concern how to steer science — or not over-steer science — in ways that are both beneficial to society *and* 'healthy' for the advancement of science.)

The second conceptual change is in thinking about the advantages and limitations of the peer review system. Peer review is still recognised as the cornerstone of any decision-making process about resource allocation for science.[27] However, the limitations of peer review are also now recognised.[28,29] In short, peer review does not function well under conditions of diminishing resources for research, when decisions need to be made across disciplines or in relation to economic and social objectives, or when the peer community is of a size or make-up that there are no 'disinterested' parties (e.g., conditions in Australia today).[30] Under these conditions, peers have been observed to become advocates

for their own disciplines or subdisciplines.[31] Management of consensus becomes necessary and difficult. Also, scientists are no better equipped than other people to make relevant judgements where social and economic parameters must be considered. The emerging view, then, is that peer knowledge is *necessary but not sufficient* in the decision-making contexts that characterise current conditions.[32]

Evolution of Methods and Techniques

In effect, the definition of 'peer review' is being broadened in response to the above factors. The OECD report defined three types of peer review currently being practiced:

i) *Direct peer review* defined as a review by a scientific peer (or peers) carried out specifically for the purpose of determining, and confined to, questions of scientific merit . . . ;

ii) *Modified peer review* is similar but the criteria are broadened to include socio-economic considerations. Balanced judgements are necessary in these circumstances;

iii) *Indirect peer review* adds information based upon peer evaluation made for different purposes and at different times. The majority of this information and analysis is bibliometric.[33]

Number iii) is becoming increasingly prevalent, particularly under the conditions described above and for *ex-post* evaluations when professional evaluators are involved.[34] In order to compensate for the limitations of peer review and to supplement the perspective of 'local' peer communities, the information base of national peer review therefore is being extended through other sources, such as international peer panels and bibliometric indicators.

There appears to be a growing acceptance of limited and controlled use of quantitative methods for evaluative purposes. Bibliometric indicators and science modelling approaches are viewed as supplementary to peer review, not competitive with it. The Advisory Board to the Research Councils (ABRC) of Britain has reportedly "endorsed citation and co-citation as worthy of further study and commended their use to research councils while cautioning that they should not be used alone . . ."[35] A recent review of research evaluation approaches commissioned by the Canadian government concluded: "on the basis of a large literature review . . . both quantitative and qualitative sources of information are necessary for a fully effective evaluation."[36]

DEFINING RESEARCH EVALUATION

Reasons for Evaluation

Research evaluation, paralleling research itself, may have multiple objectives.[37] However, there are basically two general reasons for under-

taking it. Put simply they are: 1) so you will know what you are doing and how well you are doing it; and 2) so that you can demonstrate that you are doing what you said you would do.

The first of these is often taken for granted, yet, as many researchers and research administrators know, it can be very difficult to move from a partially intuitive understanding to a specific statement of research objectives (or outcomes) and how they relate to institutional priorities, mission objectives or broader societal goals. Evaluation is partially a process of making these explicit. Also, by specifying them, it becomes possible to monitor research progress, using time-tables or milestones, and to assess how well original research goals were met. Evaluation structures these processes and, if successful, allows a systematic assessment of performance. However, assessing past performance is only constructive if the intention is to improve future performance. Reason 1), then, assumes a systematic process, a relating of goals to outcomes, and an intention to use evaluation as a means of improving performance.

Reason 2), demonstration of performance, recognises that evaluations always take place within a political context, even when they are 'in-house'. The political reality is that some entity is given a means of assessing the performance of another entity, which is given the opportunity to demonstrate performance. There is, therefore, a danger that evaluation will be used by the former party for non-constructive political motives. This means that making the decision-making context for evaluations explicit is important, particularly for those being evaluated. There is a corresponding danger that evaluation will be approached strictly as a means of self-defence and not taken seriously as a means of self-assessment and self-improvement.

The hypothesis forwarded here is that evaluation is only worth doing if both reasons 1) and 2) are at work. Therefore, *improvement of performance* and *demonstration of that improvement* will be dual motives if the evaluation is to be constructive,

Six Axioms and a Preliminary Definition

If the above points are accepted, then a number of implications may follow. The first of these concerns the relationship between research planning and research evaluation. Evaluation has typically been considered in one of three ways or phases: ex-ante, interim, or ex-post. Ex-post evaluation is often thought of as the only type of evaluation. The objective is to determine how well the research achieved the goals set for it. The interest is in outcomes. Interim evaluations are performed while the research is in progress. The objectives are largely administrative: to monitor progress, foresee potential pitfalls, provide for unforeseen requirements, and keep the project on schedule and within the budget. Ex-ante evaluation is what is normally thought of as research planning and is tied, at higher decision-making levels, with the formulation of research policy.

There is a fourth option: evaluation as a continuous, cyclical process. In this case, the findings from ex-post evaluation feed into research planning which sets the agenda for research monitoring which gathers the necessary information for the next ex-post evaluation. Though there are many situations where a one-off, ex-post evaluation is sufficient to the objectives, a longer-sighted perspective may be the one that views evaluation as a continuous process (axiom 1). Both the OECD report and a recent Commission of the European Communities document (hereafter called 'the CEC study') recommend this perspective.[38,39] This approach seems particularly appropriate if research performance improvement and institutional advancement are objectives.

In addition, there is evidence that the process of participating in planning/evaluation activities can be as valuable for performance improvement as any finding or outcome (axiom 2).[40] The hypothesis is that evaluation optimally can be conducted in such a way that it a) stimulates thinking about performance and goal setting, b) increases vertical and horizontal communication (between the R&D managers and marketing or between the Vice-Chancellor and the Faculties, for instance), and c) results in a cross-fertilisation of ideas about new opportunities. This is certainly consistent with the aim to improve performance. It also may be consistent with the aim to demonstrate performance: a recent report suggests that participation by Australian universities in the process of evaluation may itself be an indication of commitment to higher performance that deserves reward.[41]

Perhaps in contrast to this point: the general consensus from both the governmental and industrial perspectives is that evaluations need to produce policy-relevant information (axiom 3).[42] As one author puts it: "the customer must be able to take action."[43] In the context of this paper, the 'customer' includes those being evaluated, as well as the administrators and external client-audiences, and 'action' means decision-making about research priorities, directions, selectivity and concentration.

