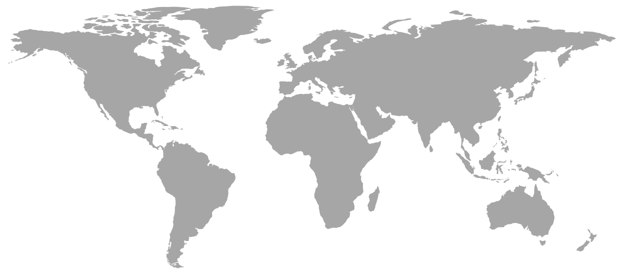
THE WORLD BANK OPERATIONS EVALUATION DEPARTMENT



The Australian experience with a centrally-driven system of performance evaluation shows how evaluation findings can support the annual budget process. In contrast, the second generation system, which has taken a devolved approach and focused on the collection and reporting of performance information, has been less successful.

Two Generations of Performance Evaluation and Management System in Australia

ECD Working Paper Series ♦ 11

Keith Mackay

www.worldbank.org/oed

March 2004
The World Bank
Washington, D.C.

Operations Evaluation Department
Partnerships & Knowledge Programs (OEDPK)
Email: eline@worldbank.org
Telephone: 202-473-4497
Facsimile: 202-522-3125

Evaluation Capacity Development (ECD) helps build sound governance in countries—improving transparency and building a performance culture within governments to support better management and policymaking, including the budget process—through support for the creation or strengthening of national/sectoral monitoring and evaluation systems. OED aims to identify and help develop good-practice approaches in countries, and to share the growing body of experience with such work.

The OED Working Paper series disseminates the findings of work in progress to encourage the exchange of ideas about enhancing development effectiveness through evaluation. An objective of the series is to get the findings out quickly, even if the presentations are somewhat informal.

CONTENTS

FOREWORD

The Operations Evaluation Department (OED) of the World Bank has a long-standing program of support to strengthen monitoring and evaluation (M&E) capacities in developing countries, as an important part of sound governance. As part of this support, OED has prepared a collection of resource material including case studies of countries which can be viewed as representing good-practice or promising-practice. This resource material is available electronically at: <http://www.worldbank.org/oed/ecd/> Considerable interest has been expressed in the case of Australia, which was the subject of an earlier paper in this series, prepared in 1998. The current paper compares and contrasts the Australian experience in this earlier period with subsequent experience.

Earlier versions of this paper have been presented at conferences in Mexico City, Seoul, Brasilia and Washington, D.C. in 2003. We are indebted to Russell Ayres, Pat Barrett (AO), Meredith Edwards, Osvaldo Feinstein and David Shand for invaluable comments on an earlier draft. The views expressed in this paper are solely those of the author, and do not necessarily represent the views of the World Bank.

Patrick G. Grasso
Acting Manager
Knowledge Programs & Evaluation Capacity Development

ABBREVIATIONS

ANAO	Australian National Audit Office
CEO	Chief Executive Officer
DEWR	Department of Employment and Workplace Relations
DoF	Department of Finance
DoFA	Department of Finance and Administration
FaCS	(Department of) Family and Community Services
GDP	Gross Domestic Product
M&E	Monitoring and Evaluation
MTEF	Medium-Term Expenditure Framework
OECD	Organisation for Economic Cooperation and Development
OED	Operations Evaluation Department
PEP	Portfolio Evaluation Plan
SCRCSSP	Steering Committee for the Review of Commonwealth/State Service Provision

EXECUTIVE SUMMARY

This paper describes and contrasts two generations of performance evaluation and management system in Australia over the past sixteen years. The first generation related to the decade from 1987 to 1997. It was centrally-driven, stressed formal requirements, and was principally concerned with ensuring evaluation findings were available to feed into the annual budget process. This system was successful in ensuring that evaluations were planned carefully, and that a growing 'library' of evaluation findings was available to support and to influence budget decision-making. However, this system had some weaknesses, including inadequate attention to the collection, use and reporting of performance information concerning government programs.

The second generation system was introduced in 1997. It has entailed a highly devolved approach, stressing principles rather than formal requirements, and largely emphasizing the collection and publication of performance information, and its provision to the parliament. However, there is evidence that this information has proved insufficient to meet the needs of parliament, and that it is insufficient for purposes of sound management. Some departments continue to devote significant resources to the conduct of rigorous evaluations, in support of their ongoing management and for reporting to parliament and others.

Finally, the paper offers lessons for efforts to develop performance evaluation and management systems in other countries. These include the pros and cons of a centrally-driven system, as compared with a more devolved approach. The critical role of powerful finance ministries is also discussed, together with the benefits from having a strong reform champion. The dangers of over-designing a performance evaluation system are also examined briefly.

1. FIRST GENERATION: 1987 TO 1997[1]

1.1 The Priority for Public Sector Reform

The reformist Labor government, which had been elected in 1983, faced a difficult macroeconomic situation including very tight budgetary constraints. In response, it successfully reduced the share of federal government outlays in GDP from 30% in 1984-85 to 23% in 1989-90. Its spending agenda was further complicated by its desire to substantially re-orient public expenditure towards poorer members of society. The new government was also determined to implement a series of public sector reforms with the objective of improving its performance significantly.

One aspect of these reforms was the desire to 'let the managers manage' by devolution of powers and responsibilities to them—encouraging better performance by providing much greater autonomy to managers. A medium-term expenditure framework (MTEF) was introduced with substantial autonomy for departments in their spending of salaries and other administrative expenses, and with greater surety about future resource availability. The annual budget became no longer zero-based; instead, the budget process focused on changes in government policy and spending priorities.

There was also a major reduction in the number of government departments through amalgamation, with the objectives of achieving less balkanized policy advice and of encouraging the internal reallocation of resources through portfolio budgeting.

The government advocated the principles of program management and budgeting, with a focus on the efficiency and effectiveness of government programs—through sound management practices, the collection of performance information, and the regular conduct of program evaluation. Guidance material on these principles was published by the Department of Finance (DoF) and the then Public Service Board, another central agency. And central departments also participated in program effectiveness reviews and joint management reviews of programs.

DoF was a major architect of many of the government's public sector reforms, reflecting its role as budget coordinator and overseer of the spending of other departments. DoF was keen to get out of the detail of spending issues, where the zero-based budget process had meant that a substantial portion of its day-to-day work was narrowly focused on minor spending bids and disputes with departments. DoF wished to focus much more on higher-level policy issues, as exemplified in its policy analysis and briefings prepared in support of the annual budget process. Its concern with budget spending encompassed not simply a priority on cutting government outlays, but also in finding ways to make spending more efficient and effective.

However, DoF and other central agencies remained unhappy with the performance of line departments in managing their performance, and so in 1987 the Minister for Finance was able to get the Cabinet's agreement to a formal requirement that all budget spending proposals ('new policy proposals') should include a statement of objectives and performance measures, as well as proposed arrangements for their future evaluation. Departments were also required to prepare plans for the systematic monitoring and evaluation (M&E) of their programs, and to report these plans to the government. At the same time, DoF expanded the advisory support it provided to line departments by provision of guidance material and a basic training course in evaluation.

By 1988 it had become evident to DoF that departments' evaluation plans were often poor, and that a more fundamental review of their M&E practices was necessary. DoF's in-depth review found:

- a lack of integration of evaluation into corporate and financial decision-making;

[1] This first generation of performance evaluation system is described in detail by Mackay (1998a).

- that evaluations tended to focus on efficiency and process issues rather than on the more fundamental question of overall program effectiveness—i.e., whether or not programs were actually meeting their objectives;
- a poor level of evaluation skills and analytical capacity; and
- that the role of central departments in evaluation, especially DoF, was unclear.

