

EVALUATION CAPACITY DEVELOPMENT

The Australian Government's Performance Framework

No. 25 | April 2011

ECD WORKING PAPER SERIES

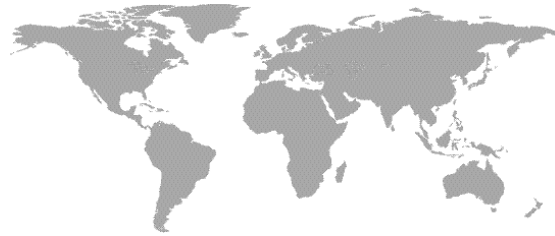


Keith Mackay



IEG

INDEPENDENT EVALUATION GROUP



ECD Working Paper Series • No. 25

The Australian Government's Performance Framework

Keith Mackay

www.worldbank.org/ieg/ecd

April 2011
The World Bank
Washington, DC



Copyright 2011
Independent Evaluation Group
Strategy, Communications, and Learning (IEGCS)
Email: ieg@worldbank.org
Telephone: 202-458-4497
Facsimile: 202-522-3125

Building monitoring and evaluation systems helps strengthen governance in countries — by improving transparency, by strengthening accountability relationships, and by building a performance culture within governments to support better policymaking, budget decision making, and management. A related area of focus is civil society, which can play a catalytic role through provision of assessments of government performance. IEG aims to identify and help develop good-practice approaches in countries, and to share the growing body of experience with such work.

The IEG 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.

The findings, interpretations, opinions, and conclusions expressed in this paper are entirely those of the author. They do not necessarily represent the views of the Independent Evaluation Group or any other unit of the World Bank, its Executive Directors, or the countries they represent.

Valuable comments on an earlier draft were received from Manuel Fernando Castro (IEG), Nidhi Khattri (IEG), Philipp Krause (PREM), Steve O’Loughlin (Australian Department of Finance and Deregulation) and Ximena Fernandez Ordonez (IEG).

How to cite this report:

Mackay, Keith. 2011. ‘The Australian Government’s Performance Framework.’ Evaluation Capacity Development Working Paper No. 25, World Bank, Washington, DC.

ISBN-13: 978-1-60244-170-5

ISBN-10: 1-60244-170-7

CONTENTS

Abbreviations.....	iv
Executive Summary.....	v
1. Context.....	1
2. First Period: 1987-1996.....	1
2.1 The Priority for Public Sector Reform.....	1
2.2 The Government's Evaluation Strategy.....	3
2.3 Other Government Monitoring, Evaluation and Review Activities.....	5
2.4 How Successful was the Evaluation Strategy?.....	6
3. Second Period: 1996-2007	
3.1 Changes in Public Sector Management.....	11
3.2 The 1996-2007 Performance Framework: Extent of Success.....	17
4. Third Period: 2007-11.....	22
4.1 Context.....	22
4.2 The Policy Cycle and Government Decision Making.....	22
4.3 Changes in Public Sector Management.....	23
4.4 Government Monitoring and Evaluation.....	24
5. Lessons From the Australian Experience.....	27
Bibliography.....	32

Abbreviations

ANAO	Australian National Audit Office
DoF	Department of Finance (until 1997); then Department of Finance and Administration (until 2007); then Department of Finance and Deregulation (currently)
GDP	Gross domestic product
M&E	Monitoring and evaluation
PBS	Portfolio Budget Statements
PEP	Portfolio Evaluation Plan
PM&C	Department of the Prime Minister and Cabinet

EXECUTIVE SUMMARY

There has long been a keen interest from countries around the world in Australia's experience in creating an evaluation system to support evidence-based decision making and performance-based budgeting. Australia's evaluation system lasted from 1987 to 1997, and during that time it was used to systematically evaluate all government programs every three to five years; these evaluation findings were used heavily by officials, ministers and the Cabinet in the annual budget process. The uses of these findings included the policy advice prepared by departments including the preparation of ministers' new policy proposals and departments' savings options submitted to the Cabinet for its consideration. More importantly, these findings were highly influential on the Cabinet's ultimate policy decisions. Finally, evaluation findings were also used widely within line departments in support of their ongoing management.

Achieving this situation required considerable effort; it is not an easy or quick endeavor to build an evaluation system. Success factors included strong champions at the most senior levels of the department of finance (DoF), which is the central budget office. The evaluation strategy was complementary to the efforts of DoF to reorient its staff away from a more traditional line-item focus toward a much greater focus on policy advice that was concerned with the efficiency and effectiveness of government spending. This in turn required a substantial evolution of DoF's staff and their skills sets. The DoF was not only the architect of the whole-of-government evaluation system but also its implementer. Having reformist government ministers was another success factor. Indeed, a number of senior and powerful ministers became strong advocates of the value of evaluation to the government.

The evaluations that were produced under the evaluation strategy were planned and conducted by line departments themselves. DoF and other central agencies would use their best efforts to influence the choice of which programs to evaluate, which issues to address in each evaluation, and in the actual conduct of these evaluations — by, for example, participating in evaluation steering committees. Thus evaluation in the Australian government was a collaborative effort.

The evaluation strategy was far from perfect. More than one-third of the evaluations it produced suffered from methodological weaknesses of one kind or another. Related to this, there were evidently insufficient evaluation skills available within the public service. Evaluations were conducted by line departments, and usually by the program areas within each department; a number of these apparently lacked the necessary skills. Some line departments avoided this problem by creating specialist evaluation units; it is these same departments that constituted the 'islands' of evaluation good practice that remained after the evaluation strategy had been abolished.

Its abolition was the result of a change in government in 1996, one of several important risk factors that can impinge on any facet of public sector management. Another such factor was the departure of key champions of monitoring and evaluation (M&E) and the arrival of M&E 'opponents' — key officials who were either hostile to M&E or to evaluation, or who pursued an extreme devolutionary approach to public sector management that virtually doomed any whole-of-government approach to M&E; a whole-of-government approach is necessary for performance-based budgeting, for example. A new performance framework was created during this period. It stressed monitoring indicators, largely in order to support accountability reporting to the parliament. The framework was principles-based, without central quality controls.

The abolition of the evaluation strategy in 1997 coincided with a substantial downgrading of DoF's involvement and skills in the areas of budget estimates and policy advising. At the same

time, large budget surpluses undermined the priority for spending discipline, for a disciplined policy process, and for evidence-based decision making. Thus Australia moved rapidly from a situation where there was a strong DoF, a fairly robust evaluation system, and a strong policy advising process, to a situation where all of these were largely undermined.

A notable feature of most of the second period, from 1996 to 2007, was the reliance on a performance framework — the *Outcomes and Outputs Framework* — that was based on a system of performance indicators. An advantage of performance indicators is that they are cheaper, simpler and quicker than evaluations. While performance indicators can be used to highlight examples of good or bad performance, a major limitation is that they fail to explain the reasons for this performance; in the absence of evaluations, it is difficult to apply the performance lessons elsewhere. Australia's performance framework during this period provides a strong example of how not to go about constructing a system of performance indicators. The framework encountered many conceptual and data difficulties. It also suffered from severe problems of implementation by line departments and agencies, and from a lack of effective oversight by the DoF.

The most recent period, from 2007 to the present time, has been marked by several initiatives that collectively are likely to increase the demand for and supply of monitoring information and evaluation findings. They include a renovated monitoring framework that focuses on government outcomes and outputs, and that reintroduces program budgeting. There is also a renewed focus on evaluation and review, with substantive interest from DoF in creating a whole-of-government system that avoids the weaknesses of the earlier evaluation system. More emphasis is also being placed on developing the policy skills of the public service, in the context of processes of government decision making that provide greater scope for policy advice from the public service.

Collectively, these changes have the potential to partially regain some of the ground lost during the period from 1996 to 2007. Only time will tell if the current efforts to renovate monitoring and evaluation in the Australian government are successful.

This paper updates two previous World Bank papers that reviewed the Australian experience with M&E and other performance-related initiatives. These papers (Mackay 1998, 2004) focused on the first two time periods addressed in this paper.

1. CONTEXT

Australia has a population of 22.3 million people in a land area of 7.7 million square kilometers. The country has enjoyed continuous economic growth over the past 18 years, and was one of the few developed countries not to experience a recession during the global financial crisis. Its economic success in recent decades is widely attributed to a combination of structural reforms that started in the mid-1980s and a prolonged boom in the demand for raw material commodities, of which Australia is a principal world supplier. This economic prosperity has helped lead to considerable budget surpluses for the federal government over most of the past decade.

The federal system of government has three tiers. In addition to the national government there are six state and two territory governments; all are based on the Westminster model. And lastly, there are local governments. Each of these governments is elected separately. Federal spending in 2010–11 is projected to be \$355 billion (currently about US\$373 billion), or 25.2 percent of gross domestic product (GDP). While the federal government collects the bulk of taxes, most public services are the responsibility of the states and territories — including education, health, welfare and community services, law and order, and infrastructure. The federal government provides considerable funding to the states and territories, and it does so in the form of block grants and tied funding. The federal government is responsible for social transfer payments to individuals, such as pensions.