If evaluations are to support reliable, informed, and defensible policy decisions, and the entire process is to be viewed as credible, then two conditions are recommended. First, the evaluation process itself needs to be systematically planned and performed (axiom 4). Otherwise, there is a danger of wasted resources and production of findings which are not useful and/or not viewed with credibility. Therefore, *ad hoc* reviews and informal discussions among peers are generally not considered evaluation.

Secondly, it is recommended, almost by definition, that evaluation involve a systematic gathering and analysis of information (axiom 5). Making sure that necessary information is available on a timely basis for those involved in the evaluation has been recognised as one important step in planning the process.[44] It may be desirable to consult multiple sources of information, depending on the type of research and the importance of the decisions in question. Intelligence about the state-

of-the-art in the field, the potential impacts (social, economic, etc.) of the research, or the potential market for the successful innovation are a few examples.

The axioms recommended thus far suggest that it may be useful to approach the conduct of an evaluation as a research problem.[45] Evaluation is policy research (axiom 6). Accordingly, it may require a willingness, on the part of peers, to put aside disciplinary loyalties (a 'willing suspension of expertise') and a willingness, on the part of policy-makers, to entertain inputs and outcomes that assail as well as validate current policy.[46] Further, it suggests that the evaluation procedure be a 'learning system', i.e., one that retains enough flexibility to incorporate what is learned as the process moves from initial intentions and designs to a fuller understanding of the social and intellectual context for the research and the policy objectives (which themselves may be evolving).

With these axioms as a basis, it is possible to formulate a preliminary definition of 'research evaluation'. The following definition is proposed:

> *research evaluation* — a systematically planned and performed process of information gathering, analysis and interpretation which directly or indirectly involves the administrators, performers and potential users of the research and which seeks to establish the inter-relationships between a) the inputs to the research, and b) the outcomes, or expected outcomes, from the research.[47]

As suggested below, the 'inputs' to the research include the political and institutional context within which the research is funded and performed, as well as capital and human resource factors. Similarly, the 'outputs' may be very diverse, e.g., cognitive, social, economic, and/or technological, and very difficult to define or measure.

THE EVALUATION PARAMETERS

From the experience in Europe and North America has emerged a rough consensus on a number of parameters and guidelines for research evaluation, at least as it applies to the Physical and Biological sciences. Different authors have emphasised different sets of concerns, and the sets have only partially overlapped. An attempt is made here to synthesise this information and to construct a unified set of parameters and guidelines.

The following parameters are proposed as those which deserve consideration in designing any research evaluation:

1. the objectives of the research being evaluated (or type of research);
2. the decision-making context;
3. the purpose of the evaluation;
4. the type of evaluation;
5. the evaluation criteria;

6. the evaluation organisation; and
7. the evaluation methods and techniques.

Figure 1 is a schematic flowchart that suggests the appropriate order for addressing these parameters.

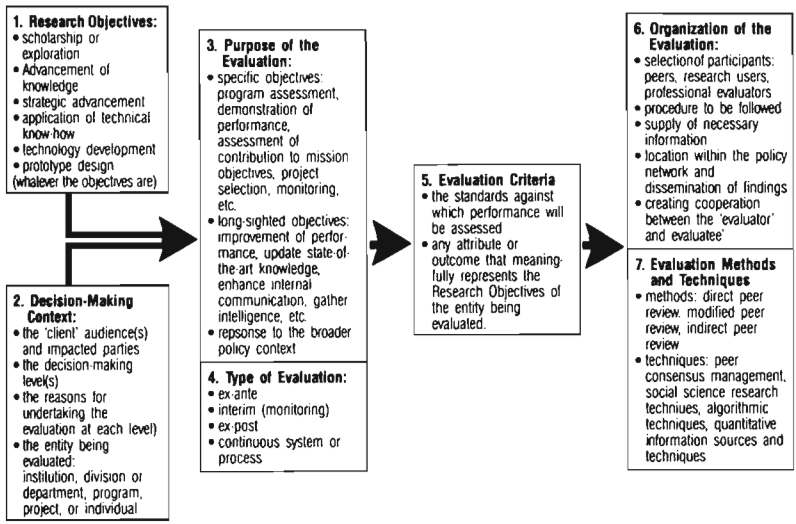

FIGURE 1
Summary Flowchart: Research Evaluation Parameters

Parameter 1: Research Objectives

Type of research refers to categories like 'basic', 'applied', and 'strategic'. Because these general categories have become less meaningful in recent years, it may be more useful to consider specific research objectives, for example: cultural enrichment (perhaps particularly for the Arts and Humanities), general exploration, advancement of state-of-the-art knowledge, advancement of enabling science, furthering strategic objectives, application of existing technical knowledge, technological development, prototype design, etc.

The first step in an evaluation, then, is to develop an understanding of the objectives of the research under consideration. This is critical, because the criteria against which the research will be assessed are derived from this understanding (see Parameter 5). As discussed below, a prevalent tendency is to reduce complex research objectives to simplistic variables. The danger from a public policy perspective is that research support will be guided by criteria that underestimate, and therefore undermine, the objectives of research that cannot be stated and measured easily.

Parameter 2: Decision-Making Context

The decision-making context can be characterised by asking: what questions will be addressed/answered by the evaluation, and who wants to know? The answer will include consideration of at least four factors:

a. the interested and impacted parties;
b. the decision-making level of the 'client';
c. the entity being evaluated, i.e. institution, division, branch or department, program, project, or individual; and,
d. the reasons for undertaking the evaluation.

There are generally multiple interested and impacted parties in the political context for an evaluation. The 'client' agency that requested and/or sponsored the evaluation is one (and perhaps the most important one for those performing the evaluation). The evaluation may be strictly an 'in-house' assessment or may have both an internal and an external audience. In controversial areas, the audience may include the general public. As discussed below, those being evaluated, i.e. the impacted researchers and research administrators, are very important participants in and users of the evaluation.

The decision-making level of the 'client' usually corresponds to the size of the entity being evaluated (institution, program, etc.) and determines the visibility and political sensitivity of the evaluation. The 'client' also generally sets the agenda or reasons for the evaluation, which are discussed below as the purposes for the evaluation.

Embarking on an evaluation without understanding the policy context is like mountain climbing at night. The risk of unexpected outcomes is greatest for those being evaluated (and those conducting the evaluation). A less severe but equally important risk is that the evaluation will simply fail to address the important issues or to produce policy-relevant information.

Parameters 3 & 4: Purpose and Type of Evaluation

The purpose(s) of the evaluation follows from the reasons for undertaking it as determined by the political context. The purpose may be difficult to neatly define, since it will usually be defined at several different levels.