1.2 The Government's Evaluation Strategy

DoF concluded that 'letting the managers manage' was insufficient; it was judged necessary to 'make the managers manage' (Keating and Holmes 1990). Thus in late 1988 the Minister for Finance secured Cabinet's agreement to a formal ***evaluation strategy*** whose underlying principle was that 'the primary responsibility for determining evaluation priorities, preparation of evaluation plans and conduct of evaluations rests ..… (with line departments)'. The strategy had three main objectives. It provided fundamental information about program performance to aid Cabinet's decision-making and prioritization, particularly in the annual budget process when a large number of competing proposals are advocated by individual Ministers. It also encouraged program managers within departments to use evaluation for the improvement of their programs' performance. Lastly, the strategy aimed to strengthen accountability in a devolved environment by providing formal evidence of program managers' oversight and management of program resources. This emphasis on transparency is of considerable interest to the parliament, particularly in the senate's processes of budget scrutiny and approval.

The evaluation strategy to which Cabinet agreed had four formal requirements for departments:
- that every program be evaluated every 3-5 years;
- that each portfolio (i.e., comprising a line department plus outrider agencies) prepare an annual portfolio evaluation plan (PEP), with a 3-year forward coverage, and submit it to DoF—these plans comprise major program evaluations with substantial resource or policy implications;
- that Ministers' new policy proposals include a statement of proposed arrangements for future evaluation; and
- that completed evaluation reports should normally be published, unless there exist important policy sensitivity, national security or commercial-in-confidence considerations, and that the budget documentation which departments table in parliament each year should also report major evaluation findings.

Cabinet also agreed that DoF would have the opportunity to make an input to PEPs and to the terms of reference of individual evaluations to ensure their consistency with government-wide policies and priorities, and that DoF would be available to participate directly in selected evaluations, subject to negotiation between DoF and the line department (or between their Ministers if a dispute arose). Line departments had expressed serious concerns with the planned role for DoF, which they regarded as intrusive. Nevertheless, Cabinet's agreement to the evaluation strategy was in the form of a formal cabinet decision; and for the federal government and its public servants, such decisions virtually have the force of a law.

In the following years, two government reports, from a parliamentary committee and the national audit office, noted the continuing unevenness in the coverage and scope of evaluation activity in line departments (Parliament of Australia 1990; Australian National Audit Office (ANAO) 1991a). Both reports argued that DoF should take a more active approach to encouraging departments to plan and undertake evaluations. DoF then created a separate branch, responsible for the provision of evaluation advice, support, training and encouragement to other departments and also within DoF itself. This branch had 9 evaluators able to provide assistance and it acted as a focal point and catalyst for evaluation throughout the Australian public service. It prepared detailed advice and handbooks on evaluation methodology, provided introductory evaluation training, identified and shared evaluation best practice, and promoted a community of evaluators within the federal public service.

One limitation of the evaluation strategy was that it paid insufficient attention to the regular collection, use and reporting of performance information, via tools such as management information systems and performance indicators (Mackay 1998a); however, it had been hoped that evaluation findings would lead to the improvement of performance indicators and the setting of performance targets. By the mid-1990s DoF was concerned about departments' poor progress in articulating clear and achievable objectives for their programs, and in collecting and reporting meaningful performance information on a frequent basis. These concerns were confirmed by two reviews which DoF commissioned, concerning departments' annual reports and their budget documentation. This situation might appear to be somewhat paradoxical, because evaluation can involve relatively sophisticated techniques, and by that time it was generally being done well, yet the setting of program objectives and the collection of frequent performance information are often conceptually easier, and they were being done poorly. Yet evaluation had been mandated, while collection of performance information had not.

Thus in 1995 DoF secured Cabinet's agreement to a rolling series of comprehensive reviews, staggered over 3 years, of the program objectives and performance information of all programs in all departments.[2] These reviews were conducted jointly by DoF and each line department, with the results reported to their respective ministers and to Cabinet as a whole. The reviews laid the basis for a much greater focus on performance information under the second generation of reforms, which commenced in 1997 (discussed below).

1.3 How Successful was the Evaluation Strategy?

(i) Evaluation Planning Since 1987-88, all government departments had prepared annual portfolio evaluation plans and these were intended to comprise the major evaluations in the department and its outrider agencies. By the mid-1990s about 160 of these evaluations were underway at any given time. The majority of these evaluations were major, in that the programs had significant policy or spending implications, although a significant minority of these evaluations, particularly for the smaller departments, were only of minor programs or of efficiency aspects of large programs. Line departments themselves decided which programs should be included in their PEPs, and also which issues the evaluation terms of reference would cover. However, DoF would usually endeavor to influence departments' choice of evaluation priorities by making direct suggestions to them. In making these suggestions DoF would attempt both to anticipate and to help create the information needs of Cabinet. Where DoF had difficulty in persuading departments, it sometimes approached Cabinet directly to seek its endorsement of proposed evaluation topics and also detailed terms of reference.

The Cabinet-endorsed, formal requirement under the evaluation strategy that portfolio evaluation plans be prepared and submitted to DoF certainly provided a powerful incentive to line departments to prepare plans and to take them seriously. Another influential factor was the issuing by DoF of formal guidelines to departments concerning the desirable content of these plans, together with follow-up monitoring and reminders to departments about the need for the plans. The evaluation branch of DoF conducted internal reviews of the content and coverage of these evaluation plans, and provided feedback and prompting to departments as well as by identifying good practice examples. DOF also supported the creation of the Canberra Evaluation Forum, which involves monthly meetings of the evaluation community to discuss topical evaluation issues. Forum meetings are organized by a steering group of departments and other interested parties.

In a number of efficiency audits and two 'better practice' guides on program evaluation and performance information, the national audit office also repeatedly reminded departments about the importance of systematically planning their evaluation activity (ANAO 1991a,b, 1992 a,b,c, 1993, 1996, 1997; ANAO/DoF 1996). Indeed, the ANAO's strong support for evaluation, which continues to the present day, and its active program of performance audits—themselves a form of evaluation—have provided invaluable support for Australia's two generations of performance evaluation and management system.

[2] See, for example, DoF (1996).

The formal requirement that all programs be evaluated every 3-5 years was also influential in creating a climate of expectation that evaluation is the norm rather than the exception. The concept of regular, comprehensive coverage of programs also encouraged a planned, staged approach to evaluation. This formal requirement should not be accepted at face value, however. It is very seldom the case that all aspects of a program are included in any single evaluation. Instead, it is usual that an evaluation will focus only on certain key problem issues or aspects of a program. The challenge is to ensure that these difficult issues are actually evaluated, and this is a role in which DoF played an active role via persuasion and via direct involvement in individual evaluations.[3]

(ii) Conduct of evaluation Most departments chose to set up evaluation units to coordinate their formal evaluation planning. At their smallest, these units comprised two or three individuals. In some departments, such as employment, there was a separate branch of 20-25 staff responsible for evaluation planning, provision of advice on evaluation methodology, participation in steering committees, and the conduct of a number of major evaluations, particularly in the area of labor market programs.

There was no standard approach by departments as to how they chose to conduct evaluations. Some evaluations involved a wide array of external and internal stakeholders, either by their participation in an evaluation steering committee, or less commonly by their participation in the actual evaluation team. Some evaluations were conducted by a central evaluation unit, but it was more common for responsibility for the conduct of evaluations to rest with the line program area. For the more important evaluations—those listed in portfolio evaluation plans—some external involvement would be typical, via provision of suggestions and comments on the terms of reference and proposed evaluation methodology, participation in the steering committee, and provision of comments on drafts of the evaluation report. But, again, there was no standard approach to this external involvement—it would be determined by the willingness of the line department to involve outsiders, and also by the interest and availability of outsiders such as central agencies to become involved. For programs with major resource or policy implications, DoF would usually be very keen to be involved, and would apply whatever pressure it could to ensure its participation.

A national audit office survey found that, for evaluations conducted over the period 1995-1997: about half examined the delivery of products or services to external clients, and a further 30% were associated with matters internal to the department. One third of the evaluations examined the appropriateness of new or established programs, and 15% were directed towards the development of policy advice for the government (ANAO 1997).

The large number of evaluations underway at any time, and the fact that over 530 evaluation reports were published between 1993 and 1997, attest to the existence of extensive evaluation activity in the Australian government. This has provided a growing 'library' of evaluation findings. DoF also published a register of published evaluation reports, and this also provided some quality assurance because the public availability of these reports exposes them to peer scrutiny. The ANAO survey found that 75% of evaluations conducted in 1995 and 1996 were released to the public and/or available on request.