The federal government is based in Canberra, and there were 164,600 public servants at June 2010. There are 20 government departments, and the three central coordinating departments are Finance & Deregulation (DoF), Prime Minister & Cabinet (PM&C), and the Treasury. The role of central departments in the budget process is described in Box 1.

Box 1: Central Departments and the Budget Process

DoF is essentially the central budget office; it coordinates the expenditure side of the budget. It oversees budget accounting including the financial framework. DoF also provides policy analysis of government outlays including all new policy proposals of line ministers; these DoF analyses accompany the spending proposals and are sent to the Cabinet committee that decides the budget (the Expenditure Review Committee). DoF also prepares savings options for budget cuts; line departments may prepare their own savings options. PM&C focuses on the governments' policy objectives and on whole-of-government policy issues. Treasury focuses on the taxation side of the budget, and on macroeconomic issues. All three central departments have desk officers who shadow the line departments. All three, and especially DoF, perform a 'challenge' function, in terms of analyzing and questioning the new policy spending proposals of line ministers.

2. FIRST PERIOD: 1987–1996

2.1 The Priority for Public Sector Reform

A reformist Labor government was elected in 1983, and Bob Hawke became the Prime Minister.¹ The new government faced a difficult macroeconomic situation including very tight budgetary constraints. One measure of the government's success is that it was able to reduce the share of federal government outlays in GDP from 30 percent in 1984–85 to 23 percent in 1989–90 — by international standards this is a very significant reduction. It did this by reducing its own spending

¹ He was replaced as Prime Minister by Paul Keating in 1991.

as well as the grants it paid to the other levels of government. At the same time, the government was committed to implementing a substantial reorientation of public spending toward the poorer members of society, and to reducing ‘middle-class welfare.’ This crisis situation provided powerful incentives for fiscal discipline and for a series of microeconomic reforms. The reforms were intended to change Australia from a highly regulated and protected marketplace to a much more flexible, open economy. The Australian dollar was floated, high tariffs reduced, the financial sector deregulated, and the flexibility of the labor market increased. Many government business enterprises were privatized or made to compete with the private sector.

The new government was also determined to implement a series of public sector reforms with the objective of improving government performance significantly. One aspect of these reforms was the desire to provide much greater autonomy to government departments and agencies. This was a ‘let the managers manage’ philosophy, and it involved the devolution of powers and responsibilities — encouraging better performance by providing much greater autonomy to managers. Departments were given autonomy in their spending of salaries and other administrative expenses, through a new system of consolidated running costs.

The budget system was changed very substantially. Australia led the world in introducing a medium-term expenditure framework in 1987, involving forward estimates of spending. They provided a spending baseline and freed up the budget process from a detailed, line item scrutiny of spending, to focus instead on changes in government policy and spending priorities. This simplified the budget process substantially, and it allowed a much more strategic approach to budget decision making (Keating and Holmes 1990; Blondal et al. 2008). The forward estimates also provided departments with greater surety about future resource availability.

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.

The departmental secretary of DoF was a major architect of many of the government’s public sector reforms. DoF’s role as budget coordinator and overseer of the spending of other departments also helped to ensure its influence on the reform agenda; DoF enjoyed the powerful support of the other central departments for this agenda. DoF was keen to get out of the detail of spending issues, where a traditional, zero-based budget process had meant that a substantial portion of its day-to-day work was narrowly focused on minor line item spending bids and disputes with departments. DoF wanted 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. The streamlined budget process facilitated exactly this kind of high-level focus. DoF’s 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 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. However, 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. This in-depth review was headed by a senior official from a line department. It found the following:

- A lack of integration of evaluation into corporate and financial decision making
- 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
- That the role of central departments in evaluation, especially DoF, was unclear.

2.2 The Government's Evaluation Strategy

DoF concluded that 'letting the managers manage' was insufficient; it was judged necessary to 'make the managers manage' and to make departments plan and conduct evaluations (Keating and Holmes 1990). Thus, in late 1988 the Minister for Finance secured the 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. The first, and arguably the most important, was that it provide fundamental information about program performance to aid the 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. Line departments are also accountable to the Cabinet and, in a sense, to central agencies such as DoF. The evaluation strategy to which the Cabinet agreed had four formal requirements for departments (Box 2).

Box 2: Evaluation Strategy — Formal Requirements

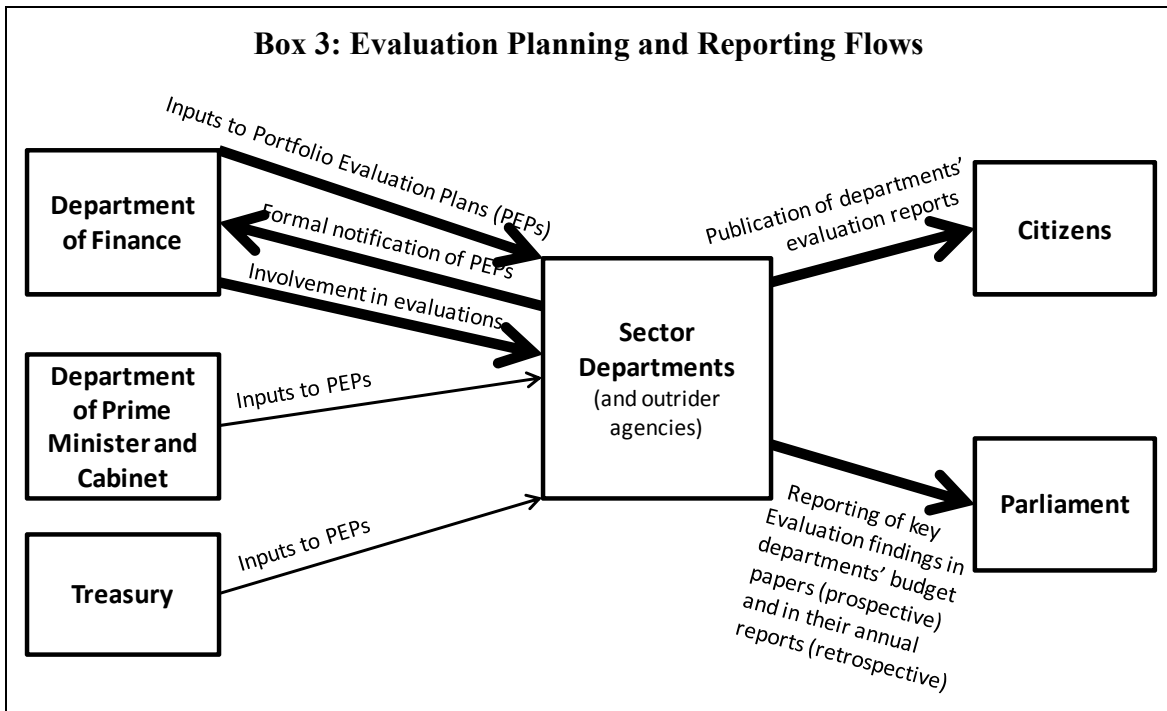
- That every program be evaluated every three to five years;
- That each portfolio (i.e., comprising a line department plus outsider agencies) prepare an annual portfolio evaluation plan (PEP), with a three-year forward coverage, and submit it to DoF — these plans were to comprise major program evaluations 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 existed 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.

The 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). The evaluations were to be conducted by the line departments (or agencies). The participation of the DoF desk officers in individual evaluations would typically involve their membership of the evaluation's steering committee, as well as their provision of comments on draft evaluation reports. The planning and reporting flows under the evaluation strategy are shown in Box 3.

Line departments had expressed serious concerns with the planned role for DoF, which they regarded as intrusive. Nevertheless, the Cabinet's agreement to the evaluation strategy was in the form of a formal Cabinet decision. An advantage of Westminster systems of government is that such decisions can be taken quickly; and for the federal government and its public servants, such decisions virtually have the force of a law. A disadvantage — compared, for example, with countries that have a Napoleonic system of government — is that such decisions can easily be reversed when there is a change in government.

While the evaluation strategy had three stated objectives, from the perspective of DoF — which was the primary architect and overseer of the strategy — the objective to which it devoted most attention was to support the Cabinet's decision making during the budget process. The senior management of DoF wanted to ensure that its line areas overseeing line departments were fully involved in the evaluation planning of departments and in the conduct of major evaluations. The immediate objective was to ensure that DoF budget officials were highly familiar with the quality and any limitations of the evaluations, were fully aware of their findings and recommendations, and were thus able to use them in their policy analysis work. Involvement of these officials in the evaluations would also substantially increase their knowledge of the evaluated program's objectives and the realities of its operating environment.