It may include specific objectives, such as: project selection, research priority setting, program comparison, comparison with competing institutions, division assessment, assessment of past performance, demonstration of performance to an external audience, assessment of research contribution to mission objectives, monitoring ongoing research, and/or assessment of research impact, dissemination or utilisation.

As suggested earlier, there also may be more long-sighted objectives, such as: stimulation of research performance, intelligence gathering to

identify research opportunities, updating knowledge of the international state-of-the-art, enhancement of internal communication about research priorities and directions, or creation of an ongoing evaluation/planning system.

In addition, the evaluation will probably also serve a broader agenda related to the overall policy context. For instance, the Australian context characterised previously sets the arena within which evaluations now will be carried out. A decision to conduct an evaluation is, in effect, a response to this context.

Further, these considerations, in addition to the 1st and 2nd parameters, determine the type or time-frame of the evaluation, which was described previously as ex-ante, interim, ex-post, or a continuous process that combines these.

To summarise, then, there may be a hierarchy of purposes for the evaluation, perhaps corresponding to different levels of decision-making. It is recommended that these purposes be responsive to the research objectives and the decision-making context if the evaluation is to be successful. As the OECD report concludes: "Yet, in the experience gained carrying out this study, there would appear to be no more important factor determining the success of an evaluation than that its purpose (or purposes) be made clear at the outset."[48]

Parameter 5: Evaluation Criteria

The evaluation criteria are the basis against which the success or failure of the evaluated research will be assessed. It is therefore important that the criteria be defined carefully to reflect accurately the objectives of the research. The criteria serve to translate the research objectives into variables that can be analysed and, if possible, measured.

This may appear to be obvious and easy. Some institutional settings may allow (what appears to be) clear definition of a small number of research objectives and, therefore, a discrete set of evaluation criteria. For instance, much of university-based research may have advancement of knowledge as its sole objective, so there may be a single evaluation criterion as represented by peer judgement. At the other extreme, it may be assumed that industrial research is concerned with market advantage and that the evaluation criteria can be defined and measured in terms of research efficiency (innovation/dollar) and research effectiveness (amount of potential market opportunity generated by innovation).

However, true extremes of these types are increasingly rare. The OECD report observes:

> Between these two extremes lie the vast majority of research activities which are neither entirely basic research nor entirely product development; which are neither entirely in the universities nor entirely in the market place, where neither the operation of the peer review system nor the market mechanism is the appropriate criterion for evaluation. In most OECD Member

> countries, there is a general trend towards developing criteria that are applicable to this vast middle ground and it is being approached from both ends of the spectrum. For example, we see attempts being made to broaden the peer review system while industry itself is beginning to experiment with the idea of strategic research as a way of coping with the technical demands of the innovation process.[49]

So, the evaluation criteria will need to be responsive to the diverse objectives of the research being examined.

There is strong consensus by evaluators and in the literature that the evaluation criteria should be set and agreed upon *before* initiating the evaluation. This of course involves a negotiation process between the researchers, research administrators and those performing the evaluation. Doing this insures an understanding of the research objectives and a commitment to a systematic evaluation. The effect is to encourage planning of the research itself. The ability to delineate the research objectives and, from those, the evaluation criteria, reflects a professionalism on the part of all parties and a commitment to "the intellectual need to specify goals in such a way that we may know what effects and outcomes we are producing."[50]

Parameter 6: Organisation of the Evaluation

Organising an evaluation consists of these activities:

1. selecting and coordinating the participants;
2. designing and setting a procedure to be followed;
3. situating the evaluation in the policy network; and,
4. creating a framework for cooperation between those conducting the evaluation and those being evaluated.

Selecting and coordinating the participants in the evaluation is the most critical task in the organisation process. There are four potential types of participants: national or 'in-house' peer specialists, international peers, professional evaluators, and the users of the research.

The evaluation team will always include scientific peers or technical specialists. An important consideration is whether to utilise expertise from outside the program or institution, which requires identifying 'disinterested' specialists. One of four conclusions of the CEC study is that evaluations should involve program-independent expertise in order to produce results that are impartial and are viewed as credible.[51] In addition, involving external specialists — perhaps through international peer panels, for instance — offers the possibility of new insights and a broader perspective on 'local' science.

It also may be appropriate to involve professional evaluators in the process. The OECD report concludes that the use of professional evaluators is increasing in the Member countries.[52] Professional evaluators serve different functions than do technical experts. They are

most useful in the situations where peers are either least useful or least interested, i.e., for ex-post evaluations, in considering levels of aggregation higher than the specialty, and in synthesizing multiple evaluation criteria (e.g., scientific excellence, as judged by peers, and relevance to mission objectives, as stated by administrators). Professional evaluators also can service the peer review process by translating quantitative information into it and interfacing between peers and policy-makers. The use of professional evaluators may serve to enhance impartiality and credibility.

Involving the intended users of the research may be particularly important for research at the more applied end of the spectrum. For universities or government laboratories conducting 'enabling' research, this means involving industrial researchers or managers in the process. It has been observed that, where mission-oriented research is concerned, failure to involve potential users increases the possibility that the research will not meet its objectives and that the evaluation will not provide useful information.[53]

Designing the evaluation process and setting clear procedures to be followed is important for obvious logistical reasons. Also, as mentioned above, this may involve gathering information and systematically interjecting it into the evaluation process.

Situating the evaluation in the policy network is partially a matter of incorporating an understanding of the decision-making context into the design of the evaluation, and partially a matter of finding the best means of filtering the results of the evaluation back into it. Predetermined political or bureaucratic factors may limit the flexibility to do these.

The most difficult, and perhaps most important, task in designing and organising an evaluation is in creating an environment of cooperation between the performers/users of the evaluation and those being evaluated. Without this cooperation, the evaluation may be so confrontational that the longer-term objectives, such as enhancing research performance, are likely to be thwarted by the evaluation process. This issue is explored in the next section.

Parameter 7: Evaluation Methods & Techniques

The choice of evaluation methods is primarily between the three types of peer review discussed previously: direct, modified, indirect. However, even direct peer review can be usefully supplemented with additional information, such as international expertise or publication and citation data — hence the interest in evaluative techniques.

Techniques are the 'tools' available for supplementing the peer process. They become increasingly important as the method is shifted from direct to indirect peer review and as the evaluation criteria swing from scientific excellence to market opportunity. Use of quantitative information, then,

is the third way of enhancing the impartiality and credibility of peer review, following use of international expertise and professional evaluators.

