University of Wollongong
Research Online

Faculty of Commerce - Papers (Archive)

Faculty of Business

2000

International trends in evaluating university research outcomes: what lessons for Australia

Samuel Garrett-Jones
University of Wollongong, sgarrett@uow.edu.au

Publication Details

Garrett-Jones, SE, International trends in evaluating university research outcomes: what lessons for Australia, Research Evaluation, 8(2), 2000, 115-124.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

International trends in evaluating university research outcomes: what lessons for Australia

Abstract

An international study compared methods used to monitor and evaluate the outcomes of university research in the United States, Canada, the Netherlands and elsewhere. It aimed to provide a foundation for improving the evaluation of research and research training in Australian universities. Evaluation methods were considered in terms of their audience, the type of outputs, outcomes or impacts being measured, and the types of research funding support schemes to which they were applied. The study found that Australian research agencies are generally in line with 'common practice' in the countries studied, and in some cases in advance of it. The study also identified six main areas where Australia could learn from overseas experience.

Keywords

lessons, australia, international, evaluating, trends, outcomes, research, university

Disciplines

Business | Social and Behavioral Sciences

Publication Details

Garrett-Jones, SE, International trends in evaluating university research outcomes: what lessons for Australia, Research Evaluation, 8(2), 2000, 115-124.

This journal article is available at Research Online: http://ro.uow.edu.au/commpapers/821

Research Evaluation, volume 8, number 2, August 2000, pages 115–124, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England

University research outcomes

International trends in evaluating university research outcomes: what lessons for Australia?

Sam Garrett-Jones

An international study compared methods used to monitor and evaluate the outcomes of university research in the United States, Canada, the Netherlands and elsewhere. It aimed to provide a foundation for improving the evaluation of research and research training in Australian universities. Evaluation methods were considered in terms of their audience, the type of outputs, outcomes or impacts being measured, and the types of research funding support schemes to which they were applied.

The study found that Australian research agencies are generally in line with 'common practice' in the countries studied, and in some cases in advance of it. The study also identified six main areas where Australia could learn from overseas experience.

Associate Professor Sam Garrett-Jones is Deputy Director of the Centre for Research Policy at University of Wollongong, NSW 2522, Australia; Tel: +61 2 4221 3256; Fax: +61 2 4221 3257; E-mail: Sam_Garrett-Jones@uow.ed.au.

This is a revised version of a paper presented at the conference Data and Strategies in Evaluating Research and Development held at the Australian National University, Canberra, Australia on November 15–16, 1999.

THE MAIN FINDINGS of an international comparative study of methods used to monitor and evaluate the outcomes and other 'results' of university research (Turpin *et al*, 1999) is summarised. The study was designed to provide a foundation for improving the evaluation of the outputs, outcomes and impacts flowing from research and research training in Australian universities.[1]

At one level, the policy and evaluation issues faced by Australia differ little from those in other developed countries. The study was driven by a desire to track the immediate outputs of research funded within the Australian higher education system; to consider the extent to which research funding is building high-quality, world-class research activities and capabilities in universities; and to assess whether society is receiving adequate returns on its investment in higher education research and research training.

On the other hand, the organisation of research in Australian universities has changed radically over the last decade. With the 'unified national system' and institutional amalgamations, the university research system is now considerably larger than in 1988. Competition for 'traditional' funds, such as the Australian Research Council (ARC) Large Grants, has become intense. There is debate about the role of the larger research universities and the smaller institutions, which make an important contribution to the national research effort in particular fields. The system is being driven towards more explicit social and economic outcomes. There has been a large increase in the funding for research grants and scholarships linked to industry and other 'users' — as evidenced by the growth of the ARC's 'SPIRT' scheme. Increasingly, university research is being organised in application-oriented research centres — such as the Cooperative

 0958-2029/00/020115-10 US$08.00 Beech Tree Publishing 2000

Research Centres (CRCs) and Key Centres — that serve to coalesce traditional research disciplines. As a consequence, the public expectations of academic research have changed considerably (Turpin and Garrett-Jones, 2000).

The environment in which university research is evaluated has changed too. Practice over recent years has required an increased degree of accountability from researchers, universities and research funding bodies. Australia now has a substantial body of material that evaluates university research outcomes at the level of project grants, disciplines, centres, institutions and funding programmes, with the opportunity to review and integrate this work. The study was also timely in the face of proposed changes to the higher education research and research training system in Australia (Kemp, 1999), and the implications these may have for academic research evaluation.

Aims and methodology

The approach used in the international comparison was both practical and pragmatic. It essentially asked what evaluation techniques were used by overseas research councils; which seemed to be more effective in particular circumstances; and whether they would also be appropriate within Australia's university research system. It is important to emphasise that the study aimed to *learn* from others, not simply to copy from them. The analysis recognises the differences between countries both in terms of the organisation of their higher education systems and the sheer size.