Achieving the necessary cultural change in DoF was easier said than done. Its budget analysts were capable, but tough minded and very conservative. Thus it was a challenge to change their

mindset from focusing on detailed line item costings to instead having much more of a high-level policy focus, concerned with the performance of government programs. The necessary cultural change was achieved by a number of means. There was strong leadership and advocacy by successive DoF secretaries and their deputies. Staff turnover was also required, with more emphasis on analytical and research skills, and less emphasis put on accounting skills. There was also some focused recruitment, so that evaluation experience became one of the selection criteria in the annual recruitment rounds for section heads in the department.

In the years following the Cabinet's agreement to the evaluation strategy, two reports, from a parliamentary committee and the national audit office, noted persistent unevenness in the scope of evaluation activity in departments (Parliament of Australia 1990; Australian National Audit Office [ANAO] 1991a). Both reports argued that DoF should be more active in 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 nine 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. It is important to note that Australia's M&E system essentially stressed *evaluation*, which was viewed as providing the necessary in-depth, reliable information on the efficiency and effectiveness of government programs. Performance information was understood to be important, but it was viewed as an issue for line departments to manage.

2.3 Other Government Monitoring, Evaluation and Review Activities

Other evaluation and review activities pre-dated the evaluation strategy. One example is ANAO performance audits. By the mid-1990s, the ANAO was producing 35 performance audits each year. Some of these focused on the evaluation activities of individual departments, and on the government's overall evaluation strategy. The ANAO's strong support for evaluation has helped to highlight and provide further legitimization to evaluation. Another set of activities related to evaluation were done by various government research bodies, such as the Bureau of Transport Economics, the Bureau of Industry Economics and the Industry Commission. Their work included research on microeconomic issues, as well as policy analysis and some evaluations.

As already noted, one limitation of the government's evaluation strategy was that it paid insufficient attention to the regular collection, reporting and use of performance information, via tools such as management information systems and performance indicators (Mackay 1998). 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 stating clear and achievable objectives for their programs, and in collecting and reporting meaningful performance information regularly. 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 regular performance information are often conceptually easier, and they were being done poorly. One explanation for this is that evaluation had been mandated, while the collection of performance information had not.

Thus in 1995 DoF secured the Cabinet's agreement to a rolling series of comprehensive reviews, staggered over three years, of the program objectives and performance information of all programs in all departments (see, for example, DoF 1996). DoF and each line department jointly

conducted the reviews, which were reported to their ministers and to the Cabinet. The reviews laid the basis for a much greater focus on performance information after 1997 (discussed in Section 3).

A parallel focus on performance information was achieved by the publication of annual reports on service delivery by the federal, state and territory governments; as noted earlier, most government services are provided at the state and territory levels. The decision to prepare these reports was taken in 1993, and the first report was published in 1995 (SCRCSSP 1995). This 700-page report covered \$38 billion in annual expenditure, or about 9 percent of GDP. It provided performance information on a range of government services, such as public hospitals; schools and vocational training; public housing; and police, court administration and prisons. The purpose of these reports was to provide greater transparency of performance and accountability for it. In addition, it was hoped that the reports would both support and spur improved performance by making comparisons across different jurisdictions — described as ‘yardstick competition’ — and help to identify best practice. Further discussion of these reports, and their uses, is provided in Section 4.

2.4 How Successful was the Evaluation Strategy?

Evaluation Planning

Since 1987–88, all government departments had prepared annual portfolio evaluation plans and these were meant to comprise the major evaluations of the department and its out-lier agencies. By the mid-1990s about 160 of these evaluations were underway at any given time. Most of these evaluations were major, in that the programs had significant policy or spending implications; however, a significant minority of these evaluations, particularly for the smaller departments, was of only 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 the Cabinet. Where DoF had difficulty in persuading departments, it sometimes approached the Cabinet directly to seek its endorsement of proposed evaluation topics and also detailed terms of reference.

The evaluation strategy’s Cabinet-endorsed, formal requirement that portfolio evaluation plans be prepared and submitted to DoF provided a powerful incentive to line departments to prepare plans and to take them seriously. Another influential factor was DoF’s 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 to departments as well as by identifying good practice examples. The DoF secretary also used this information to informally pressure line departments to improve their evaluation activities. In a number of performance 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, 1991b, 1992a, 1992b, 1992c, 1993, 1996, 1997; ANAO and DoF 1996). DoF also supported the creation of the Canberra Evaluation Forum, which involves monthly meetings of the evaluation community to discuss topical evaluation issues. The meetings were organized by a steering group of departments and other interested parties; this Forum still exists, and it attracts large audiences at its meetings.

Conduct of Evaluation

The formal requirement that all programs be evaluated every three to five years was influential in creating a climate of expectation that evaluation is the norm rather than the exception, involving the regular, comprehensive coverage of programs. 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 particular aspects of a program, or one of its sub-programs. The challenge is to ensure that the evaluation addresses the main objectives or problem issues, and this is a role in which DoF played an active role via persuasion concerning evaluation terms of reference, and via direct involvement in individual evaluations. DoF also provided guidelines on how to tailor the evaluation methodology according to the specific questions it was intended to address (DoF 1994a). These questions would depend on the size and importance of the program being evaluated, the maturity of the program, and the funds available to conduct the evaluation.

Evaluations were managed and conducted (or contracted out) by line departments. The rigor, depth and types of evaluation conducted varied considerably, as did their cost. At one end of the spectrum they comprised rapid reviews of program performance, using any available evidence, and investigating specific issues such as program efficiency, effectiveness or appropriateness. At the other end of the spectrum were rigorous impact evaluations using detailed data sets and complex statistical techniques. Evaluations also included cost-benefit analyses and performance audits, among others. All of these evaluations were subject to formal planning, terms of reference, and reporting. No statistics are available concerning how many of each different type of evaluation were conducted; however, the majority involved relatively rapid review. Of course, evaluations of major programs would often be major undertakings involving a range of evaluation tools, methods and approaches. A sample of evaluations analyzed by DoF ranged in cost (in 1993 prices) from \$56,000 to \$560,000 (DoF 1993).²

Most departments chose to set up evaluation units to coordinate their formal evaluation planning. At their smallest, these units comprised two or three individuals who provided some advice, quality review, and perhaps some training. In some departments, such as the department of employment, education and training, there was a separate branch — a specialist evaluation unit — 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 involved a wide array of external and internal stakeholders, either by participation in an evaluation steering committee, or less commonly by their participation in the evaluation team. Some evaluations were conducted by a central evaluation unit, with participation by the line program area, but it was more common for responsibility to rest with the program area. For 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 draft 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

² The annualized cost of these evaluations averaged less than one percent of the government's spending on these evaluated programs. Thus if the evaluation findings had led to only a very modest improvement in the efficiency or the effectiveness of the programs, the evaluations would have been cost-effective.

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 between 1995 and 1997, about half examined the delivery of products or services to external clients, and a further 30 percent were associated with matters internal to the department. One third of the evaluations examined the appropriateness of new or established programs, and 15 percent were directed toward 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 provided a large and rapidly 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 exposed them to peer scrutiny. The ANAO survey found that 75 percent of evaluations conducted in 1995 and 1996 were either released to the public or were available on request.

Evaluation Quality

The quality of evaluation reports is a much 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. These included: failure to adhere to the terms of reference; use of inappropriate methodologies; a divergence between data and conclusions; and unfounded recommendations. 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 be expected to vary enormously. This would be a significant problem if the intended audience of an evaluation is the Cabinet (to aid its decision making) or the parliament (for accountability 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 anger and condemnation of the 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 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 were achieved.

A disadvantage of this devolved approach was a lack of evaluation skills in many program areas and lack of experience in conducting or in outsourcing evaluations. It seems highly likely that this skills shortage was a major contributor to the reduction in quality of some 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 percent of departments were concerned about the lack of available training in advanced evaluation techniques, and this was a weakness of the Australian evaluation system.

As noted above, 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 in a number of ways: 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. That department, like some others, also put a lot of effort into training its staff to enhance their analytical and research skills.

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. It is important to appreciate the realistic limits to the influence of evaluation on ministers' or the Cabinet's decision making. Banks (2009a, p. 3) has stated this well:

'Policy decisions will typically be influenced by much more than objective evidence, or rational analysis. Values, interests, personalities — in short, democracy — determine what actually happens.

But evidence and analysis can nevertheless play a useful, even decisive, role in informing policy makers' judgments. Importantly, they can also condition the political environment in which those judgements need to be made.

Without evidence, policy-makers must fall back on intuition, ideology, or conventional wisdom — or, at best, theory alone.³

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 the Cabinet. DoF conducted several surveys of the extent of influence of evaluation findings on the budget proposals that were initiated by line ministers, prepared by their department officials, and submitted to the 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, and on the final decisions of the Cabinet. The close familiarity of DoF officers with these proposals and also with any evaluations or reviews on which they might draw, gave them an insider's perspective on the extent of influence of evaluation.

In the 1990–91, budget some \$230 million of new policy proposals submitted by line ministers was judged to have been directly or indirectly influenced by the findings of an evaluation. By

³ He also added that '*... good evidence can ameliorate or 'neutralise' political obstacles, thereby making reforms more feasible*' (Banks 2009a, p. 6).