There are at least three varieties of evaluation techniques: techniques for management of expert consensus, algorithmic approaches, and quantitative information sources. Examples of the first include questionnaire and survey techniques and the Delphi method. The opinion of the peer community (or the research-user community) can be polled using questionnaire and survey approaches. Analysts play an interpretive role in this approach, which has been employed in Britain with some success.[54] Delphi is another technique for systematically deriving a consensus from a group of experts, which can include researchers, research managers, business managers, etc. It is an iterative process of soliciting judgements from the selected experts or concerned parties, then statistically analysing and integrating those responses.[55] It usually involves anonymity among participants, but may conclude by convening the entire team to discuss and refine the emerging consensus.

Algorithmic approaches are techniques to structure the decision-making process or to synthesize information into a decision. Examples include: return on investment (ROI) methods, project scoring methods, cost-benefit analysis, and econometric modelling. These techniques are generally not appropriate when considering public support of research, but have utility in certain industrial situations. Because they tend to reduce complex scientific and technical variables to simple economic ones, they receive only limited use for evaluative purposes.[56]

Quantitative information sources, primarily in the form of S&T indicators, are being increasingly used by national policy bodies. For example:

- The US National Institutes of Health (NIH), the Dutch Advisory Council on Science Policy (RAWB), and the Canadian Commission on the Future Development of the Universities of Ontario have all directly used bibliometric indicators as sources of supplementary policy information.[57,58,59]
- The US National Science Foundation (NSF), the British ABRC and Royal Society, the Dutch Ministry of Education and Science, and the Swedish National Board for Technical Development, among others, have sponsored major investigatory studies of quantitative methods for policy applications.[60,61,62,63,64]
- Britain's ABRC recently has commissioned a bibliographic indicator profile of all universities and colleges.[65]
- Following the decision by the Dutch Ministry of Education and Science to replace non-discretionary funding of university research with a conditional financing scheme, Dutch universities have developed sophisticated indicator systems for internal assessment and priority setting.[66]

A body of knowledge about the proper use of quantitative methods is now accumulating. In most cases, indicators are being used with an understanding of their limitations and in ways that adjust for caveats to the data. For example, simple publication counts are being coupled with citation data and journal influence indicators as a means of taking research impact into account.[67] Secondly, appropriate bases for comparison are being established and followed. For example, in evaluating university departments, it is appropriate to make comparisons against past performance and against that of similar departments in other universities. It is not appropriate to compare unlike departments or fields, since different indicators will favour the Physical Sciences over the Humanities, for instance. Thirdly, raw output measures, such as number of patents or publications, are being normalised by input data, such as number of researchers or level of funding. This compensates for size differences between institutions, thereby showing where a small college has out-performed the large, established university, for example. The recent ranking of British universities and colleges by the University Grants Committee (UGC) has been criticised for its failure to do this, and more generally for its simplistic use of inadequate indicator data that may have favoured the large institutions.[68]

In general, an understanding about indicator usage is developing that recognises their 'partial'[69] nature and that they must be interpreted carefully by peer and policy specialists in relation to the context for which they were generated and to other information sources.

In addition, recent developments in thinking suggest a move away from 'bean counting' approaches to one that views the large, international databases as sources of intelligence for monitoring the dynamics of science and technology — hence the high level of interest in 'science mapping' techniques, such as co-citation bibliometric modelling and co-word analysis.[70,71] These techniques use patterns of referencing or word usage in the scientific and technical literatures to derive descriptive maps or models of the organisation and content of the international research front. A parallel development in the industrial sector may be a move away from algorithmic techniques to an approach focused on intelligence gathering, whether it be about the market or science and technology.[72,73,74]

The strongest consensus about the choice of evaluation methods and techniques is that it be determined by the evaluation purposes and criteria. One author observes that the availability of data or techniques should not determine their use.[75] Different methods and techniques are appropriate for different evaluation criteria.

One implication of this is that a certain level of knowledge of the available approaches and databases is required before an educated selection of evaluation 'tools' can be made. There is now a substantial amount of expertise and literature available for nearly all methods and techniques.

To summarise this section: it is first recommended that the parameters for evaluation be specified in advance and that they be systematically determined in the order suggested by Figure 1. The objectives of the research to be evaluated and the decision-making context will determine the purposes and type of the evaluation. These in turn will determine the criteria against which performance will be judged, and they, along with the previous parameters, will dictate the organisation for the evaluation and the methods and techniques to be used.

SOME OBSTACLES TO EVALUATION

Adversarial vs. Cooperative Evaluation

As noted earlier, one of the most important and difficult tasks in designing and organising an evaluation is in creating an environment of cooperation between the performers/users of the evaluation and those being evaluated. Evaluation can be viewed or approached either as a mode of external accounting or as a process for objective self-assessment and improvement. The former approach promotes an adversarial relationship between those being evaluated and those conducting it. It is in the interest of both parties to avoid this.

From the perspective of those requesting or conducting the evaluation, an adversarial relationship is counter-productive for at least two reasons. First, there is usually a concern with the long-term working relationship with those being evaluated. Secondly, and more importantly, if one of the objectives of the evaluation is to stimulate thinking and communication about research goals and to increase research performance, then cooperation is essential.

From the perspective of those being evaluated, an adversarial relationship is generally unnecessary, since in most cases both parties are on the same side: they each have an interest in advancement of research objectives, increased performance, and the long-term well-being of the institution. Also, greater participation in the evaluation process increases the chances of setting the terms for the policy agenda, thereby exerting more control. The evaluation also may offer a genuine opportunity to examine critically past performance, to create an atmosphere for 'team spirit' and higher performance, and to determine the best paths for achieving goals.

In Australia in general, evaluations have been practiced, used and viewed in a defensive, if not an adversarial, mode. 'Evaluation' virtually has become a dirty word. This view will need to be overcome if the benefits of evaluation are to be realised.

Several implications follow from the above observations. First, it is recommended that those undertaking evaluations express and frame them in such a way that cooperation is invited. The evaluation then can be organised to include direct participation by those being evaluated.

They can be invited, for example, to nominate some of the evaluation criteria or the performance indicators. This not only lowers the 'us/them' barrier, but is likely to produce a much more accurately representative and useful evaluation than otherwise would be possible.

Secondly, it is recommended that researchers and research institutions become much more directly involved in the processes that determine allocation of resources for research[76] and, therefore, establish internal evaluation practices. Beyond the potential of this to the institution and the fact that it demonstrates a commitment to performance, it will serve to normalise evaluation, making it part of the ongoing process of research performance and negotiation over support.