> In the United States, for example, the study found no fewer than three evaluations of the discipline of mathematics during 1998: one by the Committee of Visitors for the National Science Foundation (NSF) Division of Mathematical Sciences (using the Government Performance and Results Act [GPRA] assessment criteria); a second by an NSF Senior Assessment Panel on the international standing of US mathematical sciences; and a third international benchmarking study carried out by the National Academy of Sciences' Committee on Science, Engineering and Public Policy (COSEPUP), the policy committee of the national academies (National Academy of Sciences, 1998a). The plurality of approaches and 'triangulation' that these provide is clearly a strength, though perhaps a luxury that only the largest research systems can afford.
>
> In the United Kingdom, the Research Assessment Exercise (RAE) provides quality rankings for academic research at the university department level. Carried out by the Higher Education Funding Councils, the RAE distributes research funds primarily on the basis of quality judgements by expert panels. The RAE is expensive but the Funding Councils distribute very substantial research funds on the basis of the ratings (Garrett-Jones and Aylward, 2000).

Table 1. Main research councils and agencies investigated

Australia
- Australian Research Council
- Cooperative Research Centres Programme
- National Health and Medical Research Council
- CSIRO and selected R&D Corporations

Canada
- Medical Research Council
- National Research Council
- National Science and Engineering Research Council
- Social Science and Humanities Research Council
- Statistics Canada

European Union
- European Commission, Science R&D (Brussels)

The Netherlands
- KNAW – Royal Netherlands Academy of Arts and Sciences
- NWO – Netherlands Organisation for Scientific Research
- STW – Dutch Technology Foundation
- VSNU – Association of Universities in the Netherlands

Sweden
- FRN – Council for Planning and Coordination of Research

United Kingdom
- ESRC – Economic and Social Research Council
- Office of Science and Technology

United States of America
- National Academy of Sciences (COSEPUP)
- National Institutes of Health
- National Science Foundation

The key tasks for the study were: first, to describe and analyse the evaluation processes of a sample of overseas research councils and comparable funding agencies; and second, to identify and describe examples of what the authors considered 'better practice' in evaluating research. It also identified 'common practice', and 'breakthrough' approaches internationally. The study was also particularly concerned with how the overall evaluation practices of organisations are integrated with performance management; for example, on how programme-level performance measures are used in top-level management and priority setting.

A dual strategy was followed for data collection. First, five international research experts were commissioned to report on developments in research evaluation in their respective countries.[2] Second, in collaboration with these experts, the authors identified research funding agencies for follow-up interviews. The main agencies investigated are shown in Table 1.

Although necessarily selective, the international interviews and case studies covered a comprehensive range of the most highly regarded academic research funding agencies in the world, such as NSF in the United States, NWO in the Netherlands and ESRC in Britain. The case studies included several agencies with a mandate for application of research, such as the Dutch Technology Foundation (STW), the National Institutes of Health in the United States, and Canada's

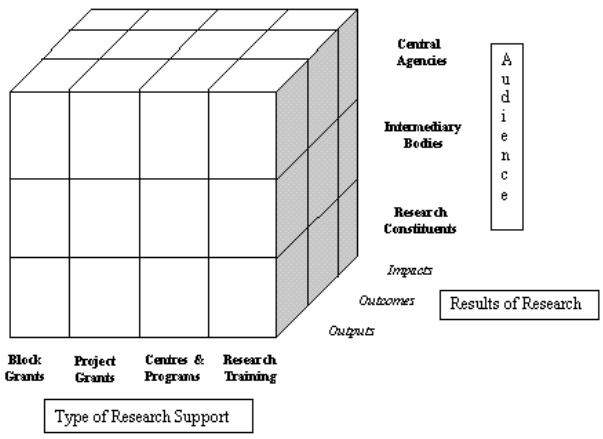

Figure 1. The 'Evaluation Cube'

National Research Council. The study team also held discussions with coordinating and professional bodies with a strong interest in evaluation.

Analytical framework

While not aimed at a theoretical analysis of evaluation methodologies, the study required a conceptual framework for the international comparison that considered the political context of the evaluation (its audience or purpose), the sort of research 'results' being considered, and the type of research support mechanism for which the evaluation was appropriate. It may be useful to consider the relation between these three sets of factors in terms of an 'evaluation cube' (Figure 1) where the *x* axis is 'type of research support', the *y* axis is 'audience' or level of evaluation, and the *z* axis is 'type of research result' being assessed, from outputs of publications to broader socioeconomic impacts.