1994–95 — the last year for which estimates were available — this had risen to \$2.3 billion.⁴ Measured in dollar terms, the proportion of new policy proposals influenced by evaluation rose from 23 to 77 percent over that period; and for most of these the influence of evaluation was judged by DoF officers to be both direct and major. These results indicate the importance that public servants, in their preparation of the details of new policy proposals, and ministers attached to having evaluation findings available. Ministers often expressed their view that it was valuable to them to have evaluation findings available for their Cabinet debates. Overall, it was very important to have had the active support of key Cabinet members in encouraging portfolios to plan and conduct high-quality evaluations. 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 \$500 million of savings options — or 65 percent of the total — was influenced by evaluation findings; again, the influence of evaluation on individual savings options 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 decision-making mechanism which made transparent the costs of competing policies and encouraged 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.

DoF officers were also surveyed for their judgments on the extent to which evaluation had influenced the Cabinet’s final decisions — as distinct from the influence of evaluation on the proposals drafted by officials and submitted to the Cabinet — 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 the Cabinet’s decision in 68 percent of the \$3.74 billion of proposals considered (new policy proposals plus savings options).⁵ The corresponding proportion for the 1993–94 budget, however, was only 19 percent 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).

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, p. 13).

There is also clear evidence that evaluation findings were used by line departments in their ongoing operations and internal management. While there are no detailed statistics concerning the use of evaluation by line departments for their own, internal management purposes, the 1997 ANAO survey found a high level of utilization of evaluation by line departments. The ANAO survey also 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

⁴ In that year, the government’s total outlays were \$123.6 billion. Of this amount, the large majority was specified in the forward estimates and was not subject to scrutiny. The ‘margin’ that was actively scrutinized was comprised of new policy proposals plus savings options.

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

resource allocation decisions and the design of service quality improvements for the benefit of clients. This high level of utilization reflected a strength of the Australian evaluation system: evaluation was essentially a collaborative effort involving DoF, other central departments and line departments. Although responsibility for evaluation was largely devolved to line departments, the involvement of the central departments in the planning and oversight of major evaluations helped achieve broad ownership of the evaluations themselves and of their findings.

3. SECOND PERIOD: 1996–2007

3.1 Changes in Public Sector Management

A conservative Coalition government was elected in March 1996, and John Howard became the new Prime Minister. The new government displayed a strong ideological preference for the private sector, which it regarded as being inherently more efficient than the public sector. The government expressed considerable unhappiness with the federal public service, and considered it to be rule-bound and caught up in red tape. The government emphasized market testing and the outsourcing of government activities wherever possible — a preference for non-government service delivery. Thus the government significantly reduced the size of the public service, from 143,000 in 1996 to 113,000 in 1999, a reduction of over 20 percent. Over the 11 years that this government was in office, it implemented a number of major changes to public sector management, discussed below. Collectively, these changes resulted in a completely new performance framework; Table 1 provides a comparison of this framework with those in the two other periods analyzed in this paper. The new framework embodied a mix of principles, expectations and formal requirements.

Role of Departmental Secretaries and the Public Service

Departmental secretaries had traditionally been career public servants, who were expected to display apolitical professionalism and impartiality. The new government, however, expected secretaries (and the public service as a whole) to be much more responsive to their political priorities, and expected them not to question government policy decisions or preferred options too closely (Podger 2005, quoted by Kelly 2006, pp. 12–13; Podger 2007). The government replaced six departmental secretaries. Reflecting a private sector paradigm, departmental secretaries became chief executive officers (CEOs), held accountable for results rather than for bureaucratic processes; it was considered undesirable to constrain the actions for CEOs by excessive administrative controls. These changes reflected what was, essentially, a ‘let the managers manage’ philosophy. It was analogous to the one adopted in the early 1980s

Departments and agencies were expected to operate on a much more business-like basis (Hawke 2007); departmental services and outputs, and in particular the outcomes that they were expected to lead to, were viewed as being, in effect, purchased by the government via the annual budget. This outcomes-based, purchaser-provider model had its intellectual basis in the New Zealand public sector reforms.⁶

⁶ This particular New Zealand reform is often viewed as unsuccessful in that country. See, for example, Schick 1996, 1998; Campbell 2001.

Table 1: Australia's Performance Framework — Key Aspects			
Key Aspect	1987–96	1996–2007	2007–10
Australian Public Service (APS)	<ul style="list-style-type: none"> • Cohesive public service; central rules, standards – e.g., pay, classifications, terms of employment 	<ul style="list-style-type: none"> • Public service downsized and balkanized; individual employment contracts; heavy use of business consultants; departmental secretaries often on three-year contracts 	<ul style="list-style-type: none"> • Efforts to renovate public service, e.g., regarding policy skills; moves to recentralize some functions, e.g., procurement, pay grades
Philosophy underlying public sector management	<ul style="list-style-type: none"> • Substantial devolution to departments; central requirements, e.g., evaluation, to 'make the managers manage' 	<ul style="list-style-type: none"> • Very high level of devolution — 'let the managers manage'; reduction in red tape; much greater reliance on private sector 	<ul style="list-style-type: none"> • Some recentralization, with heavy emphasis on encouragement; 'let the managers manage'; further reduction in red tape
Policy cycle	<ul style="list-style-type: none"> • Formalized, disciplined; heavy reliance on analysis by public service; Expenditure Review Committee (ERC) at center of budget process 	<ul style="list-style-type: none"> • Much less disciplined; greater reliance on non-APS policy advice; many policy/ expenditure decisions taken in Prime Minister's Office; ERC relatively weak 	<ul style="list-style-type: none"> • Decision-making initially in hands of 4 key ministers; now greater reliance on budget/ERC processes; APS policy skills to be strengthened
Role of the Department of Finance (DoF)	<ul style="list-style-type: none"> • Powerful, respected, high level of policy skills; heavily involved in scrutinizing new policy proposals — the 'challenge' function; responsible for budget estimates; heavily involved in evaluation 	<ul style="list-style-type: none"> • Severely downsized; small role in budget estimates and low financial management skills (until after 2002); low policy skills; little or no evaluation involvement; passive oversight of Outcomes and Outputs Framework; strategic reviews managed by DoF (from 2006) 	<ul style="list-style-type: none"> • Increase in staff numbers; refurbished financial management skills; role in reducing regulation and red tape; strategic reviews, and prospect of a rejuvenated evaluation approach
Evaluation	<ul style="list-style-type: none"> • Formal strategy and requirements (from 1987); enforcement by DoF; heavy utilization in policy advice and by ERC; evaluation use by line departments 	<ul style="list-style-type: none"> • Evaluation deregulated; only a few remaining evaluation islands among departments/agencies; small number of strategic reviews (from 2006); no systematic use of evaluation in budget process 	<ul style="list-style-type: none"> • Flurry of reviews after 2007; continuation of strategic reviews; no systematic use of evaluation in budget process, and major investment decisions taken without benefit of evaluation; agency reviews to be conducted in future; possible rejuvenation of evaluation in near future
Performance information (PI), program objectives, accountability	<ul style="list-style-type: none"> • Program budgeting (1986 on); evaluations usually published; only late attention to performance indicators via reviews of PI, program objectives (from 1994); federal/ state reporting of service delivery performance (from 1995); formal reporting requirements (annual reports, PBSs) 	<ul style="list-style-type: none"> • Program budgeting abolished (from 1999); new Outcomes and Outputs Framework for formal reporting, based on performance indicators (1999); principles-based, no quality control by DoF; accrual accounting (1999); evaluations rarely published; federal/state reporting of service delivery performance 	<ul style="list-style-type: none"> • Outcomes and Programs Framework, based on performance indicators, and now including program budgeting; evaluations rarely published; federal/state reporting of service delivery performance; citizen surveys planned

The Policy Cycle and the Role of DoF

The election of the Howard government in 1996 resulted in greater contestability in policy advice and debates. The government diversified its sources of policy advice and preparation, including business consultants, think-tanks and academics. It displayed an ideological preference for advice from outside the public sector. Consulting companies not only provided policy inputs but also prepared detailed policies (Banks 2009a). Thus by 2007, the total expenditure on consultants was \$484 million, or about the same amount as the entire senior executive service in government, which comprised some 2,700 staff. At the same time, more and more policy and budget decisions came to be taken in the Prime Minister's office, with less reliance on the Cabinet or the policy process.

For most of this period, Australia enjoyed large budget surpluses, mainly as a result of a strong economy. This undermined the need for budget discipline, and thus the government often decided a large proportion of budget spending right at the end of the budget, and outside of the formal budget (i.e., policy) process — *'controlling expenditure has become ever more difficult'* (Blondal et al. 2008, p. 29). With this lack of budget discipline, M&E information about the performance of government programs was essentially irrelevant; the large budget surpluses after 1998–99 removed a main driver of a performance orientation by government.