Operationalisation of Multiple Evaluation Criteria

Operationalisation of the evaluation criteria is a central problem for evaluations.[77] It requires: 1) understanding the objectives of the research (Parameter 1); 2) translating those into criteria or performance variables (Parameter 5); and, 3) identifying souces of information or indicators that represent the criteria (Parameter 7).

The first barrier is recognising the potential diversity and heterogeneity of research objectives. This may be a particular challenge in Australia for historical and cultural reasons. The long-standing separation of university and industrial research has resulted in a persistent tendency in considering the objectives of research, e.g., either it is supposed to advance knowledge *or* it is supposed to contribute to the immediate solution of an industrial problem. This attitude greatly hinders the development of adequate evaluations, because it insists that the evaluation criteria also be simplistic.

The second barrier, then, is admitting the appropriateness of multiple criteria of performance. For example, the CSIRO has a very broad mandate: to produce knowledge that benefits Australian society and advances Australian industry. The emphasis is on longer-term 'applied' research, yet the difference between this and 'basic' science in fields like biotechnology is indistinguishable. CSIRO's research objectives are multiple, and the political context places multiple and indistinct expectations on it. (This may be even more the case for universities). It is therefore only appropriate that multiple evaluation criteria be used, including both advancement of knowledge and contribution to solving immediate industrial problems. However, this approach generally has not been taken in assessing CSIRO's performance. A representative, convincing, and productive evaluation of CSIRO has yet to be produced.[78]

The third barrier or difficulty is figuring out how to judge or measure the criteria once they are agreed upon. Some criteria may be fairly straight-forward, such as 'contribution to state-of-the-art knowledge', which can be adequately judged by peers in most cases. The difficulty

arises with criteria such as 'creation of new knowledge in pre-competitive topic areas' or 'cultural enrichment' or 'advancement of community development'. No single judgement or measure can adequately represent whether the research has met these criteria or not. In relation to the first criterion, measures of degree of researcher consultation with related industry, extent to which the research is subsequently taken up by industry, or extent to which it results in patented know-how may all apply.

The approach suggested here is first to state each performance/evaluation criterion clearly, then ask: what would accurately represent or measure whether this objective/criterion has been (or will be) achieved? The most suitable answer may be a creative one that is responsive to the specific context in which the research is being performed. In cases where 'standard' indicators are not sufficiently representative — such as when US NSF-type indicators are applied to developing nations[79] or when publication and citation data are expected to be fully representative of university department performance — it may be desirable to customise appropriate indicators. The point is that attention to the specific research context and inventiveness in creating representative indicators are recommended if the evaluation criteria are to be operationalised meaningfully.

Structural Location of Evaluation

As noted by the OECD report, one of the difficulties facing the implementation and use of evaluations is locating them within the relevant decision-making networks.[80] This is a structural consideration. It has to do with locating the evaluation in relationship to the agency/program being evaluated and to the decision-making body(ies). The objective is to foster an effective flow of information and then of findings.

The question of structural location of evaluations may be particularly relevant within the Australian context. Other than ASTEC, there have been no governmental agencies in a position or with the capability to perform evaluations. The major Commonwealth funding schemes and research performing institutions were not structured to include evaluative functions. Policy research activities have taken place largely in isolation from actual policy decision-making. For example, though the Australian Research Grants Scheme (ARGS) was housed within the former Department of Science (DoS), and the DoS performed policy research, there was no real framework for tying that research to policy considerations of relevance to the ARGS.

There are alternatives to the structural arrangement in Australia between decision-making, policy research, and peer review. For instance, in some US agencies, research funding decision-making is structured at the program level around 'program managers'. Program managers

are typically Ph.D.s in the sub-disciplinary areas they oversee and are also aware of agency objectives and priorities. In some cases, they rely on an extended network of peer specialists to respond to research proposals via the mail in a standard format. What is structurally important about their role is that they act as intermediaries between the scientific community and the policy community, translating information from one to the other. The agencies house policy research divisions that are responsive to the decision-making demands placed on the program managers (and higher-level policy questions). Research evaluations can involve input from the program managers, and the results from evaluations can be filtered back into the decision-making process through the program managers and division heads.

This is not to suggest that a similar set-up is right for Australia, but only that it is important to consider the structuring of these components. Policy research, including evaluation, has typically been the ignored component. If the location of evaluative activities and policy research is not considered in the structuring of the new Australian Research Council (ARC), for example, then some difficulties encountered with the old systems undoubtedly will reappear. Evaluations will continue to be performed strictly on a one-off, *ad hoc* basis and will always be viewed as hostile, 'outside' tampering with the system. Simply changing the make-up of expert committees (to include industrial researchers, for instance) does not guarantee a more systematic or objective decision-making process, nor represent a greater commitment to accountability in government spending for research.

Practical Concerns

Two practical concerns deserve mentioning: the cost of evaluation, and evaluating the evaluation. Evaluation is not free. Peer review alone is a costly activity, both in real dollars and in terms of potential research time lost. Purchase of data for indicators and professional evaluator consultation are also costs. The developing perspective is that research evaluation should be considered part of the investment in research. A recent study of evaluation activities in the United Kingdom concludes: "Evaluation is becoming increasingly institutionalised in the UK, with the realisation that though it is resource-intensive, the resources it consumes are negligible (less than 1 per cent) compared to the scale of the budget decisions to which it contributes."[81] Similarly, the Dutch RAWB and the US NIH are two agencies committed to spending up to 2 per cent of the total research budget on evaluation.[82,83]

Finally, an evaluation of the evaluation is the last step in a complete process. The CEC study supports this in concluding that evaluations need to have permanent feedback mechanisms built into them so that the process can be revised in response to developing policy interests, research objectives and evaluative methods.[84] The question to ask, then, is: did the evaluation fulfil the purpose for which it was undertaken,

i.e., did it answer the questions, provide the intelligence, demonstrate performance, stimulate performance, etc.? If "no", then a review of the evaluation parameters and redesign of the evaluation is called for.

SUMMARY AND CONCLUSION

Selectivity and concentration in research funding has become unavoidable in Australia today. The environment within which researchers and science policy-makers now operate is one of increased competition for resources, pressure to demonstrate social value for support, and the need for systematic planning and evaluation of research. In the same way that we hear of the 'innovative culture', we may now hear of the 'evaluative culture'. These are hopefully more than just catch-phrases. This paper has attempted to demonstrate the importance of developing such a culture now.