(iii) Evaluation quality Quality of evaluation reports is a more difficult dimension to measure. The rigor of program evaluations depends on the expertise and objectivity of the evaluators. The ANAO assessed the quality of a sample of evaluation reports in 1997 and found that over a third of them suffered from methodological weaknesses of one kind or another. It is certainly the case that some published evaluations were of low quality, and the suspicion is that some of these were produced for self-serving purposes, such as to provide a justification for the retention or expansion of the program. DoF's own perspective was that the quality of evaluations can vary enormously. This would be a significant problem if the intended audience of an evaluation is Cabinet (to aid its decision-making) or the parliament (for accountability

[3] There was only modest success with the requirement that Ministers' new policy proposals include an evaluation plan of action that would be undertaken if the proposal was accepted. Feedback from portfolios indicated that this requirement was onerous for portfolio managers during the busy budget period. Only about 30% of proposals broadly met this requirement in the 1993-94 budget, for example, although an additional 50% of proposals included a clear undertaking to evaluate the proposal if accepted (DoF 1994b). These percentages were only achieved after considerable prodding by line areas within DoF.

purposes). In such circumstances DoF would certainly be willing to inform the Cabinet that it considered an evaluation to be unreliable. Line departments would typically try hard to avoid such criticism, which would be virtually guaranteed to attract the ire and condemnation of Cabinet.

The national audit office consistently argued that departments should set up central oversight procedures to achieve quality assurance of evaluations conducted by line areas within the department. There is certainly evidence from those few departments which have followed this approach that it is an effective means of making available needed evaluation skills and expertise, and of ensuring evaluation quality. But most departments chose to rely on program managers and their staff for the actual conduct of evaluations. This devolutionary approach helped ensure that the evaluations drew on the program expertise of staff, and that there was a high level of 'ownership' of the evaluation findings—both of these may be difficult to achieve with external evaluations. DoF's philosophy was to try to achieve the benefits of self-evaluation while ensuring, via its involvement in the steering committees of major evaluations, that sufficient objectivity and rigor was achieved.

A disadvantage of this devolved approach was a lack of evaluation skills in many program areas and lack of experience in conducting evaluations. Basic training in evaluation skills was widely available in the Australian government—provided by DoF in particular—and DoF and departments also prepared guidance material such as evaluation handbooks (e.g., DoF 1991, 1994a, 1996). There is also a fairly large community of evaluation consultants in Canberra, including numerous academics with either subject area knowledge (e.g., health issues) or with specialist research and analysis skills. Nevertheless, the 1997 ANAO study also revealed that 20% of departments were concerned about the lack of available training in advanced evaluation techniques, and this does appear to have been a weakness of the Australian evaluation system.

Some departments addressed the need for more advanced skills and experience by setting up a central evaluation unit to provide advice on methodology and to participate in evaluation steering committees. The then department of health pursued evaluation quality assurance in a devolved environment by ensuring adequate skills and resources were available to program managers, and also by ensuring there were structural arrangements to support line program areas, such as technical panels and steering committees.[4] That department, like some others, also put a lot of effort into training its staff to enhance their analytical and research skills.

(iv) Use of evaluation A bottom-line issue is the extent to which evaluation results are actually used. If their use is patchy or poor then there really is little point in conducting evaluations. There is clear evidence that evaluations were used intensively in the budget process—they provided a substantial contribution to the development of policy options and their consideration by cabinet. DoF conducted several surveys of the extent of influence of evaluation findings on the budget proposals submitted to Cabinet for its consideration (e.g., DoF 1994b). These were surveys of DoF officers, who typically attended all Cabinet meetings concerned with budget issues, and their judgments were sought concerning the extent of influence of evaluation on the budget proposals of line Ministers. The close familiarity of DoF officers with these proposals and also with any evaluations or reviews on which they might draw, gives them an insider's perspective on the extent of influence of evaluation.

In the 1990-91 budget, some A$230 million (then about US$175 million) of new policy proposals submitted by line Ministers were judged to have been directly or indirectly influenced by the findings of an evaluation. By 1994-95—the latest year for which estimates were available—this had risen to A$2300 million. Measured in dollar terms, the proportion of new policy proposals influenced by evaluation rose from 23% to 77% over that period; and for most of these in turn, the influence of evaluation was judged by DoF officers to be both direct and 'major'. These results indicate the importance which public servants, in their preparation of the details of new policy proposals, and Ministers attached to having evaluation

[4] The department of health encouraged quality evaluations through: selection of good quality officers to manage the evaluation; involvement of internal and external stakeholders; ensuring technical advisory panels were available to help assess the work of consultants; having steering groups available to help manage consultants; and ensuring sufficient resources were available for the evaluation.

findings available. Overall, it was very important to have had the active support of key Cabinet and other Ministers in encouraging portfolios to plan and conduct high-quality evaluation. This support was also reflected in the many Cabinet decisions which called for evaluations of specific programs or issues.

It is also the case that evaluation can have a significant influence on the 'savings options' put forward by DoF or by portfolios for Cabinet consideration in the budget process. (Savings options are areas of government expenditure which could be trimmed or abolished entirely.) In 1994-95 about A$500 million of savings options—or 65% of the total—were influenced by evaluation findings; again, the influence of evaluation was usually judged to be major. This emphasis on evaluation findings was encouraged by the nature of the budgetary system in the Australian government. Australia had a well-functioning policy making mechanism which makes transparent the costs of competing policies and encourages debate and consultation among stakeholders within government. In this 'marketplace of ideas' evaluation findings can provide a competitive advantage to those who rely on them.

One issue which it is important to appreciate is the realistic limits to the influence of evaluation on Ministers' or Cabinet's decision-making. The evaluation paradigm in an investment project is typically that of cost-benefit analysis: a project is warranted if, but only if, its benefit-cost ratio is greater that one. But program evaluation is a more qualitative science: it can help identify the efficiency or effectiveness of existing, ongoing programs but it can rarely provide an overall conclusion that the activity is worthwhile or not.

DoF officers were also surveyed for their judgments on the extent to which evaluation had influenced Cabinet's decisions in the 1993-94 and 1994-95 budgets. While the evidence is mixed, it indicates that evaluation played a substantive role. In 1994-95, evaluation was assessed to have influenced Cabinet's decision in 68% of the A$3,740m of proposals considered (new policy proposals plus savings options).[5] The corresponding proportion for the 1993-94 budget, however, was only 19% of proposals. One important reason for this difference was the substantial revision of labor market, industry, regional and aboriginal policies in the 1994-95 budget—the major policy review on which these decisions were based had been heavily influenced by a number of evaluations commissioned specifically to help guide the policy review (DoF 1994b).

Campbell (2001) has observed that DoF 'functioned as the nerve center for cabinet on how programs actually functioned. It spearheaded public service reform throughout the 1980s. Its signature initiatives were devolution of authority over running costs to departments and the promotion of an evaluation culture that allowed ministers to monitor and gauge the effectiveness with which departments employed such discretionary spending power'.

The observation of the Auditor-General is particularly noteworthy: 'In my view, the success of evaluation at the federal level of government …. was largely due to its full integration into the budget processes. Where there was a resource commitment, some form of evaluation was necessary to provide justification for virtually all budget bids.' (Barrett 2001). And in comparing the relative success of Australia's and Canada's approaches to incorporating evaluation into the budget process, Schick (2001) has stated that:

> 'Canada organized a vast (evaluation) effort around the Comptroller General in the 1970s, Australia adopted an ambitious evaluation strategy in the late 1980s. Canada's effort bore little fruit, Australia's produced significant reallocation of budget resources. Canada is thought to have failed because it centralized evaluation, thereby dampening cooperation by spending departments which may have been adversely affected by the findings. Australia is thought to have succeeded because it gave affected departments a big stake in designing and using evaluations.'[6]

[5] Moreover, where policy proposals were supported by evaluation, the evaluation was judged to have also influenced Cabinet's decision in 89% of cases.