DoF was an early casualty in this changed environment, and it went through a 'traumatic' period in the late 1990s and early 2000s (Blondal et al. 2008). The government had appointed a new departmental secretary to head DoF in 1997, and he was a key proponent of the outcomes and outputs framework (discussed below). He disbanded the DoF unit responsible for advising on evaluation and public sector management reform issues, in order to focus the department on accrual budgeting and the contracting out of government activities. He wanted departments and agencies to take the main responsibility for budget estimates, including forward estimates, and for reporting their performance via the outcomes and outputs framework (Wanna and Bartos 2003).

'The then Department of Finance wanted agencies to assume still greater responsibility and saw its role then as akin to an investment bank that only needed the broadest of overviews... The number of analytical budget branches — the ones mirroring spending ministries — was reduced from 13 to 4, creating a vacuum at the centre. The Department lost much of its corporate knowledge of expenditure policy issues, such that it became difficult for it to exercise an effective budget 'challenge' function vis-à-vis spending ministries and agencies' (Blondal et al. 2008, p. 11).

This situation led the normally diplomatic 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 them to adequately advise the Minister' (OECD 2002, p. 4).

For a finance ministry, with its traditionally leading role in the budget and related policy formulation processes of government, this criticism is fundamental.

A new DoF departmental secretary was appointed in early 2002; most senior budget officials were also replaced at that time. A review of various Howard-era public sector management changes was conducted later that year, and it was decided to increase the number of staff in DoF who had financial skills necessary to monitor agencies' financial performance, cash flows,

financial management compliance, and ability to estimate expenditures (Hawke 2007). These are normally the most basic functions for a department of finance, and it is a telling observation that even these had been degraded in the early years of the Howard era. The number of staff in the budget branches has since risen to about the level that it was before 1996, although their focus was reportedly much more toward financial issues, rather than on the provision of policy analysis and advice. It is somewhat ironic that, in an era when much greater attention was intended to be paid to government *results*, the central budget office had shifted its focus from a substantial emphasis on government results and performance, to instead a focus on financial inputs. This represents a much more old-fashioned view of the role of a DoF.

Abolition of the Evaluation Strategy

There had been considerable opposition on the part of line department secretaries to the creation of the evaluation strategy in 1987, mainly on the grounds that they viewed it as an intrusion on their areas of responsibility. However, once the strategy had been established, there was little opposition to it during the following decade. All this changed after there was a change in government. With the advent of the Howard government in 1996, and consistent with the government's push for less 'red tape,' line departments pressed for less oversight by, and reporting to, DoF. They also took the opportunity to highlight the burden to them of planning and conducting evaluations. One particular concern was the requirement 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 the Cabinet, detailed and elegantly worded plans were not necessary to achieve that objective. These arguments immediately found a receptive audience with the new government, which therefore decided in 1997 to abolish the evaluation strategy, including its formal requirements. The current DoF departmental secretary has recently argued that the decision to abolish the evaluation strategy reflected a view that it was too cumbersome, too resource intensive for all parties, and that there was a problem with obtaining suitable skills to undertake the evaluations (Tune 2010).

In addition to the abolition of the formal strategy, the evaluation support which the Department of Finance — renamed the Department of Finance and Administration (DoF) — had provided to line departments was discontinued. DoF still encouraged line departments to conduct evaluations of key policies and programs, with a recommendation that they be conducted on a five-year cycle, but there was no longer any formal requirement for this. This principles-based approach left the decision to departmental secretaries as to whether or not to conduct any evaluations.

Annual reports, which are tabled in parliament, were meant to report the implications of any evaluations that have been conducted in the past year; however, an ANAO report (2003) concluded that they often did not report evaluations or their findings.⁷ Portfolio budget statements were also meant to indicate any planned evaluations in the coming year. In addition, new policy proposals were required to report any evaluation evidence that existed. But there was no requirement that all new policy proposals should include an evaluation plan if the proposal were accepted. Instead, the proposing Minister was expected to present the 'business case' for the proposal, as well as the strategy for monitoring results — reflecting the government's business-oriented mindset.

⁷ It is not known if such omitted findings reflected adversely on the performance of the departments.

There was a new requirement to review ‘terminating programs’ — those with a defined end date — and ‘lapsing programs’ — those that were expected to continue but where funding would have to be reauthorized — and to report the findings to the Cabinet’s Expenditure Review Committee (McPhee 2008). These reviews focused on asking if and how the government should continue to be involved in a program. A team of Organisation for Economic Co-operation and Development (OECD) reviewers concluded that these reviews *‘became a mechanical and ineffective exercise which rarely resulted in any significant changes to the programmes, despite an abundance of reviews — there were 149 lapsing reviews between the 2004–05 and 2006–07 budgets’* (Blondal et al. 2008, p. 19). This requirement has been abolished.

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

In this environment, the decision whether or not to conduct and use evaluations was essentially left to individual departmental secretaries to make. In the absence of strong accountability pressures on line departments, and in an environment where policy and budget processes had been significantly weakened, any incentives for line departments to take monitoring and evaluation seriously would have to be internally generated. By 2003, some six years after the deregulation of evaluation, there were still some departments that devoted considerable priority to evaluation, and some could be considered good practice in a number of respects; these included the departments of family and community services; employment; and health (Box 4). These departments can be considered to be islands of good-practice. However, it seems that even these departments tended to conduct evaluation less frequently, in order to address particular issues on a selective basis. It not clear that any departments continued to conduct evaluation as regularly or as systematically as they did under the evaluation strategy of the Hawke/Keating era. There has been no investigation of the reasons why these departmental islands of good-practice evaluation persisted after the abolition of the government’s evaluation strategy — this is an issue that merits further research. However, one likely reason includes the personal commitment of some key individuals in these departments — i.e., champions of M&E. Another is a corporate culture and mindset of professional staff in the areas of health and education, whose professional training underscores the value of research, evaluation, monitoring and statistics.

Toward the end of the Howard government era (1996–2007), another DoF departmental secretary was able to secure the Cabinet’s agreement in October 2006 to create a ‘strategic review framework.’ The objective of these reviews — which continue to the present day — is to assess the efficiency, effectiveness and policy alignment of a limited number of government initiatives (DoF 2007); however, the reviews are not intended to identify possible savings. Review topics are chosen by a Cabinet committee, based on advice from DoF and the other central agencies. The topics are typically large, high-priority, complex and cross-agency in nature, such as the strategic review of the government’s 62 climate change programs (Wilkins 2008). Five strategic reviews were conducted in 2009, for example. These reviews are usually led by eminent persons, such as retired senior public servants, assisted by four to six staff seconded from various departments. Review teams are located in DoF, where there is a secretariat to manage the reviews. The current secretary of DoF considers that some of the reviews have been of excellent quality, although other reviews have been hampered by a lack of agency involvement and cooperation (DoF 2010).

The Outcomes and Outputs Framework

The main formal performance-related innovation of the Howard government was its *Outcomes and Outputs Framework*, introduced in 1999 (Box 5). This was intended to further strengthen the

shift in focus from inputs to results. It required departments to agree with their ministers the outcomes — the government objectives — toward which they were working, and to also agree the departmental outputs, such as service delivery for specific target groups, that would be produced to help achieve these outcomes (DoF 2000; Chan et al. 2002; Russell 2003). By 2004 the 145 departments and agencies collectively had 199 outcomes, and usually between 1 and 10 outputs each⁸ (McPhee 2005; Blondal et al. 2008). CEOs were given the authority to allocate appropriations across different types of output to achieve the desired outcomes. DoF provided guidance regarding performance management and performance reporting. However, DoF did not enforce compliance with the framework; it did not specify minimum or mandatory requirements (ANAO 2007). Thus the definition of individual outcomes and outputs was left to departments and agencies to decide. The reporting flows under this framework are shown in Box 6.

Box 4: Good-Practice Departments in 2003 — Evaluation Islands

Department of Family and Community Services. This department provided income support to the disabled, rent assistance, and housing grants to the states, with annual budget appropriations of about \$46 billion. It reported 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 included a substantive use of evaluation findings (FaCS 2002a, b). The department regularly prepared a research and evaluation plan, including 145 research/evaluation projects — the total budget for this work was \$26 million per annum, equivalent to about 1.4 percent of the department's operational budget (FaCS 2001, 2002c). Funding was provided to 6 university research institutes, and for 3 longitudinal data sets. The research emphasis was on behavioral outcomes, and this was an important dimension of the evaluations which were conducted.

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

Department of Employment and Workplace Relations. It had a wide range of performance indicators derived from its administrative information system and from surveys, and these provided information on outputs and outcomes (DEWR 2002). The department also had an active evaluation branch which had 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 formed the basis for the evaluations conducted by the department (or sometimes conducted by universities); these evaluations were typically overseen by steering committees comprising central ministries (such as DoF) 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 during the Hawke/Keating era (Crossfield and Byrne 1994).

The department had a service charter with service standards such as response times. Regular surveys of client satisfaction were 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 was the feedback sought systematically from the ministers who had responsibility for the department — they were asked to rate each piece of policy advice work according to 11 criteria, such as logic, accuracy, creativity, and timeliness. This type of performance feedback can be expected to provide very strong incentives to public servants to provide high-quality policy advice. The ANAO has commended the department's public reporting of this performance information (ANAO 2001b).