Toward that goal, a working definition of 'research evaluation' that may be suitable within the Australian context was proposed. It was suggested that evaluations will be most productive when motivated by both of two goals: improvement of research performance *and* demonstration of that performance within the policy context.

Six axioms for research evaluation have been proposed. The suggestion is that evaluation will be most productive and constructive when:

1) it is conducted as an ongoing process in which research planning (ex-ante evaluation), interim assessment, and ex-post evaluation are performed in an iterative cycle;
2) participation in the evaluation process is considered as valuable for improving research performance and institutional advancement as any end-product from it;
3) it is undertaken with an explicit objective to produce policy-relevant information;
4) it is systematically planned, performed and documented;
5) it involves an information gathering and analysing component;
6) it is approached as a reseaarch problem in itself.

In the discussion of evaluation parameters, a framework for designing an evaluation has been supplied. The process is first to specify clearly the objectives of the research to be evaluated and identify the conditions and expectations of the decision-making context within which the evaluation is to be undertaken. The purpose(s) of the evaluation then can be defined clearly and agreed upon among the involved parties, and the type of evaluation determined (ex-ante, interim, ex-post, continuous system). These parameters will set the evaluation criteria, which will in turn dictate the factors to be addressed in organising the evaluation and determine the evaluative methods and techniques to be used.

In addition to the parameters, a number of guidelines for evaluation were discussed. These are only a few of those that have been suggested for consideration:

a) a cooperative relationship between those conducting the evaluation and those being evaluated is critical for productive results;
b) multiple or indistinct research objectives will require multiple evaluation criteria and a corresponding array of appropriate evaluation methods and techniques;
c) reliance on a sole information source, evaluative method or technique is often not sufficient, since each has different applications, capabilities and limitations;
d) determining the appropriate basis for comparison of performance is important — e.g., target objectives, past performance, comparison with a similar program or institution, the international state-of-the-art, etc.;
e) use of 'disinterested' expertise, supplementary quantitative tools, and/or professional evaluators may be desirable in order to avoid limitations of perspective and to maintain credibility;
f) involving the intended users of the reseaarch in the evaluation process is particularly important when considering 'strategic' or 'enabling' research; and,
g) optimally, an evaluation procedure will be a 'learning system', retaining the flexibility to incorporate what is learned in the process about the social and intellectual context for the research, the policy objectives, and new approaches to evaluation as these evolve.

Most of the axioms, definitions and guidelines forwarded here are not new. However, few efforts have been made to synthesise and examine the wealth of experience and knowledge that has recently accrued. This is perhaps the sign of a topic area in the early stages of investigation. Indeed, evaluation, as an area of scholarly inquiry, is newly and rapidly developing. It is an area at the borders of several subdisciplines or specialties, drawing on the emerging related literature from the social studies of science, science policy studies, research management studies and social-program evaluation research.[85] It is also a very 'applied' topic, one in which appiications have and will continue, for better or worse, to precede theoretical understanding. So, while recognising the need for more theoretical and empirical research on evaluation, this analysis has taken a somewhat practical, policy-oriented approach. The reason for this is the current need for greater understanding and appropriate application of evaluation in Australia today.

The pressure today is for those who receive public support for research — which directly or indirectly includes almost every major performer of research located in Australia — to be more accountable. The meaning of 'accountability' is perhaps misunderstood. It is less a matter of accounting for each dollar than it is an awareness of the need to make the overall research effort responsive to the "national thrust for advancement."[86] Demonstrating this awareness is part of demonstrating accountability. Accountability, then, is partially a matter of economics (efficiency), partially one of research effectiveness, and partially one of politics, but it also is a state of mind.

The notions of accountability and of an innovative and evaluative culture also have implications for those who set national priorities and directions and determine the allocation of resources. If the Australian government is going to be an innovative "world-class government", as was recently suggested,[87] and if the tax-payer and the research community are going to support and retain confidence in government decisions about selectivity and concentration of funding, then two things probably need to follow.

First, the decision-making process needs to be made more systematic and objective than it perhaps ever has been. It is not sufficient that major priority decisions, such as those concerning strategic IR&D targets by DITAC and selection of national research priorities by the new ARC, be made in an *ad hoc* fashion through informal consultation among a small number of people. Too much is at stake for the nation for priorities to be set on the basis of "advocacy, 'bartering', and 'horse trading' ".[88]

Secondly, and following from the above, the decision-making process needs to be more explicable and transparent. As a case in point: the selection decisions that presumably will soon face the ARC effectively represent a highly visible test of government judiciousness and decision-making integrity. The UGC in Britain failed this same test and has been attempting to repair the situation and its image for the last few years.[89] That the UGC may have made some misjudgements in discriminating between universities is perhaps less critical than the fact that the UGC apparently did not think hard enough about the criteria for the decisions or the need to justify those decisions.

In other words, the very demand for accountability made by government on those receiving public support also applies to government. If Australia is going to develop an evaluative culture (or an innovative culture), then government will have to lead the way by example.

NOTES AND REFERENCES

1. J.S. Dawkins, "The challenge for higher education in Australia", a speech by the Minister for Employment, Education and Training, Canberra: Department of Education, Employment and Training, 22 September, 1987, pp. 11-12.
2. Commonwealth Tertiary Education Commission, *Review of Efficiency and Effectiveness*, Canberra: CTEC, October, 1986..
3. Australian Science and Technology Council, *Improving the Research Performance of Australia's Universities and Other Higher Education Institutions*, A report to the Prime Minister, Canberra: ASTEC, February, 1987.
4. Department of Employment, Education and Training, "Government moves quickly to establish Australian Research Council", A public bulletin, Canberra: DEET, 23 September, 1987.
5. UGC, *A Strategy for Higher Education Into the 1990's: Criteria for Rationalisation*, London: HMSO, University Grants Committee, 1984.