[6] Canada introduced a new, whole-of-government results management framework in 2000. This includes a rejuvenated evaluation strategy, with additional funding of C$20m over four years to help departments strengthen their evaluation capacities. See, for example, Treasury Board of Canada Secretariat (2000).

The large volume of evaluation activity also provides reassurance that evaluation findings were used by line departments in their ongoing operations—in an era of very tightly limited administrative expenses, departments would simply not bother to conduct such a large number of evaluations unless they were going to be used. The 1997 ANAO survey found that the impact or use of evaluations by line departments was most significant with respect to improvements in operational efficiency, and to a lesser extent with respect to resource allocation decisions and the design of service quality improvements for the benefit of clients.[7]

[7] See also Crossfield and Byrne (1994).

2. SECOND GENERATION: 1997 TO PRESENT

2.1 Context

The election of a conservative Coalition government in March 1996 marked a change in direction of public sector reform in Australia. The new government has a strong ideological preference for the private sector, with a bottom-line emphasis on results, and appears to regard the private sector as being inherently more efficient than the public sector. The government placed particular emphasis on its objectives to reduce the overall size of the public sector while also increasing its efficiency.

The government has expressed considerable unhappiness with the federal public service, and considered it to be rule-bound and caught up in red tape. This has led to an emphasis on the market testing and outsourcing of government activities wherever possible—a preference for non-government service delivery; one result has been a significant downsizing in the numbers of public servants.[8] And reflecting a private sector paradigm, the heads of government departments became chief executive officers (CEOs), held accountable for results rather than for bureaucratic processes; it was considered undesirable to constrain the actions of CEOs by excessive administrative controls. This was essentially a 'let the managers manage' philosophy, analogous to the one adopted in the early 1980s.

One example of simplification relates to the requirement under the first generation system for preparation of portfolio evaluation plans. Some of these had increased in size from a recommended 20 or 30 pages in length, to over 120 pages—with a concomitant increase in administrative workload necessary to prepare them. A consensus had emerged within the bureaucracy that while it was important to have evaluation findings available to assist decision-making by program managers and by Cabinet, detailed and elegantly-worded plans were not necessary to achieve that objective. At the same time, line departments were pressing for less oversight by, and reporting to, DoF.

2.2 Nature of the Reforms

Thus a new performance management system was created in 1997, embodying a mix of principles, expectations and formal requirements—overall, the number and nature of formal requirements was simplified considerably from those of the previous government (Russell 2003). All departments were required to measure their performance—with performance measured in terms of outcomes, based on relatively simple performance indicators—and to report annually to the parliament:

- ex ante reporting in portfolio budget statements, which relate planned spending appropriations to promised performance; and
- ex post reporting in departmental annual reports, discussing actual spending and actual outcomes.

Accrual accounting was introduced to the federal budget in 1999-2000. One reason was to ensure that budget appropriations reveal the full cost of government activities. Another was to facilitate benchmarking comparisons of the unit cost of government outputs and outcomes, in areas such as service delivery. This in turn was seen as a prerequisite for market testing and contracting out.

The dollar amounts of budget appropriations are now based solely on outcomes sought—thus making Australia unique, in contrast to the traditional focus on the amount of spending (inputs) or on outputs (as

---

[8] In this context it is worth considering the comments of Allen Schick (2003): 'A public service ethic is the bedrock of governmental performance which depends at least as much on people as on machinery and process..….In most developed countries, the ethic of public service still exists, but not as robustly as in the past. It has been eroded by powerful socioeconomic forces including…increased reliance on private markets and contractors to provide public services, and decline in the esteem with which public employees are held'.

per the approach of countries such as New Zealand). This also provided the parliament with a unique opportunity to debate and hold government accountable for the extent to which its activities and outputs succeed in achieving desired outcomes. (Parliament's evident unwillingness to assume this role is discussed below.) And CEOs have been given the authority to allocate these appropriations across different types of output to achieve desired outcomes. The accountability of CEOs is reflected in their having a formal performance agreement with their minister. Those departments and agencies which provide services directly to the public are required to put service charters in place, with actual levels of service performance to be included in their reports to parliament.

At the same time, the formal evaluation requirements of the first generation system were abandoned, and the evaluation support which DoF—renamed the department of finance and administration (DoFA)—had provided to line departments was essentially discontinued. DoFA still encourages line ministries to conduct evaluations of key policies and programs, ideally on a five-year cycle, but there is no longer any formal requirement for this. All funding submissions to Cabinet are required to report whatever evaluation evidence exists; but as there is no requirement to undertake evaluations, the requirement to report available evaluation findings is weak. However, from 2004-05 onwards, line ministers who wish to extend program funding (via a submission to Cabinet) for fixed-duration programs are required to review the program's performance to see if there is evidence that the program is achieving its objectives. DoFA has prepared generic terms of reference to guide such reviews. Unfortunately, there is no such requirement for the review of ongoing programs.

DoFA has interpreted this new generation of performance evaluation and management system as a devolved approach (Russell 2003). The Australian Auditor-General has characterized these reforms as a deregulation of evaluation (Barrett 2001).

Other than the introduction of accrual accounting, which the previous government had been developing over a number of years, the main innovation of the new government was its ***outcomes and outputs framework***. Taking effect in the 1999-2000 budget, this framework requires departments to agree with their ministers the outcomes—the government objectives—towards which they are working, and to also agree the departmental outputs[9] which would be produced to help achieve these outcomes. This framework also requires clear measures of departments' performance in achieving outcomes and outputs; this performance measurement is to be achieved by the regular collection, analysis and reporting of performance information (DoFA 2000; Chan et al. 2002; Russell 2003).

The three objectives of the outcomes and outputs framework, which mirror those of the earlier evaluation strategy, are:

- to support the government's policy development;
- to support and strengthen departments' internal management, including staff learning; and
- to strengthen external reporting, for accountability purposes.

On paper at least, the outcomes and outputs framework has much to commend it in addition to these three objectives. It has the potential to:

- clarify and ensure agreement among key stakeholders concerning expected standards of performance—this shared understanding is also crucial to ensure purchaser-provider arrangements work well;
- facilitate benchmarking comparisons across departments, and between the public and private sectors. This in turn would make it easier for the government to select least-cost service providers, thus yielding efficiency gains. Accrual accounting can be expected to facilitate considerably such an approach;
- encourage departments to understand well the results chains underlying their work activities—i.e., the logical (and desirably evidenced-based) links between departmental spending, their activities, their outputs, and their contribution to desired outcomes; and

[9] The term 'outputs' is used here in its generic sense, as distinct from the definition under the framework, which distinguishes between departmental outputs and administered items such as transfer payments and grants to the states.

- help departments to cope better with management complexity, particularly when the work of several departments jointly impinges on achievement of a particular outcome objective. This relates to the concept of 'joined-up government'.[10]

While this new generation of reforms was being implemented, the government reduced the role of DoFA in the policy process which the budget cycle represents. In particular, the divisional units in DoFA responsible for oversighting each line ministry have had their staffing levels and role substantially diminished. The government has instead put more emphasis on alternative sources of advice outside the public service, including ministers' offices and private sector sources.

2.3 What Have Been the Results from the Second Generation Reforms?

There is no simple answer concerning the results of the second generation performance evaluation and management system—the government has not published any systematic evaluation of these reforms. Of course, any new reforms have a period of implementation and bedding down before 'success'—however defined[11]—can reasonably be expected to emerge. But there are already some indications of strengths and weaknesses of the reforms.

The emphasis on outsourcing, competitive tendering and contracting, the creation of purchaser-provider relationships for the delivery of services, and the creation of service charters, have the potential to lead to a stronger focus on performance and its explicit measurement. They also have the potential to lead to considerable efficiency savings. For these approaches to work well there is a need for: a clear understanding of program objectives and target levels of performance; careful ongoing and ex post measurement of actual performance; and close attention to relationship management. Notwithstanding these priorities, the Auditor-General has noted the ongoing difficulties experienced by a number of agencies in contract management—these include lack of procedures for regular review of contractor performance, inadequate management information systems, and difficulties in selecting appropriate performance indicators (Barrett 2001).