Alternatively, the *x* axis could be reduced to 'basic research' (aimed at advancing knowledge) and 'applied research' (having likely practical use). But 'type of research support' was more useful in practical terms, since the study wished to make comparison between, for example, the evaluation of industry-university collaborative research centres in the US with that of the CRC programme in Australia.

In relation to the timing of the evaluation — i.e. whether the evaluation is *ex ante*, 'real time' or *ex post* — the emphasis on outcomes led to a focus on *ex post* evaluation practices, but the study also looked at *ex ante* evaluation issues where relevant.

The authors contend that these parameters taken together largely determine or at least constrain the sort of evaluation methodology that is appropriate. Indeed, a major conclusion of the study is that:

> 'Many of the methods available for identifying and quantifying the outcomes and impact of research are useful in particular circumstances. Evaluations of research outcomes are carried out for different audiences and for different purposes. The approach for a particular evaluation is usually to adopt a methodology that matches the driving force behind the evaluation (by including evidence that is familiar to the audience), the purpose of the evaluation and the use to which it will be put.'

Table 2. Summary of typical evaluation mechanisms by type of research support

Type of research support	Typical-practice evaluation mechanism	
	Ex ante (assessment for selection)	*Ex post* (including mid-term/ ongoing)
Block grants to institutions	• Bibliometrics • Institutional profiles	• Bibliometrics • Modified peer review • User panels • Self-evaluation • Commissioned evaluation studies • Benchmark comparisons – national and international • Disciplinary reviews
Project grants	• Peer review	• Peer review • Sector evaluation panels • Researcher 'end of term' reports • Publications and 'user' related outputs
Research centres and larger programme grants	• Modified peer review • Bibliometrics	• Modified peer review • Self-assessment • Bibliometrics • Benchmark institutional comparisons • Disciplinary reviews • User surveys • Alliance monitoring
Research training and fellowships	• Peer review	• Postgraduate completions • Career tracking • Publications outputs • Scheme-based evaluations

In theory, each of the 36 cells of the cube implies a somewhat different selection of evaluation approaches. In practice, there is much overlap and gradation in the evaluation methodologies appropriate in each case and the study did not attempt a strict cell-by-cell comparison. However, in developing evaluations of academic research, it is useful to consider 'where within the cube' the evaluation lies (see Table 2).

To take two examples:

> An evaluation of the results of a large project grant by a research council is likely to focus on the

quality of the research as measured by international standards, and the broader recognition that the research has received. It will commonly use counts of publications in prestigious journals, honours and prizes awarded to the investigators, and community or commercial recognition of the value of the research.

At the opposite corner of the cube, an evaluation of particular research programmes for a central budgetary agency would concentrate on 'return on investment' and beneficial social impacts. For outcomes, this might mean measuring 'user benefit or satisfaction' with the programme, or a comparison with the outcomes from similar programmes run by other agencies.

Research results

The terminology of research results evaluation is by no means standardised and is continuing to develop. Most authors agree that the term 'results' covers the spectrum of outputs, outcomes and impacts. But others regard 'outcome' as a more general term than 'impact', or as synonymous. The study used the following working definitions:

Outputs are the routine *products* of research activity, which may include publications, conference papers, data sets, training courses and research degrees, etc.

Outcomes are the *achievements* of the research activity, whether conceptual (a new theory), practical (a new analytical technique) or physical (a new device or product – although some authors regard this as an output). Research outcomes are potentially available for *use*.

Impact is a measure of the *influence* or *benefit* (either realised or expected) of the research outcomes, either within the research community itself (through advancement of knowledge), or in the wider society. The economic, social or environmental benefit to the community may be termed 'non-academic impact' — some authors use the term 'impact' only in this sense. Impact measures the scale, effects or implications of use.

As one moves from outputs to impacts, the results of research activity are generally broader in their effect, take longer to manifest themselves, are harder to quantify and are less readily traceable to particular research projects, funding programmes or agencies (the 'attribution factor').

As one moves from outputs to impacts, the results of research activity are generally broader in their effect, take longer to manifest themselves, are harder to quantify and are less readily traceable

The audience for the evaluation

A further observation is the need to tailor the evaluation to the audience, and to use the sorts of evidence that they are likely to find convincing. The 'evaluation cube' identifies three broad 'levels' of audience.

Central agency evaluation audiences

In most countries there are strong demands from government budget agencies to justify allocations of public funds to higher education research. At this level, governments are seeking socioeconomic justifications for their investments in research, with two basic imperatives for evaluation: economic and political. The economic imperative requires justifying budget applications to treasury and justifying allocation decisions made within treasury.

The political imperative concerns the broader public perceptions of the value of research investments to the community. This broader audience is seeking assurance that public investments in higher education research 'pay-off'.