6. S.S. Blume *et. al., Evaluation of Research: Experiences and Perspectives in the Netherlands,* Report on a study commissioned by the OECD Directorate for Science Policy, Ad Hoc Group on University Research, Paris: OECD, 1985.
7. Office of Technology Assessment, *Research Funding as an Investment: Can We Measure the Returns?,* A Technical Memorandum, Washington, D.C.: Congress of the United States, April 1986.
8. M. Gibbons, *Evaluation of Research: Evaluation of Research in Sweden,* Report on a study commissioned by the OECD Directorate for Science Policy, Ad Hoc Group on University Research, Paris: OECD, 1985.
9. Australian Science and Technology Council, *Future Directions for CSIRO,* A report to the Prime Minister, Canberra: ASTEC, November, 1985.
10. See for example: Jane Ford, "Govt research drive disappoints," *Financial Review,* July 17, 1987, p. 53. See also: "ABS survey shows industry R&D growth", in *Laboratory News,* July, 1986.
11. Michael Gibbons and L. Georghiou, *Evaluation of Research, A Selection of Current Practices,* A report prepared for the Secretary-General of the OECD, Paris: 1987, p. 58.
12. For further discussion see: M.G. Taylor, "Evaluation of research and resource allocation", *International Journal of Institutional Management in Higher Education,* 9, 1, March 1985, p. 89.
13. The university referred to is the University of Wollongong.
14. Paul Bourke, *Quality Measures in Universities,* A study commissioned by the Commonwealth Tertiary Education Commission, Canberra, Australia: CTEC, 1986, p.20.
15. Some observations made here about changes in the nature and perception of 'science' apply more to the Physical and Biological sciences than to the Arts and Humanities. The Physical sciences have generally served as the model around which the theories relevant to this paper were developed. The cultural and intellectual role of the Arts and Humanities may be quite different from that of the Physical/Biological sciences and, therefore, deserves separate consideration. This was not possible in this paper.
16. C. Ganz Brown, "The technological relevance of basic research," in B. Bartocha, *et al.* (eds), *Transforming Scientific Ideas into Innovations: Science Policies in the United States and Japan,* Tokyo: Japan Society for the Promotion of Science, 1985, pp. 113-134.
17. F. Narin and E. Noma, "Is technology becoming science?", *Scientometrics,* 7, 3-6, 1985, pp. 369-381.
18. Gibbons *et al., op. cit.,* note 11, p.14.
19. R. Johnston, "Why scientists don't get more money," *Metascience,* 3, 1985, p. 46.
20. Department of Science, *Science and Technology Statement 1985-86,* Tables 5 and 18, Canberra: DoS, November 1986, pp. 14, 88.
21. Department of Science, *Submission to ASTEC Review of Higher Education Research Funding,* Tables 1, 6 and 10, Canberra: DoS, 1986.
22. Australian Science and Technology Council, *Improving the Research Performance of Australia's Universities and other Higher Education Institutions,* Canberra: ASTEC, February 1987, p. 19.
23. P.S. Chen, "Evaluation in biomedical research at the National Institutes of Health", in G. Goggio and E. Spachis-Papasois (eds), *Evaluation of Research and Development,* Proceedings of the European Community Seminar, Brussels, October 17-18, 1983, Dordrecht, Netherlands: D. Reidel, 1984, p. 115.
24. See for example: John Ziman and Peter Healey, *International Selectivity in Science,* A working paper from the Science Policy Support Group, London, SPSG, 1987.
25. Johnston, *op. cit.,* note 19, p. 49.
26. The reader is referred to the literature of the sociology and history of science for more detail. The following provide an introduction to a large and varied literature: K.D. Knorr-Cetina and M. Mulkay (eds), *Science Observed,* London: Sage, 1983; Steven Shapen and S. Schaffer, *Leviathan and the Air Pump: Hobbes, Boyle, and the Experimental Life,* Princeton: Princeton University Press, 1985; Bruno Latour and S. Woolgar, *Laboratory Life: The Construction of Scientific Facts,* Princeton: Princeton University Press, 1979.

27. S.E. Cozzens, "Expert review in evaluating programs", *Science and Public Policy,* 14, 2, April 1987, pp. 71-81.
28. S. Cole, J.R. Cole and G.A. Simon, "Chance and consensus in peer review", *Science,* 214, 20 November 1981, pp. 881-86.
29. A.L. Porter and F.A. Rossini, "Peer review of interdisciplinary research proposals", *Science, Technology & Human Values,* 10, 3, Summer 1985, pp. 34-38.
30. B.R. Martin and J. Irvine, "Assessing basic research", *Research Policy,* 12, 1983, p. 72.
31. Gibbons *et al., op. cit.,* note 11, p. 27.
32. Gibbons *et al., op. cit.,* note 11, p. 26.
33. *ibid.,* p. 10.
34. *ibid.,* p. 46.
35. *ibid.,* p. 57.
36. Thomas E. Clarke, *The Evaluation of R&D Programs and Personnel: A Literature Review,* Ottawa, Ontario, Canada: Stargate Consultants Ltd., December 1986, p. 56.
37. Most of the writing about research evaluation is targeted more for the Physical and Biological sciences than for Arts and Humanities. However, this does not mean that Arts and Humanities research cannot be systematically evaluated in the general way suggested here, only that further consideration than has generally been given is required as to the specific nature of research performance in those fields.
38. Gibbons *et al., op. cit.,* note 11, p. 16.
39. P. Fasella, "The evaluation of the European Community's research and development programmes", in G. Goggio and E. Spachis-Papasois (eds), *op. cit.,* p. 5.
40. J. Irvine and B. Martin, *Foresight in Science: Picking the Winners,* London: Frances Pinter, 1984, p. 141.
41. Bourke, *op. cit.,* note 14, p. 15.
42. One example is Lewis Branscomb, "Industry evaluation of research quality: edited excerpts from a seminar", *Science, Technology & Human Values,* 7, 39, Spring 1982, pp. 15-22.
43. J.A. Snow "Research and development: programs and priorities in a United States mission agency", in G. Goggio and E. Spachis-Papasois (eds), *op. cit.,* p. 95.
44. O.T. Fundingsland, "Perspectives on evaluating federally sponsored research and development in the United States", in G. Goggio and E. Spachis-Papasois (eds), *op. cit.,* p. 100.
45. Daryl E. Chubin, "Designing research program evaluations: a science studies approach", *Science and Public Policy,* 14, 2, April 1987, p. 82.
46. *ibid.,* p. 88.
47. The definition of 'research evaluation' is partially derived from V. Stolte-Heiskanen, "Evaluation of scientific performance on the periphery", *Science and Public Policy,* 13, 2, April 1986, p. 85.
48. Gibbons, *et al., op. cit.,* note 11, p. 19.
49. Gibbons, *et al., op. cit.,* note 11, p. 21.
50. Bourke, *op. cit.,* note 14, p. 23.
51. Fasella, *op. cit.,* note 37, p. 5.
52. Gibbons, *et al., op. cit.,* note 11, p. 46.
53. Fundingsland, *op. cit.,* note 44, pp. 109-11.
54. J. Irvine, B. Martin and G. Oldham, *Research Evaluation in British Science: A SPRU Review,* A paper commissioned by the Centre de Prospective et d'Evaluation, Ministère de la Recherche et de l'Industrie, Paris, France, Sussex: University of Sussex, SPRU, April, 1983, p. 5.
55. For an introduction to the Delphi method see H. Sackman, *Delphi Assessment, Expert Opinion, Forecasting, and Group Process,* US: Rand Corporation, 1974, as discussed in A.L. Porter *et al., A Guidebook for Technology Assessment and Impact Analysis,* New York: North-Holland, 1980, p. 126.
56. J.D. Roessner, "The multiple functions of formal aids to decision-making in public agencies", *IEEE Transactions on Engineering Management,* 1985.
57. One aspect of NIH evaluation activities is exemplified by Francis Narin, *Subjective*

vs. Bibliometric Assessment of Biomedical Research Publications, A US National Institutes of Health program evaluation report, Bethesda, MD: US Departent of Health and Human Services, April, 1983.