⁸ The ANAO (2007) found that one agency had 52 outputs.

Box 5: Outcomes and Outputs Framework — Objectives and Potential Benefits

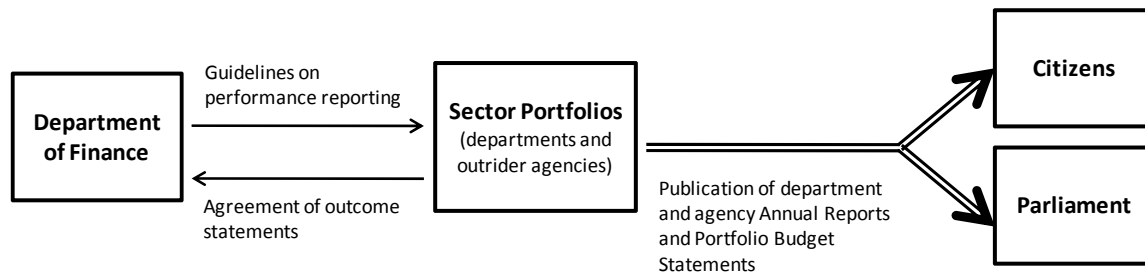
The principal objectives of the framework mirror the objectives of the earlier evaluation strategy:

1. To support the government's policy development;
2. To support and strengthen departments' internal management, including staff learning;
3. To strengthen external reporting, for accountability purposes.

On paper at least, the framework also has the potential to:

4. Clarify objectives and expected standards of performance — this understanding is crucial to ensure purchaser-provider arrangements work well;
5. Clarify for departments their results chains — i.e., the logical (and desirably evidence-based) links between spending, activities, outputs, and contribution to desired outcomes;
6. Facilitate benchmarking comparisons across departments, and between the public and private sectors. This would make it easier to select least-cost providers; and
7. Promote shared objectives among departments — the concept of 'joined-up government.'

Box 6: Outcomes and Outputs Framework — Performance Reporting Flows



The framework required the formal reporting by departments of their performance in achieving these outcomes and outputs. This performance reporting relied on the publication of performance indicators that measure departments' outcomes and outputs — including their quantity, quality and prices — as well as departmental efficiency and effectiveness. Those departments and agencies that provide services directly to the public are required to have service charters, with actual levels of performance to be included in their reports to parliament. Examples of outcome statements and performance indicators under the framework are shown in Table 2.

3.2 The 1996–2007 Performance Framework: Extent of Success

The Outcomes and Outputs Framework

The core part of the government's performance framework was the formal *Outcomes and Outputs Framework*. It relied on the collection and reporting of performance information, to aid policy development, departments' internal management, and external reporting for accountability purposes.

One inherent limitation of a reliance on performance information for these purposes is that while it can help provide a useful overview of program and agency performance, 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 sound analysis and interpretation of the data. A very good example of this type of analysis is provided by the federal/state/territory performance comparisons for a wide range of government services, discussed in Section 2. Fortunately, these published comparisons continued uninterrupted during the Howard era. Of course, while analyses of performance information are useful in making cross-jurisdictional comparisons, and can help suggest some possible reasons for good or bad performance, they do not provide depth of understanding nor can they provide definitive explanations of causal factors that explain performance. Thus while performance indicators can measure the different stages of a program's results chain — inputs, processes, outputs, outcomes and their final impacts — they do not demonstrate causal relationships. Such understanding can only be provided by evaluation; hence there is a complementarity between performance information and evaluation.

Table 2: Examples of Outcome Statements and Performance Indicators under the Outcomes and Outputs Framework — 2006–07 Budget

Portfolio	Example of an Outcome	Examples of Performance Indicators
Communications, Information Technology, and the Arts	Development of services and provision of a regulatory environment which encourages a stable and effective communications sector for the benefit of all Australians and an internationally competitive information economy and Information and Communications Technology industry	<ul style="list-style-type: none"> • Qualitative evaluation, analysing the department's contribution to developing the ICT sector • High quality program administration as measured by (i) number (and trend) of funded grants and grants administered; (ii) percent and number of applications processed within agreed time frame; (iii) percent and number of key processes completed within target period (plus analysis where applicable)
Families, Community Services and Indigenous Affairs	Families and children have choices and opportunities	<ul style="list-style-type: none"> • Percentage of funded projects assessed as meeting the objectives set for the programme • Number of families assisted directly and indirectly • Percentage of outside school hours care services satisfactorily participating in quality assurance
Immigration and Multicultural Affairs	Contributing to Australia's society and its economic advancement through the lawful and orderly entry and stay of people	<ul style="list-style-type: none"> • The extent to which entry of migrants results in a positive impact on living standards as measured by models of the economic impact of migration • Number of onshore applications (persons) finalized • Performance against service standards (median processing times)
Transport and Regional Services Portfolio	Assisting regions to manage their own futures	<ul style="list-style-type: none"> • Regions are assisted to manage their own futures • Information on Australian Government programmes and services is available to all Australians • Australian Government interests in all Territories are managed
Sources: Portfolio Budget Statements, 2006–07 Budget, for each portfolio. See www.budget.gov.au/ .		

Problems with the performance information presented in the two main reporting vehicles — annual reports and portfolio budget statements — became evident soon after the Outcomes and Outputs Framework was introduced (Box 7). These were highlighted frequently in subsequent years, but were never satisfactorily addressed during the Howard government (1996–2007).

Blondal et al. (2008) have argued that the outcomes focus was not integrated into budget decision making or in agencies' daily operations. Instead, the focus of departments and agencies, of the Cabinet, and of parliamentarians, remained on programs.⁹ The main, and probably the only, purpose of portfolio budget statements and annual reports was as an external reporting requirement, for accountability purposes. But even as an accountability tool, the Outcomes and Outputs Framework appears to have failed. This is evidenced by the various ANAO performance audits of aspects of the framework, and by the various parliamentary enquiries cited above. The ANAO noted, for example, that *'Performance information generally 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'* (ANAO 2003, p. 13). Thus it is no surprise that the

Box 7: Outcomes and Outputs Framework — Significant Problems

- 'Outcomes' were usually defined in a single sentence, in very broad, aspirational terms, rather than trying to state in specific terms the desired impact of the government's activity. There was a lack of performance information to tell if outcomes had been achieved or not (ANAO 2001a, 2007; SSCFPA 2007; Podger 2009).
- There were 200 outcomes in total, and so they had a high level of aggregation. Departments and agencies did not have any shared outcomes; they each preferred to have their own outcomes, for which they alone were accountable (ANAO 2007; AGRAGA 2010).
- There were poor logical links between many outputs and outcomes (Podger 2009).
- Targets or benchmarks were typically not specified for outputs (ANAO 2003, 2007).
- Departments and agencies tended not to report unmet targets, and often did not discuss areas where performance was poor (ANAO 2003; JCPA 2004).
- There was insufficient performance information concerning efficiency and effectiveness, and too much focus on activities undertaken (ANAO 2007).
- The majority of agencies with purchaser-provider arrangements did not include performance information on them in their portfolio budget statements (ANAO 2007).
- The specification of outputs and outcomes differed between departments and agencies, making comparisons very difficult (Blondal et al. 2008).
- Definitions continued to change over time, even a decade after the framework was introduced (Blondal et al. 2008; Webb 2010).
- Portfolio budget statements reported the forward estimates of spending, but presented no information concerning forward estimates of outputs or outcomes (ANAO 2007).
- There was no 'clear read,' i.e., there was a lack of corresponding and comparable performance information between the performance promised in portfolio budget statements and the performance actually delivered and reported in annual reports (Murray 2008).
- The Senate found the outcomes structure confusing. It strongly preferred program-based

⁹ Programs in Australia are objective based. They comprise a set of related activities with a common objective (i.e., a desired outcome). While they have a conceptual relationship with the outcomes focus of the Outcomes and Outputs Framework, in practice programs have been defined in a much more detailed manner. Programs in the Australian government are also organization based.

performance reporting and budgeting (SSCFPA 2007, quoted by Mulgan 2008).

Auditor-General noted the dissatisfaction of parliamentary committees with the performance information provided by departments (Barrett 2003). The extent to which the various objectives of the Outcomes and Outputs Framework were achieved is summarized in Table 3.