The ANAO has criticized government management of one important type of purchaser-provider relationship—federal special-purpose grants to state governments for hospital health care (ANAO 2002). These grants amount to about US$20 billion over five years. The ANAO noted the existence of good information on financial flows and on program outputs. But it found that additional information is needed to interpret the available data, and that the performance information available on issues of the efficiency, effectiveness and access to hospitals is far from complete. The ANAO concluded that available data are insufficient to assess performance against objectives.[12]

More significantly, the ANAO has also conducted performance audits of a sample of departments' and agencies' reports to the parliament via the portfolio budget statements—the ex ante statements of government spending and performance targets—and also their annual reports—the ex post reporting of actual performance (ANAO 2001, 2003). The 2001 ANAO audit found that while the performance reporting of agencies did focus on outcome indicators, these indicators were inadequate to measure actual outcomes of government activities, because reported indicators are often influenced by factors beyond the

[10] New Zealand has taken a more systematic approach than Australia in this regard, and has formulated a number of strategic results areas which transcend narrow departmental perspectives—see for example Schick 1996; incidentally, his critique of the New Zealand reforms includes a warning against their uncritical application to developing countries (Schick 1998). A complementary critique of New Zealand's reforms appears in Campbell 2001.

[11] The definition of 'success' is debatable and is not considered explicitly here: is it the achievement of the government's explicit reform objectives? Is it the achievement of what might be considered good-practice standards for a performance evaluation and management system, such as performance budgeting?

[12] For another, albeit small government activity, the family relationships program, the ANAO is much more positive about the adequacy of available performance information. This program entails government support being provided via contracts with over 100 non-government organizations, costing US$28 million per annum (ANAO 2000).

agencies' control and which mask any direct effects of the agencies themselves.[13] This provides a strong argument for conducting evaluations and reviews, as a complement to performance indicators, so that the relative importance of the agencies' performance vis-à-vis external factors can be assessed. The ANAO also noted that the performance information did not always include targets, and where it did they were often vague or ambiguous.

These adverse findings were confirmed and elaborated on by the ANAO's more recent 2003 report, which focused on performance reporting in annual reports to the parliament:

- the reports usually define outcomes (and their expected impacts) clearly;
- but while agencies were able to produce accurate performance information, most did not have the necessary data standards and procedures in place to assure data quality and coherence;[14]
- a number of agencies used different systems and procedures to produce internal and external reports on performance;
- performance information was often not structured to allow an assessment of operational efficiency nor of the cost-effectiveness of outputs actually delivered;
- targets and other benchmarking comparisons were not widely used;
- the annual reports only provided descriptive information, with a lack of substantive analysis and a lack of targets—thus difficult to know if performance had or had not met expectations. Most of these reports failed to discuss areas where performance was poor;
- agencies continued to experience difficulty in measuring the quality of their outputs and their effectiveness or impact;
- there was little or no performance information available relating to the contributions of individual departments towards the achievement of shared outcomes; and
- finally, the ANAO noted that 'Although most agencies undertook a range of evaluations, the results of these were frequently not discussed in the annual report'.

The ANAO concluded that 'Performance information had not been presented and analysed in annual reports in a way that would allow Parliamentarians and other stakeholders to interpret and fully understand results'. Thus it is no surprise that the Auditor-General has noted[15] the dissatisfaction of parliamentary committees with the performance information provided by departments, with a particular difficulty being definitional changes from year to year—these make it difficult to analyze departmental performance over time, or to make efficiency comparisons between departments. These difficulties suggest that the government has failed to achieve one of its reform objectives, i.e. greater transparency via benchmarking comparisons of unit costs for all departments and agencies. It can be argued that one advantage of a *systemic and centralized* approach to performance measurement is that it enables exactly such comparisons to be made. A devolved approach, in contrast, is inherently unlikely to achieve this.

Performance information can provide the basis for a useful understanding of program and agency performance, but it requires careful analysis for this to be achieved. Simple reporting of data is insufficient; it is typically necessary to supplement this with qualitative information and especially with sound analysis and interpretation of the data. A good example of this type of analysis is provided by the federal/state performance comparisons for a wide range of government services, including schools, vocational training, public hospitals, police, court administration, prisons, emergency management, and community services such as public housing and support for the elderly, disabled and children (SCRCSSP 2003). These

---

[13] The relationship between government objectives (i.e. its desired outcomes), the strategic planning of line departments and agencies, and their ongoing management is fundamental to the achievement of a sound performance management system within agencies. See, for example, Edwards (2002).

[14] If performance information ever does become important for accountability purposes, this deficiency will become significant. This is because incentives will arise to 'game the system', notably by producing inaccurate or misleading performance information. Such a situation has arisen with the United States government's *No Child Left Behind* education program, which links federal government funding to school performance (The Washington Post, 8 November 2003). This is highly relevant to the Australian government's decision to link US$5 billion in school funding to students' academic performance (The Sydney Morning Herald, 14 November 2003).

[15] Presentation at a World Bank seminar on 'Australian Public Sector Financial Management Reforms', 25 June 2003.

comparisons, which have been published annually since 1995, are prepared by a number of federal/state working groups. Fortunately, this reporting has continued under the second generation reforms. Stakeholder surveys reveal that central and line ministries (in federal and state governments) find these data helpful in policy analysis and formulation, in assessing the resource needs of agencies, and in briefings for ministers and parliaments.

There are of course limits to the depth of understanding which can be derived from performance information. Issues of causality can be answered much more definitively via in-depth evaluations; hence there is a complementarity between performance information and evaluations. It is difficult to reach a judgment about the overall scale, nature and quality of evaluation activity underway in the Australian government. The devolved approach to performance management means that there is, in a sense, not one common system but instead 16—one for each of the portfolio departments. Among this 16 it is clear that some still devote considerable priority to evaluation, and some can be considered good practice in a number of respects. However, it appears that even these departments tend to conduct evaluation less frequently, to address particular issues on a selective basis. It not clear that many departments continue to conduct evaluation as regularly and systematically as under the first generation reforms.

One good-practice example is the department of family and community services, which provides income support to the disabled, rent assistance, and housing grants to the states. Its annual budget appropriations are about US$35 billion. It reports a wide range of performance information to parliament, particularly in terms of the quantity, quality and cost of outputs, together with some information on intermediate outcomes. The department's reports to parliament also include a substantive use of evaluation findings (FaCS 2002a,b). The department regularly prepares a research and evaluation plan, the latest version of which includes 145 research/evaluation projects—the total budget for this work is US$20 million per annum, equivalent to about 1.4% of the department's operational budget (FaCS 2001, 2002c). Funding is provided to 6 university research institutes, and for 3 longitudinal data sets. The research emphasis is on behavioral outcomes, and this is an important dimension of the evaluations which are conducted.

A significant aspect of this department's emphasis on research and evaluation is that it is a self-generated priority. The department receives no specific funding for this work; rather, it chooses to fund this work on its own authority within the overall budget envelope which it is provided. Thus even in a deregulated environment, evaluation is still viewed as important by this department.

Other noteworthy departments include, for example, education; employment; and health. The department of employment and workplace relations, for example, has a wide range of performance indicators derived from its administrative information system and from surveys, and these provide information on outputs and outcomes (DEWR 2002). The department also has an active evaluation branch which has a system for producing outcomes data on a regular basis, using longitudinal data on current and former program participants, with the data interpreted using econometric models to estimate the net effects of program interventions. Such data also form the basis for the evaluations conducted by the department (or sometimes conducted by universities); these evaluations are typically overseen by steering committees comprising central ministries (such as DoFA) and other line ministries. This model of internal, independent evaluations is the same as the one followed by this and a number of other departments under the first generation reforms.