Typically, 'good news stories' of research outcomes or achievements serve to provide such reassurances. However, 'stories' are not sufficient to satisfy the demands of treasury officials. This audience is increasingly seeking accountability in terms of 'performance monitoring'. This requires carefully documented definitions of agency objectives, strategies and performance indicators.

In countries such as Canada and Australia each department must provide an annual performance report to parliament on strategies and progress toward stated objectives. In Australia, familiar examples include the strategic plans of the Research and Development Corporations, and the performance indicators included in the triennial funding agreements for federal research agencies.

The Canadian Treasury Board requires each department to state its 'commitments to Canadians'. There, the Medical Research Council (MRC) has given four such 'commitments', with broad (though not quantifiable) indicators for each. The MRC worked on a 'large inventory of indicators' and consulted with the medical research community about them.

In the UK, the Office of Science and Technology has been pressing the six main research councils to develop common output measures and performance indicators against the objectives of: research; research training; industrial competitiveness; quality of life; and promotion of science.

Box 1. The US Government Performance and Results Act

In 1993, the US Congress passed an act called the 'Government Performance and Results Act' (GPRA, or the Results Act). It closely resembled accountability legislation already in place in a number of other countries. Crafted by analysts with long experience in formal programme evaluation, the Results Act aimed in part to overcome the limitations of formal programme evaluation by supplementing it with short-term performance monitoring. The Act required each agency to submit three documents to Congress: a strategic plan, covering all major agency functions over a minimum five-year period; a performance plan, setting specific, quantified target levels of performance for a particular fiscal year; and a performance report, giving actual versus targeted performance. The Act was one of several related pieces of legislation linking budgets to performance, and the GPRA reports were gradually to be integrated into the regular flow of budget and audit submissions to Congress.

Source: Cozzens, 2000

In several countries, *government legislation* is providing the driving force for systematic agency performance evaluation. The South African *Reporting by Public Entities Act* applies to the Foundation for R&D in that country. A better known example is the *Government Performance and Results Act* (GPRA) in the United States (see Box 1).

The 1993 *Results Act* and the implications that followed for the US National Science Foundation illustrates the impact that formal legislative requirements are having on research evaluation reporting to audiences at this level (Cozzens, 2000). While there was clearly initial resistance to application of the GPRA to an academic research funding agency, the influential COSEPUP has reported its conviction that both applied and basic research can be evaluated in a meaningful way, provided the right measures are used (National Academy of Sciences, 1998b).

One of the implications of centralised reporting demands enshrined in legislation, such as GPRA, is to force academic research funding agencies to adopt regular (and to some extent common) reporting mechanisms. Perhaps the major challenge, however, is to develop a systematic approach for *integrating* the results of different evaluation approaches so that this can inform public policy at this 'top level'.

Intermediary body evaluation audiences

The Dutch evaluation system (Van der Meulen and Rip, 1999) is characterised more by a delegation of responsibility for evaluation to intermediary organisations such as the funding foundations (the research council — NWO, and Royal Netherlands Academy — KNAW) or the Association of Dutch Universities (VSNU). In this case the evaluation audience is generally the governing boards or the programme directors. A consequence is that there is a greater focus on the management of research programmes, of which quality of outcomes is one aspect. An advantage of the less centralised system is that definitions of value of research are not subject to sudden changes resulting from political ideology. Rather, they remain and evolve with the priorities and objectives of the agency. This is not to suggest that agency priorities are not linked to national priorities, but rather that they act as intermediaries between government and research.

The 'Dutch Model' is applied, with variation, to institutional and discipline-base evaluation by a wide range of agencies in the Netherlands. The salient features of the model are:

evaluation by self-assessment combined with a visiting committee of international peers;
the existence of a 'protocol' for the evaluation, defining aims, criteria, content of the self-assessment and the information base for the evaluation, expertise of the committee, responsibilities of the actors involved and the form of publication of the results;
evaluations serve a reflective rather than a prescriptive purpose; and
evaluations cover performance over a set period of 4–5 years.

Unlike the United Kingdom's RAE, evaluations (such as the discipline evaluations of the VSNU) do not directly influence government resource allocations to the universities. Rather they are influential in helping the research groups and universities to manage themselves. They are, however, leading to a stratification in the university system as universities reward (and protect) the best-performing research groups. To this extent, the open methodologies of protocols and self-assessments support local research management rather than centralised decision making.

Research constituents as an evaluation audience

Apart from the government agencies responsible for funding research programmes there are institutions and 'user groups' that comprise a further level of research evaluation audience. This audience includes universities, researchers and research users from business and the broader community. In some countries, industry sector groups have been established to assess the quality and 'useability' or 'transferability' of research outcomes. In cases such as the Swedish and Dutch foundations, modified peer-review panels have been established to review research evaluation reports. For these audiences the technical content and transferability of output of research groups or projects are of particular concern rather than the broad socio-economic impact of the programme.