Table 3: Outcomes and Outputs Framework — Extent of Success		
Objectives	Results	Extent of Success
Support the government's policy development	Framework information does not appear to have been used for policy development	Weak
Support and strengthen departments' internal management, including staff learning	Framework information was often not used for internal management	Mediocre
Strengthen external reporting, for accountability purposes	External reporting was weakened significantly	Very poor
Clarify objectives and expected performance	Unclear results; a majority of agencies with purchaser-provider arrangements did not report performance information on them publicly	Unclear
Understand results chains	There were poor logical links between many outputs and outcomes	Weak
Facilitate benchmarking comparisons	Agencies did not use common performance indicators; insufficient measurement of efficiency and effectiveness	Very poor
Promote shared objectives	Departments and agencies had no shared outcomes	Very poor

It is unclear whether any of the problems with the framework arose because of inherent deficiencies in the data — in the performance information collected by departments and agencies. The evidence concerning data quality appears to be mixed, and it is not possible to reach a clear conclusion in the absence of detailed data audits. However, there do appear to have been at least two fundamental problems with the framework. The first is its conceptual — some would say its ideological — basis. The framework assumed that government objectives could easily be captured in simple outcome statements, and that outcome-based purchaser-provider relationships — in effect between the elected government and departments and agencies — would be able to drive government performance. The first assumption is technically difficult to achieve, if not impossible, while the second assumption evidently did not work.

The second fundamental problem with the framework was the quality of its implementation, and the fault would appear to lie with departments and agencies, which were responsible for the performance reports that were published, and with the DoF, which was responsible for the design and oversight of the framework. The ANAO has noted the failure of agencies to comply with a number of the mandatory performance reporting requirements (ANAO 2007). The apparently low

priority paid by a number of departments to external reporting indicates the lack of strong incentives for them to take it sufficiently seriously.

DoF clearly displayed a ‘hands-off’ approach: it issued reporting guidelines and co-authored with the ANAO a better-practice guide on annual performance reporting (ANAO and DoF 2004), but it was not prepared to enforce the reporting requirements in an environment where the absence of ‘red tape’ and a high level of devolution to departments were seen as being more fundamentally important. DoF reportedly shrugged off repeated criticisms, such as those made by a Senate committee in 2000, concerning the inadequate information content of portfolio budget statements (Thomas 2009). The government-sponsored Murray inquiry into the implementation of the Outcomes and Outputs Framework concluded that DoF’s oversight had been ‘lackadaisical’, i.e., unenthusiastic (Murray 2008). It is ironic that, despite the abolition of the evaluation strategy, the new framework resulted in new bureaucratic requirements for line departments to produce and report a large volume of performance information. Thus, overall the new reporting requirements increased rather than reduced the reporting burden on departments and agencies. These reforms in the 1996–2007 period, thus ‘threw out the baby but kept the bathwater.’

Accountability and the Role of Parliament

The parliament, particularly the Senate, plays an important role in budget review and scrutiny, and in holding the government to account for its performance. Senators are elected by a system of proportional representation, and this virtually guarantees that the political party in power — typically the party with a majority of seats in the House of Representatives — will be in a minority in the Senate. Although the Senate has the constitutional power to block the government’s proposed budget, in practice it makes only relatively minor changes to the budget; the government would often accept these. The main opportunities for the Senate to scrutinize government performance are its review of portfolio budget statements (PBSs), and to a lesser extent, of annual reports. Despite the substantial weakening of the information content of PBSs as a result of the Outcomes and Outputs Framework, they remain the Senate’s main source of information on government performance. The Senate has eight committees that scrutinize the budget bills for portfolios. In addition to examining the information content of PBS, the committees also question portfolio ministers and senior officials of departments and agencies. Although the committees are supported by 150 staff, they reportedly have limited analytical resources available for budget scrutiny. In 2006–07 the committees collectively spent 700 hours on budget hearings, and heard more than 4,000 witnesses (Thomas 2009). Parliamentary committees also consider ANAO performance audits into specific issues (Box 8). Thomas has observed, *‘The Senate takes the function of scrutiny very seriously. The idea of the Senate as a house of review for both legislation and spending has over time become entrenched in the culture of the institution.’* Moreover, *‘...the Australian Senate is probably entitled to claim that among western democratic legislative bodies it takes the oversight function more seriously than any other, with the possible exception of the Senate in the United States’* (Thomas 2009, p. 388).

The quality of the Senate’s scrutiny of government performance is unclear. Some former departmental secretaries have been frustrated with what they regarded as superficial review of policy and performance issues by the Senate (e.g., Podger 2009). Others have argued that the use by Senate committees of performance information from portfolio budget statements and annual reports varies enormously: some committees have many questions on actual vis-à-vis promised performance; other committees have few such questions (e.g., Hawke 2007). In 2010 the government announced plans to create a Parliamentary Budget Office to support parliament by providing fiscal analysis and policy costing advice on budget-related matters.

4. THIRD PERIOD: 2007–11

4.1 Context

A Labor government was elected in November 2007, with Kevin Rudd as the new Prime Minister. He was committed to making substantial changes to the policies of the previous government, and was less favorably disposed toward the private sector. A former senior official in the Queensland government, he professed a belief in ‘evidence-based policy making.’ In the first eight months in office, he commissioned 140 government reviews of various issues, ranging from higher education to innovation policy (Symonds 2009). The collective reasons for this flurry of reviews have not been clearly articulated but appear to include the new government’s need — particularly for a government that had been out of power for almost 12 years — to clarify a range of policy issues and options, to gather evidence on these, and to engage in public debate on them.

Box 8: ANAO Performance Audits

The primary client of the Australian National Audit Office is the parliament. The ANAO provides the parliament with an independent assessment of selected areas of public administration, as well as assurance about public sector financial reporting, administration and accountability. The ANAO also views the executive government and public sector entities as important clients. The Auditor-General is appointed directly by the parliament.

With a staff of about 360, and a budget of \$87m (in 2010–11), the ANAO conducts performance audits and financial audits; the former are a type of evaluation, and they account for about one-third of ANAO resources (McPhee 2008). About 50–60 performance audits are performed each year. The ANAO explicitly does not pursue a ‘fault-finding’ role with these. Instead, it takes a ‘client approach,’ with the objective of improving public sector management by identifying ‘better practice’ and by making recommendations to improve the efficiency and effectiveness of government programs. About 90 percent of the ANAO’s performance audit recommendations are accepted by government departments; the remainder are usually partially accepted. Performance audits have been conducted on a wide range of topics, such as:

- AusAID’s aid to tertiary education
- The Australian Federal Police’s management of new policy implementation
- The management of overseas leased estate
- The Australian Taxation Office’s administration of the luxury car tax
- The service delivery of the Commonwealth Rehabilitation Service (for the disabled)
- Therapeutic goods regulation
- Indigenous employment in government service delivery.

For more information, see www.anao.gov.au.

4.2 The Policy Cycle and Government Decision Making

One feature of this period has been the growing willingness of senior officials to speak out publicly on their concerns about how the public service — its capacities and policy advising role — had deteriorated significantly since the mid-1990s, and about the lack of evidence-based

decision making by the government. Some of these criticisms relate to the Howard era, and to the need to address the accumulating problems that had arisen during that time. The Rudd government tried to address various problems with public sector management, including with the performance framework broadly defined. However, there were also growing concerns expressed over the style and approach to decision making of the Rudd government itself, with a widespread view that decision making was concentrated in his hands and with several close colleagues, rather than with the Cabinet or based on policy processes involving the public service in detailed policy analysis, advice and deliberation. These concerns culminated in the replacement, by Labor members of parliament, of Kevin Rudd as Prime Minister by Julia Gillard in June 2010.

The quality of the government's decision making has been criticized both implicitly and explicitly (e.g., ANAO 2010; Banks 2009b). The most criticized government investment in Australia in recent years is the decision to construct a National Broadband Network that will deliver high-speed fiber optic internet cabling to almost every home in Australia. The government initially estimated its total cost to be up to \$43 billion, or 3.5 percent of GDP. A rapid cost-benefit analysis conducted informally by respected economists indicated that in present value terms, the costs of this investment would exceed the benefits by between \$14 billion and \$20 billion (Ergas and Robson 2009). Despite repeated recommendations by the Productivity Commission, Organisation for Economic Co-operation and Development, Business Council of Australia, and Reserve Bank of Australia (Stutchbury 2010) that this investment should be subject to an in-depth economic appraisal using cost-benefit analysis, the government has insisted that a '*cost-benefit analysis would waste time and money*' (Communications Minister Stephen Conroy 2010).

The head of the Productivity Commission has criticized the government over its failure to use evidence in its industry policy (Banks 2009b, quoted by Mitchell 2009). Banks' concerns focus on the structural arrangements that support evidence-based policy making; and with the converse, that their absence constrains such policy making. He has listed these necessary structural arrangements as including: good data; sufficient time for evidence-building; good analytical methodology; evidence being open to scrutiny; capable, expert public servants; independence; and a receptive policy-making environment (Banks 2009a). Among this list, he regards the last as fundamental. However, he regards all these steps in an evidence-based decision-making chain as having significant weaknesses in Australia. As we shall see below, the government has announced several initiatives that should ameliorate these weaknesses.