58. IDEA Corporation, *A Comparison of Scientific Research Excellence at Selected Universities in Ontario, Quebec and the United States, 1982,* A technical background paper for The Commission on the Future Development of the Universities of Ontario, Ontario: IDEA Corporation, September, 1984.
59. Blume, *op. cit.,* note 6, p. 10.
60. One of the many examples of US NSF investigations: M.P. Carpenter, *Updating and Maintaining Thirteen Bibliometric Data Series Through 1982,* A final report to the US National Science Foundation, Science Indicators Unit, New Jersey: Computer Horizons, 19 November, 1985.
61. H.R. Coward, J.J. Franklin and L. Simon, *ABRC Science Policy Study: Co-Citation Bibliometric Models,* Final report to the Advisory Board to the Reseaarch Councils of the United Kingdom, Philadelphia: Center for Research Planning, July, 1984.
62. Royal Society Policy Studies Unit, *Evaluation of National Performance in Basic Research — A review of techniques for evaluating performance in basic science, with case studies in genetics and solid state physics,* ABRC Science Policy Studies, No. 1 performed for the Economic and Social Research Council, London: Department of Education and Science, 1986.
63. H.F. Moed, W.J.M. Burger, J.G. Frankfort, A.F.J. van Raan, "The use of bibliometric data for the measurement of university research performance", *Research Policy,* 14, 1985, pp. 131-149.
64. J.J. Franklin, H.R. Coward, and L. Simon, *Identifying Areas of Swedish Research Strength: A Comparison of Bibliometric Models and Peer Review Evaluations in Two Fields of Science,* Final report to the National Swedish Board for Technical Development, Philadelphia: Center for Research Planning, 23 April, 1986.
65. Referred to in B.R. Martin and J. Irvine, *Final Report on the Three-Year SPRU Programme on Research Evaluation by the Leverhulme Trust,* Sussex: University of Sussex, SPRU, November 1986, p. 17.
66. H.F. Moed, W.J.M. Burger, J.G. Frankfort, A.F.J. van Raan, *On the Measurement of Research Performance. The Use of Bibliometric Indicators,* Leiden, Netherlands: The University of Leiden, 1983.
67. F. Narin, *Measuring the Research Productivity of Higher Education Institutions Using Bibliometric Techniques,* Report to the OECD Workshop on Science and Technology Indicators in the Higher Education Sector, 10-13 June, 1985, Paris: OECD, 20 May, 1985.
68. Raphael Gillett, "No way to assess research", *New Scientist,* 30 July, 1987, pp. 59-60.
69. Martin *et al., op. cit.,* note 30.
70. J.J. Franklin and R. Johnston, "Co-citation bibliometric modelling as a tool for S&T policy and R&D management: issues, applications, and developments", forthcoming in A.F.J. van Raan (ed.), *Handbook of the Quantitative Study of Science and Technology,* Amsterdam: Elsevier, 1987-88.
71. M. Callon, S. Bauin, J-P. Courtial and W. Turner, "From translation to problematic networks: an introduction to co-word analysis", *Social Science Information,* 22, 1983, pp. 191-235.
72. L.A. Myers, "Information systems in research and development: the technology gatekeeper reconsidered", *R&D Management,* 14, 4, 1984, pp. 199-206.
73. N. Cooray, "Knowledge accumulation and technological advance", *Research Policy,* 14, 1985, pp. 83-95.
74. S. Ghoshal and S.K. Kim, "Building effective intelligence systems for competitive advantage", *Sloan Management Review,* 49, Fall 1986, pp. 49-58.
75. Gibbons *et al., op. cit.,* note 11, p. 10.
76. Johnston, *op. cit.,* note 19, p. 52.
77. M. Gibbons and L. Georghiou, *Evaluation of Research: Evaluation of Research and Development in the United Kingdom,* Report on a study commissiond by the OECD Directorate for Science Policy, Ad Hoc Group on University Research, Paris: OECD, 1985, p. 19.

78. For related discussion see Ken Green, "Research funding in Australia: a view from the North", *Prometheus,* 1, June 1986, p. 85.
79. Stephen Hill, "From dark to light: seeing development strategies through the eyes of S&T indicators", *Science and Public Policy,* 13, 5, October 1986, pp. 275-84.
80. Gibbons *et al., op. cit.,* note 11, p. 24.
81. Gibbons *et al., op. cit.,* note 77, p. 31.
82. Blume *et al., op. cit.,* note 6.
83. Chen, *op. cit.,* note 23.
84. Fasella, *op. cit.,* note 39, p. 5.
85. The only one of these subdisciplines that has perhaps not been referenced here is social evaluation research. See L. Rutman and G. Mowbray, *Understanding Program Evaluation,* Beverly Hills: Sage, 1983 or Marvin C. Alkin, *A Guide For Evaluation Decision Makers,* Beverly Hills: Sage, 1985.
86. Hugh Preston, "The new Australian Research Council — its objectives, structure and implications", A speech given by the Assistant Secretary of the Research Grants Branch, DEET, University of Wollongong, 14 October, 1987.
87. Terry Hillsberg, an untitled speech given at the conference "Innovation Outlook '87" by the First Assistant Secretary of the Technology and Business Efficiency Division, DITAC, Sydney, 17-18 September, 1987.
88. J. Ronayne, *The Allocation of Resources to Research and Development: A Review of Policies and Procedures,* A report to the Australian Science and Technology Council, Canberra: ASTEC, 1980, p.iv.
89. For discussion see Bourke, *op. cit.,* note 14, pp. 4-5.