Type of research support

Institutional evaluation/block grants

The balance between block funding and other sources of research funding tends to determine the level of effort invested in 'institutional evaluation'. Experience

in the US suggests a lesser role for evaluation of research outputs at the institutional level than in other countries, at least in the university sector. As Cozzens (2000) points out, questions of the evaluation of university units, which constitute at least half the assessment problems in many other countries, do not arise at national level in the USA. Instead, the federal government has evolved a strong quality control system and 'extraordinarily effective' *ex ante* selection system for the research it funds in the form of the peer-review system for competitive project grants. This reduces the demand for *ex post* evaluation, Cozzens concludes. Even so, research funding agencies have made use of institutional output data, at least for the research they fund.

> The National Institutes of Health (NIH) are the largest federal funder of academic research in the USA. In the 1970s, the NIH pioneered the collection of publications data by institutions and its evaluation office produced a series of reports on the quantity and character of publications ('basic' or 'clinical'), institute by institute. By the 1980s, this exercise proved too expensive and was cancelled.

One of the obvious problems in the evaluation of the research outputs of institutions and research groups is how to compare the results for different institutions or groups. Generally this has been done by simplifying and standardising the evaluation process as far as sensible.

> A good example of this 'mass evaluation system' is the Research Assessment Exercise for universities in the UK. In theory, RAE panels are primarily concerned with making quality judgements based on selective reading of the listed research outputs as well as other evidence of reputation and track record. In practice, as Bourke (1997) commented, 'relatively little first order evaluation can occur … except in the smaller fields'. As a result, 'panels rely heavily on *surrogates*, of which the most important is … the evaluation processes of scholarly journals and book publishers'.

Project grants

The assessment of the worth of fundamental research project proposals remains the domain of *ex ante* selection by peer-review panels, based on the track record of the principal researcher or research group.

> The US federal government, through agencies such as the NSF, runs a 'classic' model of *ex ante* research project evaluation and selection, described by Cozzens (2000), which is well known and well regarded. This *ex ante* assessment at the project level is complemented by a series of broader *ex post* evaluations, both regular and *ad hoc.* NSF Committees of Visitors (COVs) audit the project peer-review process every three years. Their main concern is the integrity and efficacy of the proposal process, but Committees are now turning their attention increasingly towards programme results, integration with the NSF's long-term strategies and the requirement of the GPRA.
>
> The UK Economic and Social Research Council is one of the few research councils studied that carries out formal evaluation of the outcomes of all the research projects it funds (Garrett-Jones and Aylward, 2000). Research projects are evaluated on the basis of an 'end of award' report provided by the researcher. The report is assessed usually by two to four independent peer 'rapporteurs' (more if there is disagreement) and given a grade on a four-point scale.

Research centres and larger research programmes

In one way the outputs and outcomes of formal research centres and programmes are easier to evaluate than is fundamental, project-based research. The research carried out in centres and programmes is commonly 'applied' in nature and the centres usually have agreed social, economic or technological objectives against which their performance can be evaluated. Significant effort is spent in Australia and internationally on *ex post* evaluation of research centres and programmes against their nominated objectives and outcomes. On the other hand, specific objectives vary considerably across centres and programmes and require a range of different evaluation measures and approaches.

The evaluation is tailored to the goals of the centre or programme. Again, the preferred means of evaluation of performance is the expert peer-review panel. Given the large size of the research centres, this is likely to be very senior, with international representation. More importantly, it will include significant non-academic partner or 'user' representation. A wide range of methodologies is employed to augment the central 'modified peer-review' process: bibliometrics, patent counts and network tracing. Benefit-cost analysis may be used in limited cases.

The study found two main issues in these evaluations, common across several countries. First, how to ensure that the 'reach' of the programme is covered and the actual or potential beneficiaries of the research are recruited into the evaluation process; and second, how to achieve a comparison (preferably quantitative) of the outcomes and impacts of quite different centres or programmes.

> The National Research Council of Canada (NRC) operates many public research laboratories. NRC's general evaluation approach is to examine 'Resources-Reach-Results'. This has been achieved by strengthening the 'modified peer-review' processes already in place. Increasingly, stakeholders other than researchers (e.g. industrialists) are co-opted as peer reviewers. The review committees receive both oral and written terms of reference from the Council. Typically the evaluation involves peer

review (a two-day intensive meeting), benchmarking against comparable UK/European/US/Asian groups, socioeconomic analysis/case studies and client responses. There is a five-year rolling cycle of programme evaluation.