The department has a service charter with service standards such as response times. Regular surveys of client satisfaction are undertaken with respect both to departmental services delivered directly by the department, and to outsourced services delivered under contract. An interesting feature of the department's client-orientation is the feedback sought systematically from the ministers who have responsibility for the department—they are asked to rate each piece of policy advice work according to 11 criteria, such as logic, accuracy, creativity, and timeliness. The ANAO has commended the department's public reporting of this performance information (ANAO 2001b).

The conduct and use of monitoring and evaluation (M&E) findings by line departments constitutes an important issue on which to judge the extent of success of the second generation of performance evaluation and management system in Australia—although, as noted earlier, no whole-of-government review of this is publicly available.[16]

Another issue is no less important—the utilization of M&E findings in the annual budget cycle. DoFA has not conducted any reviews of this issue, so it is not possible to draw definitive conclusions. The continuing effort devoted to evaluation by at least some departments provides a useful supply of evaluation findings, available to be drawn on in line ministries' budget submissions to Cabinet, and in the policy advice of central ministries concerning these submissions.

However, Campbell (2001) has provided evidence that the policy formulation process has been substantially weakened in recent years as a result of the irregularity of cabinet meetings and poor preparation for them, and that DoFA no longer serves as a 'nerve center' for cabinet concerning the performance of government programs—these findings are based on a number of interviews of senior executives he conducted in the latter part of the 1990s. Based on this evidence, he concluded that the policy process had been undermined due to '… the current government's tendencies to operate less collegially and collectively than did the Labor Government, to trust less and not include in decisions top officials, and to introduce a budgeting technology which appears to focus on outputs—often even inputs—at the expense of attention to outcomes'.

One senior official interviewed by Campbell lamented that '…if Finance keeps on the way it is with no emphasis on policy and no access to (cabinet), then its debates will not be on policy issues and you will lose the expertise—both the officials' and the ministers' '. This prediction is backed up by more recent, albeit anecdotal evidence that the extent of DoFA's involvement in the budget process, the quality of its policy advice, and the reliability of the evidentiary basis of this advice, has declined markedly during the second generation reform period. This has led the Organisation for Economic Cooperation and Development to conclude that:

> *In Australia, the deregulation of the public service and the adoption of an arm's-length posture by the central agencies allowed management freedom but is currently considered to have deprived the Finance Ministry of the information necessary for it to adequately advise the Minister.* (OECD 2002)

For a finance ministry, with its traditionally leading role in the budget and related policy formulation processes of government, this criticism is significant.[17] More in-depth review would be desirable to investigate the nature and full extent of these problems. Prima facie, however, it appears that the devolved approach to evaluation has helped to undermine the ability of central ministries to adequately advise government.

[16] DOFA has apparently conducted at least two reviews, including an inter-departmental *budget estimates and framework review* in 2002 (DOFA 2003). However, its findings have not been made public, and there are no clear implications for the government's approach towards performance management.

[17] The DOFA unit responsible for advising on public sector management reform issues was also disbanded, in order to focus the department more fully on accrual budgeting and the contracting out of government activities.

3. COMPARISON OF FIRST AND SECOND GENERATION REFORMS

It is difficult to reach a balanced judgment about the relative success of the two generations of performance evaluation and management reform in Australia. The first generation, which emphasized the planning, conduct and use of evaluations, particularly during the budget process, started in 1987 and has been shown to have had a significant influence on the policy advice and budget process by the mid-1990s. There is also evidence that evaluations were being used quite intensively by departments in support of their ongoing management. Yet a continuing weakness in this first generation was the inadequate attention given to the collection, reporting and use of performance information on programs.

The second generation reforms have now been underway for six years. They have relied on a much more devolutionary approach to public sector management, and have sought to influence public sector performance by influencing the underlying incentives faced by CEOs and their departments—consistent with the philosophy that if the incentives are appropriate, better performance will result. However, the strength of existing incentives for good performance appears weak. And there is some evidence that the regular performance information available to line managers remains insufficient for purposes of sound management, particularly in the more challenging environment where outsourcing has become more common, and that it is still insufficient to meet the information needs of the parliament. Some departments continue to emphasize evaluation in support of their internal management and for reporting purposes—this can reasonably be interpreted partly as a continuing benefit from the first generation reforms. However, it is not known to what extent evaluation findings are still used by line departments in the annual budget process. Finally, the ability of DoFA to provide sound policy advice appears to have been significantly reduced.

The centerpiece of the government's performance management system is the outcomes and outputs framework, and it is important to understand if the continuing difficulties it is experiencing are due to implementation problems, such as insufficient training for public servants or inadequate central guidance to (or policing of) departments, or whether they reflect more fundamental design issues. The findings from the ANAO performance audits reveal that the information base necessary for achievement of the strategic objectives of the framework remains substantially inadequate—despite the framework having been in place for over five annual budget cycles. These deficiencies cannot reasonably be interpreted as simple 'teething problems'.

A fundamental weakness in the design of the framework appears to be the weak incentives for individual departments to take it seriously—or at least seriously enough to warrant the necessary effort on their part.[18] The key accountability relationships which underpin the framework are between departments and their ministers, on the one hand, and the parliament on the other. Accountability of line departments to DoFA and to the other central ministries appears to have fallen largely by the wayside.

Accountability to parliament has always been an important part of governance relationships, and thus in one sense this is nothing new. But parliament apparently continues to scrutinize budgets largely at the level of inputs—i.e., spending allocations. There is no evidence of any real parliamentary 'ownership' nor capacity for the role which DoFA and other architects of the framework have assigned to them. The ANAO, in its direct reporting to parliament, continues to play an important role in this accountability relationship, but this role cannot be expected to be a substitute for the type of ongoing, detailed review, oversight and policy advice on departmental performance by central departments which occurred under the first generation of reforms.

[18] The World Bank's experience in support of public sector reform in the 150 borrower countries with which it works is that incentives are key to such reform being successful and sustainable (see, for example, World Bank 1997).

In the absence of strong accountability pressures on line departments, and in an environment in which the policy and budget processes have been weakened, any incentives for line departments to take performance monitoring and evaluation seriously will have to be internally generated. This in turn will depend on the priorities and commitment of individual department secretaries.

The genesis of evaluation in the first generation reforms was a response to the difficulties encountered with a simple 'let the managers manage' approach, with the realization that this devolutionary philosophy needed to be balanced with tighter information requirements to demonstrate performance—to support resource allocation decision-making in the budget process and for reasons of accountability. The second generation reforms have applied the devolutionary approach to evaluation itself, with the unfortunate result that the quality and depth of the information available concerning government performance has actually declined. This appears to have repeated some of the key mistakes of the early 1980s;[19] however, one difference with this earlier period is that there are now bureaucratic requirements for line departments to produce and report a large volume of performance information. At the risk of being overly glib, it appears that the second generation reforms 'threw out the baby but kept the bathwater'.

The government's ability to measure its outcomes and to report this performance to the parliament and other stakeholders appears to have been compromised. Sharper incentives will be needed to improve the performance information and evaluation findings being produced by line departments and to ensure that such information is used effectively; there is little point in producing better information if it is not going to be used intensively. For performance information to yield reliable insights into outcomes attributable to government activities will require greater reliance on evaluations, if issues of causality, the role of external factors and the reasons for good/bad performance are to be much more fully understood. Whether the Australian government will choose to place a renewed emphasis on evaluation—either via a principles-based approach or a compliance-based approach —remains to be seen.

[19] Indeed, Campbell (2001) has characterized the government as having undertaken budgeting and management reform 'through an unreconstructed 1980s lens'.

4. SOME IMPLICATIONS FOR EFFORTS TO DEVELOP PERFORMANCE SYSTEMS

This section draws on the lessons from Australia presented in earlier sections, together with experience from other countries which have endeavored to strengthen or build a whole-of-government M&E system (for example, Mayne and Zapico-Goni 1997; Mackay 1998b, 2002; Boyle and Lemaire 1999).