Research centre evaluations have made significant use of quantitative indicators of outputs or outcomes — publications quality through citations analysis (especially in engineering and natural sciences), patents granted, and career paths of researchers trained in the centres. But these indicators support, rather than lead, an evaluation that rests on the expertise of the 'blue ribbon' panel. Yet there is some indication in Australia and internationally of the growing importance of quantitative measures of outputs and outcomes, and of common measures that allow some comparison between different centres, research programmes and research council schemes with similar objectives. Five or seven years ago, it was possible to find expert panel reviews using few or no quantitative output indicators. Today, such a review report seems naked without them, and in many cases they contribute to the 'system level' performance indicators that are increasingly being gathered by research councils.

Research training

All the countries studied collect considerable data on the inputs and outputs of research training activities (numbers of postgraduate research students, postdoctoral fellows, PhD completions, qualifications and national origin of scholars, etc.) and had done so for many years. Most have used these data, augmented by surveys of students, fellows and institutions, to gauge the outcomes and impact of their investment in these training programmes and to make management changes to the way the programmes operate. In Australia, as elsewhere, these evaluations have tended to be 'one-off' surveys, and as a consequence it has been difficult to make comparisons between them.

Within industry-university centres in the United States, centre monitoring systems typically contain basic information about students and some outputs including publication rates (Gray, 1999). The Engineering Research Center Program's monitoring system attempts to maintain current records on graduates and their employers in order to facilitate post-graduation follow-up studies. Typical output/outcome assessments for students are similar to centre technology transfer assessments in that they involve collection of survey data from the target audience (students). Surveys usually ask students to retrospectively evaluate the quality of the training they received and how well prepared they are to carry out their jobs. They also ask students about outcomes like promotions and publications. A study carried out by the Stanford Research Institute reported results for students in industry, academic and government employment studies.

In relation to Canada, Holbrook (1999) comments that, in the area of human resources there is a desperate need for evaluation past the simple counting of graduates. StatCan and Human Resources Development Canada are working together on this, using the services of university alumni associations. Currently there are two graduate careers evaluations under way. In evaluating the supply and demand for scientists and engineers, by looking at career tracking of scientists and engineers, StatCan is limited by Canada's privacy laws. However, the regular Census gives information on level of qualification, salary and field of study. For example, unemployment rates by field of study are available. There is also a National Graduate Survey (two and five years out from graduation) that shows how graduates are moving into and around the job market by industrial sector and field of study.

Lessons for Australia

The study found six main areas where, the team believes, Australia can learn from the overseas experience in evaluating academic research outputs, outcomes and impacts. This is not to imply that the evaluation of university research in Australia is especially deficient — indeed the study showed that Australia meets or exceeds 'common practice' in many areas of evaluation. Rather, in these areas, international practice and/or theory is moving ahead and should be taken into account in any future Australian strategy for the evaluation of academic research. The six 'themes' are summarised below.

Evaluation for quality: strengthening peer review and national and international benchmarking

The study emphasises the strong primacy of peer-review panels for the evaluation of the quality of most academic research. For the most significant reviews, internationally regarded experts are invariably used. Panels use a mix of qualitative and quantitative evaluation methods. Where quantitative techniques like bibliometrics are used, it is to inform the peer-review process; for example, in 'mapping' a field. Agencies may develop 'guidelines' or 'toolkits' to assist their peer-review panels and to achieve some conformity of approach to the evaluation of different disciplines and institutions. For academic research, a wider range of outputs is being taken into account in the evaluation process, and these are being more systematically collected over a longer period than previously.

There is scope in Australia, as elsewhere, to develop the use of evaluation methods and indicators that allow greater comparability between disciplines and between institutions, while recognising their unique features. One possible approach is to use inter-institutional and international benchmarking of the quality of research outputs. For

Evaluation of research outputs is no longer seen as a one-off exercise. Agencies are developing a 'learning loop', linking *ex ante* project selection, *ex post* discipline reviews and ongoing programme monitoring

example, one might benchmark performance in the social sciences across institutions, using a kind of 'mini- RAE' process.

Agencies are using modern communications systems to strengthen the peer-review process by reducing the cost of using national and international experts.

For example, the COSEPUP review of mathematics research was carried out by a panel of very senior US and foreign academics and industrial researchers. The study was experimental with a relatively small budget. Much of the work was done through electronic networking, video-conferences and the like.

Evaluation as an integral input to agency planning and national priorities

Evaluation of research outputs is no longer seen as a one-off exercise. Agencies are developing a 'learning loop', linking *ex ante* project selection, *ex post* discipline reviews and ongoing programme monitoring. In turn, these results are 'processed' through the 'evaluation loop' and feed into agency strategic planning. This trend is driven in part by a strong push at the national level to require common 'performance indicators' for research councils and other funding agencies. These indicators may have application in national budget negotiations and in priority-setting for research. A challenge is to articulate the 'agency level' indicators with the outcomes of the evaluation of grants and programmes.

Many agencies included in the study have attempted to learn, through *ex post* evaluations, of the effectiveness and impact of their *ex ante* assessment mechanisms and processes. Better practices, however, have developed more systematic approaches for ensuring that *ex post* analyses are linked to ongoing refinements in *ex ante* processes. In some cases this has been approached simply through the dissemination of reports to research assessment committees. More effective learning processes have included representatives of *ex ante* assessment committees in the *ex post* evaluation process.

In respect of learning between institutions and agencies, the study noted the role of intermediary organisations like research councils and the use of standardised evaluation reporting or indicators, in some cases driven by a legal requirement to provide them.

Normative or formative? — evaluation as a complex tool in the performance management of research

The better evaluations, the study concluded, are a productive discourse between the funding agency and the research institutions/groups. They are rarely normative, nor do they have a direct and immediate consequence for research funding. Rather, they inform the management of the research process on both sides. They employ agreed 'protocols' and are 'owned' by both parties.

The Dutch experience — based on a longer experience with formal institutional or discipline evaluation than exists in Australia — shows the value of having a clearly agreed protocol for the evaluation from the outset. This is helped by having a 'menu' of indicators or criteria of evaluation that are commonly used by evaluation committees. The parties may then select those most relevant to the institution or discipline. But in emerging areas, such as socioeconomic impact, what is perhaps more important is stimulating learning processes that allow improved practices to evolve over cycles of evaluation.

This sharing of experience does not appear as ingrained in Australian academic evaluation as it is in the Netherlands. The 'Dutch model' of cooperative evaluation appears to deliver a formative result for both the evaluator and those subject to the evaluation, and encourages learning and application of the results of the evaluation.

Better measures of S&T human capital

There is considerable interest in measuring the human capital impact of research funding. Several research agencies are attempting to extend work on the outcomes and impacts of research training by examining parameters such as career track, salary status, and type of work (whether still in a research position) for postgraduates for some years following graduation. There are both practical and legal problems in carrying out this kind of study.

All research activity contributes in some way to building the national science base. A conscious effort is now being made in several countries to quantify these 'system impacts' for different types of research. S&T human capital studies attempt to capture not only the quantity but also the quality of individual researchers and the 'capacity' of teams and networks of researchers.

Bozeman and Rogers (1998) present a breakthrough approach to the human capital building outcomes of research. In particular, they emphasise the broad measurement of scientific, technical and human capital as a legitimate indicator of outcomes for a wide

range of R&D programmes. These authors point out that many (non-academic) government R&D programmes — even the more applied ones — are also concerned with 'nurturing science'. This, they argue, should lead to a different 'balanced portfolio' approach which integrates a focus on discrete outputs with an emphasis on maintaining and extending the research community's capacity, especially its scientific and technical human capital. These authors define S&T human capital as 'the sum total of scientific and technical knowledge and social skills embodied in a particular individual'. This comprises not only the formal educational qualifications encompassed in traditional human capital studies, but skills, know-how, 'tacit knowledge', experimental knowledge, and social and research management skills. 'The knowledge of how to manage a team of junior researchers … is part of human capital. Knowledge of the expertise of other scientists (and their degree of willingness to share it) is part of S&T human capital' (Bozeman and Rogers, 1998, page 6).

> Clearly, measuring S&T human capital thus defined presents challenges for the evaluator. However, Australia is quite well placed by virtue of its experience with career track evaluation and its existing agency expertise in data collection in human resources to develop and employ measures of the human capital outcomes and impacts of academic research training.

Reaching the 'non-academic' users of research

The study also found much interest in measuring the 'non-academic' or socioeconomic impacts of research. For some industrial technology research, this can be done using various benefit-cost analyses, at least to measure the 'direct' benefits. In other applied research, attempting to identify the 'reach' or potential users/stakeholders of the research is becoming a central tenet of the evaluation of research outcomes and impacts. International experience provides examples of how 'users' are identified and recruited to the evaluation process, and the circumstances where this is appropriate.

> In the UK, the ESRC has been looking at 'networks and flows', user panels and tracing post-research activity (e.g. career paths) as possible methodologies for assessing research impact on 'non-academic audiences'.

Capturing research outputs and outcomes in the information age

All research funding agencies are recognising that electronic information systems and the Internet can make the collection of data on research outputs and outcomes easier and more systematic. On the other hand they also entail significant investment in systems and procedures and there may be risks in unnecessarily 'integrating everything'. Overseas agencies provide models of successful systems, some of which are 'at arm's length' from the research council. Other models are being developed as an integral component of agencies' assessment and monitoring systems.