The evaluation strategy under the first generation reforms was a centrally-driven initiative of the powerful finance ministry, linked closely to its main area of influence—the annual budget process. The strategy's success can reasonably be interpreted as a result of that ministry's influence and power. There are advantages and disadvantages of a centrally-driven approach, with the latter including often weak commitment by line departments. DoF tried to soften such disadvantages by relying on persuasion wherever possible, and by providing a range of positive support and assistance, rather than by using more forceful methods. In contrast, the limitations of a reliance on advocacy of good-practice principles can be seen from the results of the second generation reforms (and from the initial effects of the 'let the managers manage' reforms in the mid-1980s). Thus the appropriate interpretation appears to be that a balance of 'carrots, sticks and sermons' works best, rather than only one of these (see, for example, Bemelmans-Videc et al. 1998). In Australia's case, the carrots included the advisory support provided by DoF, as well as possible access to resource agreements for line departments which put a convincing case for up-front funding (more than balanced by the later funding cuts which would agreed ab initio) to help address under-performing programs. The sticks were related to DoF's ongoing relationship with line departments and DoF's ability to influence their budget allocations. DoF also had the option of embarrassing departments by releasing the comparative rankings it prepared concerning departments' approaches to the planning and conduct of evaluation. The sermons included persistent advocacy by the permanent secretary of DoF and his senior executives.

Finance ministries may not always be supporters of reforms designed to strengthen the amount of information available on government performance. Before the reforms, the Australian DoF was heavily involved in the detailed scrutiny of departments' spending activities. The danger is that this traditional focus on spending can mean relatively little attention is paid to the results of that spending. And powerful finance ministries can even act as roadblocks to reform. Having DoF responsible for evaluation oversight ensured that there was a direct influence on the divisional units within DoF which oversight line departments. However, achieving the needed cultural change within DoF was a slow process over a number of years, and involved a substantial staff turnover. DoF's greater focus on issues of value for money (rather than on spending issues) flowed through to the nature and quality of policy advice which DoF provided to Cabinet; and that advice increasingly drew on available evaluation findings.

The Australian experience shows that the policy process, involving the annual budget cycle, provides a particularly powerful vehicle for achieving a focus on government performance. Central to this is the question of what is regarded as good-quality policy advice: what standards are applied within ministries, by central ministries, and by the Cabinet?[20] What policy advisory processes and mechanisms in support of the cabinet can best meet its information needs? And what are the required skills and qualifications of those public servants who seek to work as policy analysts?

Another lesson is the importance of having a reform effort spearheaded by a ministry willing to champion it, and to invest time and effort to making it a success. Simple advocacy of evaluation as being 'worthwhile' has been found unsuccessful—as per the principles-based approach to evaluation in the second generation Australian reforms, and as per the early experience of the Canadian government, for example (Mackay 1998; Schick 2001; Barrett 2001). It is invaluable to have additional agencies, such as the ANAO and the prime minister's department, also champion evaluation; the work of these agencies provides synergy and adds to the reform momentum. This is exemplified in the ANAO's ongoing support

[20] The dimensions of sound policy advice, and the role which evaluation can play in supporting this in the federal government, are discussed by Uhr and Mackay (1996).

for the Canberra Evaluation Forum, which has helped the forum continue in existence following DoFA's dismantling of the evaluation strategy.

Most areas of government reform, including the strengthening of performance information and evaluation systems, require a lengthy period of implementation. They also require ongoing review and adjustment, to ensure they are working in the manner intended. It takes a number of years before the benefits (and limitations) of a reform become evident; thus persistence and continuing commitment are prerequisites for success. A danger for any reform effort is when support for it is narrowly-based, dependent on the advocacy of only a few individuals such as the head of a central ministry. If that individual changes job, then the push for reform can easily become stalled. This suggests the merit of pushing rapidly to institutionalize reforms, with considerable effort necessary to raise awareness and to strengthen the breadth and depth of commitment to them. The urgency of this push for institutionalization can be assuaged if a broader coalition of support—in effect, a diversification of champions—has been created within the civil service.

The Australian experience provides a complementary insight: while it is difficult and time-consuming to build up a performance evaluation and management system, and to strengthen the role of a finance ministry in the policy advising and budget process, these can be dismantled or diminished relatively quickly.

Under the two generations of reform in Australia, the formal requirements and principles for performance information and evaluations were less important than how they were implemented. It is certainly useful to have a Cabinet decision (or in other countries a law or decree) to provide the authority for such information but this is not enough. More important is the nature of incentives to produce this information and to actually use it for the purposes for which it is intended—to support budget decision-making, for sound management, and for accountability. This illustrates the importance of not being overly supply-driven when it comes to the specification of what performance information and evaluation findings should be produced. Ensuring there is strong demand for this information is a key condition for success. Thus it helps to have a sound understanding of who needs what information, for what purposes, and when. Only then will it be possible to design a performance evaluation and management system which meets these needs—such a system will need to answer the question: 'what mix of performance information and evaluations will meet the needs of key stakeholders?' And as awareness of, and demand for, such information grows, so the types of M&E information which is available may need to grow.

It is a mistake to view the implementation of a public sector reform simply as a technical issue to be resolved. Information needs of central agencies or the parliament can be satisfied if there are sufficiently strong external pressures on line departments to provide this information. But if a reform is to be truly internalized by a department—thus sufficient to change their own planning and management processes—then substantive cultural change will be required. Such a change was achieved within DoF because of the high level of commitment of two successive departmental secretaries. Unfortunately, there are no easy ways to achieve cultural change either within a department or in a civil service as a whole.

Another common mistake when designing a performance system is to over-engineer the information —in terms of both quantity and quality—which it will provide. This mistake can be made by having an overly-comprehensive system of performance indicators, or by commissioning a large and costly number of sophisticated evaluations. In many countries this has proved not only wasteful but also counterproductive. Those responsible for providing the M&E information will have no incentive to produce quality information of a timely nature if they perceive the information is not being fully used. And if the quality of information starts to decline, this can further undermine the demand for it.

A final issue which is difficult to resolve is that of the objectivity and independence of evaluation (Feinstein 2003). External evaluations are often regarded as providing objectivity, but they can suffer from inadequate program knowledge; and the level of ownership by the agency being evaluated is often low, which can lead to poor internal utilization of evaluation findings. On the other hand, self-evaluations are often viewed as lacking objectivity, as being self-serving, and with insufficient evaluation rigor. In Australian departments, evaluations are often conducted by specialist units within the department but independent of the line areas responsible for program management.[21] Important evaluations are often overseen by steering committees which include central ministries and other line ministries—these would be relied on to ensure objectivity, rigor, and that difficult issues are addressed in the evaluation. This approach also ensures a high level of ownership and utilization of the evaluation findings by the key evaluation stakeholders.

[21] By contrast, the World Bank employs a comprehensive and systematic system of M&E which includes both self-evaluation by operational units and independent evaluation by the Bank's Operations Evaluation Department (OED). The architecture of this system is analyzed in OED (2003a) and the extent of independence actually achieved is discussed in OED (2003b).

REFERENCES

ANAO (Australian National Audit Office). 1991a. *Implementation of Program Evaluation—Stage 1.* Efficiency Audit Report No. 23, 1990-91. Canberra: Australian Government Publishing Service (AGPS).

___. 1991b. *Evaluation in Preparation of the Budget.* Efficiency Audit Report No. 13, 1991-92. Canberra: AGPS.

___. 1992a. *Program Evaluation in the Departments of Social Security and Primary Industries and Energy.* Efficiency Audit Report No. 26, 1991-92. Canberra: AGPS.

___. 1992b. *Auditing Program Evaluation—ANAO Performance Auditing Guide.* Canberra: AGPS.

___. 1992c. *Department of the Treasury—Procedures for Managing the Economic Policy Program.* Efficiency Audit Report No. 36, 1991-92. Canberra: AGPS.

___. 1993. *Program Evaluation—Strategies, Practices and Impacts—Industry, Technology and Regional Development Portfolio.* Efficiency Audit Report No. 35, 1992-93. Canberra: AGPS.

___. 1996. *Performance Information—Department of Employment, Education, Training and Youth Affairs.* Performance Audit Report No. 25, 1995-96. Canberra: AGPS.