> The ESRC systematically collects a wide range of information on the outputs and outcomes of the research that it funds through the Research Grants Archive and Database (REGARD). REGARD covers publications, and other outcomes such as consultancies, media appearances and academic honours and allows researchers to submit or to access information via an Internet website.[3]

Conclusions

Australian experience with 'results' evaluation for academic research has been quite extensive. Formal evaluation of research outputs and outcomes is now common at many levels. However, practical application of such evaluation to higher education research policy has been patchy. For example, there has been limited coordination across agencies in developing system-wide evaluations.

The Australian government's 1999 green paper on higher education research and research training appears to favour a larger degree of institutional autonomy and 'self-assessment' in the planning and evaluation of university research, while suggesting possible indicators that might be used in common (Kemp, 1999). It formalises the 'linkage' role within the ARC's research support programmes (implying an evaluation criterion of 'user' or partner satisfaction), and emphasises ARC's responsibility to report on both the performance of Australian research and on national returns on investment in research.

The greater involvement of universities themselves in research evaluation and performance management and assessment is certainly consistent with international trends, and with the 'learning loop' model of evaluation discussed in the study. On the whole, though, the new Australian proposals provide a challenge rather than a solution for the evaluation of university research outputs, outcomes and impacts. It is unclear, for example, how far the ARC can or should play a more 'intermediary' role in *ex post* evaluation, in the Dutch sense. Nor is it clear how formalised the 'possible common indicators' might (or indeed should) evolve to become.

But, if the issue is extending the evaluation of the quality of university research, developing new approaches to measuring its relevance or impact, recruiting intermediaries and users to the evaluation process, or tracking the broader social contribution of research trained graduates, it will undoubtedly pay Australian research councils and agencies to keep a weather eye on the evaluation strategies of their overseas counterparts.

Notes

1. The study was commissioned by the Australian Department of Education, Training and Youth Affairs.
2. The international reviewers were: Susan Cozzens (School of Public Policy, Georgia Institute of Technology, Atlanta, USA), Denis Gray (North Carolina State University, USA), J Adam Holbrook (Centre for Policy Research on Science and Technology, Simon Fraser University, Vancouver, Canada) and Barend van der Meulen and Arie Rip (Centre for Studies of Science, Technology and Society, University of Twente, the Netherlands).
3. http://www.regard.ac.uk

References

P Bourke (1997), *Evaluating University Research: The British Research Assessment Exercise and Australian Practice*, NBEET/ARC/HEC Commissioned Report No. 56, July (Australian Government Publishing Service, Canberra).

B Bozeman and J Rogers (1998), 'Strategic management of government-sponsored R&D portfolios: lessons from Office of Basic Energy Sciences projects', *APEC Symposium on the Evaluation of S&T Programmes among APEC Member Economies, 2–3 December 1998, Wellington, New Zealand* (National Centre for Science and Technology Evaluation, Beijing).

S E Cozzens (2000), 'Assessing federally-supported academic research in the United States', *Research Evaluation,* 8(1), April, pages 5–10.

S Garrett-Jones and D Aylward (2000), 'Some recent developments in the evaluation of university research outcomes in the United Kingdom', *Research Evaluation*, 8(1), April, pages 69–75.

D O Gray (1999), 'Evaluation of university/industry linkage programmes in the United States', in Turpin *et al*, 2, page 151.

J A Holbrook (1999), 'Higher education research evaluation in Canada', in Turpin *et al*, 2, page 175.

D Kemp (1999), *New Knowledge, New Opportunities: A Discussion Paper on Higher Education Research and Research Training* (Commonwealth of Australia, Canberra).

National Academy of Sciences (1998a), *International Benchmarking of US Materials Science and Engineering Research: Panel on International Benchmarking of US Mathematics Research*, NAS/NAE/IOM.

National Academy of Sciences (1998b), *Evaluating Federal Research Programs —Research and the Government Performance and Results Act*, Committee on Science Engineering, and Public Policy, NAS/COSEPUP.

T Turpin, S Garrett-Jones, D Aylward, G Speak, S E Cozzens and R Iredale (1999), *Valuing University Research: International Experiences in Monitoring and Evaluating Research Outputs and Outcomes*. A report prepared for the Research Evaluation Committee, Department of Education, Training and Youth Affairs/Australian Research Council, Centre for Research Policy, University of Wollongong, May.

T Turpin and S Garrett-Jones (2000), 'Mapping the new cultures and organization of research in Australia', in P Weingart and N Stehr (eds.), *Practising Interdisciplinarity,* University of Toronto Press, pages 79–109.

B van der Meulen and A Rip (1999), 'Evaluation of public sector research: the Dutch evaluation research system and the idea of societal quality', in Turpin *et al*, 2, page 195.