___. 1997. *Program Evaluation in the Australian Public Service.* Performance Audit Report No. 3, 1997-98. Canberra: AGPS.

___. 2000. *Family Relationships Services Program—Department of Family and Community Services.* Performance Audit Report No. 24, 2000-01. Canberra: AGPS.

___. 2001a. *Performance Information in Portfolio Budget Statements.* Performance Audit Report No. 18, 2001-02. Canberra: AGPS.

___. 2001b. *Developing Policy Advice*. Performance Audit Report No. 21, 2001-02. Canberra: AGPS.

___. 2002. *Performance Information in the Australian Health Care Agreements—Department of Health and Ageing.* Performance Audit Report No. 21, 2002-03. Canberra: AGPS.

___. 2003. *Annual Performance Reporting*. Performance Audit Report No. 11, 2003-04. Canberra: AGPS.

ANAO/DoF (Department of Finance). 1996. *Performance Information Principles.* Canberra: ANAO.

Barrett, Pat, Auditor-General of Australia. 2001. 'Evaluation and performance auditing: sharing the common ground. A review of developments.' Address to the Australasian Evaluation Society, Canberra, 10 October.

Bemelmans-Videc, Marie-Louise, Ray Rist and Evert Vedung (eds). 1998. *Carrots, Sticks & Sermons: Policy Instruments and Their Evaluation.* New Brunswick and London: Transaction Publishers.

Boyle, Richard and Donald Lemaire. 1999. *Building Effective Evaluation Capacity: Lessons from Practice*. New Brunswick and London: Transaction Publishers.

Chan, Matthew, Mark Nizette, Lisa La Rance, Charles Broughton and Derek Russell. 2002. 'Australia'. *OECD Journal of Budgeting* 1(4): 35-69.

Crossfield, Len and Anne Byrne. 1994. *Review of the Evaluation Function in DEET.* Canberra: Department of Employment, Education and Training, and Department of Finance.

DEWR (Department of Employment and Workplace Relations). 2002. *Annual Report 2001-02.* Canberra: AGPS.

DoF (Department of Finance). 1991. *Handbook of Cost-Benefit Analysis.* Canberra: DoF.

___. 1994a. *Doing Evaluations—a Practical Guide.* Canberra: DoF.

___. 1994b. *The Use of Evaluation in the 1994-95 Budget.* Finance Discussion Paper, DoF, Canberra.

___. 1996. *Performance Information Review: Overview of the First Year of the Review.* Canberra: DoF.

DoFA (Department of Finance and Administration). 2000. *The Outcomes & Outputs Framework: Guidance Document*. Canberra: DoFA.

___. 2003. *Annual Report 2002-2003*. Canberra: AGPS.

Edwards, Meredith. 2002. 'Public sector governance—future issues for Australia.' *Australian Journal of Public Administration* 61 (2): 51-61.

FaCS (Department of Family and Community Services). 2001. *Research and Evaluation Framework.* Canberra: FaCS.

___. 2002a. *Annual Report 2001-02.* Canberra: AGPS.

___. 2002b. *Portfolio Budget Statements 2002-03. Family and Community Services Portfolio.* Canberra: Commonwealth of Australia.

___. 2002c. *Research and Evaluation Digest.* Canberra: FaCS.

Feinstein, Osvaldo. 2003. 'The World Bank's evaluation experience: evolution and lessons'. In Jay-Hyung Kim (ed.). *Developing a Performance Evaluation System in Korea*. Seoul: Korea Development Institute.

Keating, Mike, and Malcolm Holmes. 1990. 'Australia's budgeting and financial management reforms'. *Governance: an International Journal of Policy and Administration* 3(2): 168-185.

Mackay, Keith. 1998a. *The Development of Australia's Evaluation System.* ECD Working Paper No.4. Washington D.C.: Operations Evaluation Department, The World Bank.

___ (ed). 1998b. *Public Sector Performance—The Critical Role of Evaluation.* Washington D.C.: Operations Evaluation Department, The World Bank.

___. 2002. *2002 Annual Report on Evaluation Capacity Development.* Washington D.C.: Operations Evaluation Department, The World Bank.

Mayne, John and Eduardo Zapico-Goni (eds.). 1997. *Monitoring Performance in the Public Sector: Future Directions from International Experience.* New Brunswick and London: Transaction Publishers.

OECD (Organisation for Economic Co-operation and Development). 2002. *Public Sector Modernisation: A New Agenda*. GOV/PUMA (2002), 2. Paris: OECD.

OED (World Bank Operations Evaluation Department). 2003a. *2003 Annual Report on Operations Evaluation*. Washington D.C.: World Bank.

___. 2003b. *Independence of OED*. OED Reach. Washington D.C.: Operations Evaluation Department, The World Bank.

Parliament of Australia, House of Representatives Standing Committee on Finance and Public Administration. 1990*. Not Dollars Alone—Review of the Financial Management Improvement Program.* Parliament of Australia, Canberra.

Russell, Derek. 2003. 'Performance management and evaluation: the Australian experience'. In Jay-Hyung Kim (ed.). *Developing a Performance Evaluation System in Korea*. Seoul: Korea Development Institute.

Schick, Allen. 1996. *The Spirit of Reform: Managing the New Zealand State Sector in a Time of Change*. Wellington: State Services Commission.

___. 1998. 'Why most developing countries should not try New Zealand's reforms'. *World Bank Research Observer* 13: 123-31.

___. 2001 *Does Budgeting Have a Future?* OECD (PUMA/SBO (2001) 4). Paris: OECD.

___. 2003. *The Performing State: Reflection on an Idea Whose Time Has Come But Whose Implementation Has Not.* OECD (GOV/PUMA/SBO (2003) 17). Paris: OECD.

SCRCSSP (Steering Committee for the Review of Commonwealth/State Service Provision). 2003. *Report on Government Services 2003*. Canberra: AusInfo.

The Sydney Morning Herald. 14 November 2003. *Threat to Schools: Perform or Be Shut Down*.

Treasury Board of Canada Secretariat. 2000. *Results for Canadians: A Management Framework for the Government of Canada.* Ottawa: Treasury Board Secretariat.

Uhr, John and Keith Mackay (eds.). 1996. *Evaluating Policy Advice: Learning from Commonwealth Experience*. Canberra: Federalism Research Centre of the Australian National University and DoF.

The Washington Post. 8 November 2003. *Education 'Miracle' Has a Math Problem*.

World Bank. 1997. *The State in a Changing World*. World Development Report. Washington, D.C.: The World Bank.

Other Papers in This Series

#1: Keith Mackay, *Lessons from National Experience*.

#2: Stephen Brushett, *Zimbabwe: Issues and Opportunities*.

#3: Alain Barberie, *Indonesia's National Evaluation System*.

#4: Keith Mackay, *The Development of Australia's Evaluation System*.

#5: R. Pablo Guerrero O., *Comparative Insights from Colombia, China and Indonesia*.

#6: Keith Mackay, *Evaluation Capacity Development: A Diagnostic Guide and Action Framework.*

#7: Mark Schacter*, Sub-Saharan Africa: Lessons from Experience in Supporting Sound Governance*.

#8: Arild Hauge, *Strengthening Capacity for Monitoring and Evaluation in Uganda: A Results Based Management Perspective*.

#9: Marie-Hélène Adrien, *Guide to Conducting Reviews of Organizations Supplying M&E Training*.

#10: Arild Hauge, *The Development of Monitoring and Evaluation Capacities to Improve Government Performance in Uganda.*

Other Recommended Reading

OED, *2002 Annual Report on Evaluation Capacity Development (AREC)*.

OED, *Monitoring and Evaluation: Some Tools, Methods and Approaches*, 2002.

Development Bank of Southern Africa, African Development Bank and The World Bank, *Developing African Capacity for Monitoring and Evaluation*, 2000.

K. Mackay and S. Gariba (eds.), *The Role of Civil Society in Assessing Public Sector Performance in Ghana*, OED, 2000.

Other relevant publications can be downloaded from OED's ECD Website:
>http://www.worldbank.org/oed/ecd/>