4.3 Changes in Public Sector Management

The government has provided \$70 million over four years to the three central departments — PM&C, Treasury and Finance — to enable them to strengthen their policy capacities. The government accepted a detailed blueprint — *Ahead of the Game* — for the reform of government administration (AGRAGA 2010). These reforms constitute an attempt to fix a number of problems caused by the public service changes made by the Howard government over the 1996–2007 period. All departments and agencies are now being encouraged to strengthen their policy capabilities, and to do more research and evaluation. To support these efforts, the Australian Public Service Commission will make training available, from sources such as the Australian National University, in core areas such as strategic policy, policy analysis, implementation and regulation. The government also accepted that Australia lags other countries in systematically measuring departmental and agency performance, in areas such as strategy development, program delivery and organizational operations. Thus each department and agency will be reviewed at least once every five years. The reviews will be led by eminent external reviewers, and the review teams will have representatives from the departments of PM&C, DoF, the Public Service Commission, the agency itself and other agencies as appropriate. The reviews will also identify

ways to improve data on agency performance; this will also help in the development of performance benchmarks for public service agencies. Agencies will be chosen for review on the basis of ‘risk management principles.’ However, the reviews will not be published, at least initially.

The Outcomes and Programs Reporting Framework

Lindsay Tanner, who became the Labor government’s finance Minister after the 2007 election, had long criticized the quality and lack of transparency of budget documentation, including portfolio budget statements. As finance Minister, he implemented *Operation Sunlight* to improve the transparency of government budgetary and financial management, and to promote good governance (Tanner 2008). This included the replacement in 2009 of the previous government’s Outcomes and Outputs Framework with a new *Outcomes and Programs Reporting Framework*. It improved the specification of outcomes, to make them more specific and tangible. DoF played a leading advisory role in this by agreeing outcomes/programs structures with each department and agency. It was recognized that changes to outcome statements would introduce a break in time series of performance indicators, making comparisons with previous years even more difficult. Thus the 2009–10 portfolio budget statements provided a cross-walk between the information provided by the outcomes/outputs framework and that of the outcomes/programs framework. There is also better reporting in annual reports of targets and comparisons with actual performance, with effect from FY10. Targets are provided for the budget year and for the three out-years of the forward estimates period. DoF is to publish an omnibus report comparing portfolios’ performance with their targets. Program budgeting was also reintroduced. The ANAO is conducting a performance audit of the new Outcomes and Programs Reporting Framework.

4.4 Government Monitoring and Evaluation

Performance Reporting on Government Service Delivery

A strong feature of performance monitoring and reporting in Australia is the annual reports on government service delivery. These relate to sectors such as: hospitals and primary health care; schools and vocational education; police, court administration and prisons; and community services such as public housing and support for the elderly, disabled and children (SCRGSP 2010). These services involve total spending of \$136 billion, or 13.1 percent of GDP; most of these services are delivered by the state and territory governments. The reports contain about 1,600 pages and include about 930 performance indicators. The indicators relate to all parts of the results chain for each service, and especially: expenditure; staffing; outputs (including their efficiency, effectiveness and equity); and outcomes/impacts. At least as important as the performance information is the discussion of the reasons for performance differences among different jurisdictions. The discussion refers to contextual differences that help explain the differences in performance; these include issues such as population size, remoteness, and so on.

User surveys reveal the continuing usefulness of these reports (Productivity Commission 2007). About 80 percent of respondents are satisfied or very satisfied with them. The main uses of the reports are to brief management and ministers (69 percent), as a research tool (68 percent), to assess performance (58 percent) and to develop policy (43 percent). The reports are used by parliaments and state audit offices. However, the surveys also note that improved comparability of data, better data quality and more timely data are all areas that need improvement. The reports rely on data provided by each jurisdiction (i.e., the federal government, the six state governments and the two territory governments). While efforts are reportedly being made to improve the quality of these data (McLintock 2010), this appears to be the main weakness of the reports. For

example, the health Minister of New South Wales has repudiated his government's data on hospital performance, which indicated very poor performance on the part of some of his government's hospitals (Steketee 2008). A former auditor-general of New South Wales has stated that the governments of New South Wales and of Victoria have falsified their hospital performance data (Harris 2009).

There is also a national initiative, launched in early 2010 and supported by the federal, state and territory governments, that provides a report card on the performance of almost 10,000 primary and secondary schools in Australia. The My School Web site (www.myschool.edu.au) provides contextual information on each school, as well as information on the school's outcomes — student average test results for reading, writing, language (spelling, grammar, punctuation) and numeracy. It also provides Australia-wide averages for school outcomes, as well as for up to 60 comparable schools — these schools are identified using an index of socioeconomic characteristics of student households. This Web site was created to promote transparency and accountability for school performance. It has proved to be highly popular with parents and others, attracting very high Web traffic — with over 9 million hits on the first day. The federal government has been a strong promoter of the Web site, notwithstanding strong opposition to it (on philosophical grounds) from teacher unions. The federal government launched a similar Web site for hospitals in late 2010 (www.myhospitals.gov.au).

A Rejuvenated DoF Evaluation Strategy

The secretary of the department of finance recently briefed the government noting that¹⁰

'The quality of performance monitoring and evaluation information across agencies and policy/programs is variable. While there are some programs (particularly some large ones) that benefit directly from good evaluation practices, average quality appears to be low. Even where good quality evaluation and review exists, this information is not readily available to inform government decision making, especially on cross-portfolio matters' (DoF 2010, p. 3–5).

In other words, evaluation is being done for internal portfolio decision making and management, and it may also be done to provide support for ministers' new policy proposals (when the evaluation findings are favorable). But it is not being done to support the government's overall decision making nor for accountability purposes. The formal M&E requirements remain focused on monitoring, but there are some developments that suggest that considerably more attention could soon be paid to evaluation, with the department of finance once again taking a leading role.

The secretary of the department of finance advised the newly re-elected government in 2010 that *'processes for evaluation and review ... need to be reinvigorated'* (DoF 2010). He has noted the range of evaluations and reviews that are conducted by different entities within the government — these include DoF's own strategic reviews, DoF's ad hoc savings reviews, ANAO performance audits, reviews conducted by the Productivity Commission, evaluations conducted by departments and agencies, parliamentary enquiries, and so on (Tune 2010). However, these evaluative activities are ad hoc and piecemeal. Even DoF regards its own strategic reviews as being of variable quality and at times limited usefulness, due to the disinclination of some departments to become actively involved in them or to provide needed information. In addition, departmental secretaries tend to be disinclined to make evaluation findings available — either publicly or to DoF, and so DoF is unable to use these findings to inform its policy advice. The

¹⁰ This brief, which was initially confidential, was prepared for the Labor government re-elected in September 2010. The comments would be equally applicable to evaluation during the Howard government.

DoF secretary believes that this partly reflects concerns that DoF might use any adverse evaluation findings for savings options to cut programs.

In this environment, the DoF secretary has stated that *'Finance has retained its overall responsibility for the Government's monitoring, evaluation and review'* (Tune 2010), and he has advised his new Minister that *'a robust formal framework for evaluation and management of performance information would support fiscal strategies and improve delivery of services'* (DoF 2010, p. 3–5). DoF is still considering options for such a framework, and one model that is attractive to the DoF secretary is the Canadian approach; this involves the evaluation of all government programs on a four-year cycle in order to determine the efficiency, effectiveness and alignment with government priorities of all programs. He has indicated that the components of a rejuvenated DoF evaluation strategy are likely to include:

- Reforming evaluation and review practices by (i) positive incentives to departments to conduct robust, transparent evaluations (e.g., by sharing any subsequent program savings); (ii) some centralization of evaluation by means of DoF oversight of departmental and agency evaluations, in order to enhance evaluation quality; (iii) a governance framework encompassing both evaluation and performance information to provide transparency to government
- Enhancing the strategic review processes via, for example, a more comprehensive program of reviews, and by provision of greater authority for the reviews, such as by ensuring their terms of reference are explicitly endorsed by the Cabinet (DoF 2010).

The DoF secretary is keen to avoid what he perceives as the problems with Australia's evaluation strategy from 1987 to 1997: that the requirements to prepare detailed portfolio evaluation plans and to evaluate every program on a three- to five-year cycle were too cumbersome and resource intensive (for portfolios and for DoF policy analysts), and that the level of evaluation skills both inside and outside the public service was inadequate for the evaluation task. However, the challenges that he will face will include, among others:

- How to build demand for performance information and evaluation in support of more informed public debate and government decision making?
- How to ensure a sufficient supply of evaluation skills to enable the regular evaluation of government programs?
- How to change perceptions that evaluation is more of a fault-finding threat to line departments and to public servants, rather than an opportunity to improve program performance?
- How to persuade reluctant departmental secretaries to conduct more, and higher quality, evaluations, and to allow greater DoF involvement?
- How to ensure a strategic approach to departments' evaluations, so that the right programs are evaluated at the right time, that the right issues are evaluated, and that they are evaluated rigorously?
- How to ensure that DoF budget analysts — i.e., the DoF officers who advise the Cabinet on new policy proposals and savings options — become fully familiar with departments' evaluations, and that they view it as a core part of their work, rather than as a 'politically correct distraction'?

One final observation is that the challenge facing any greater emphasis on performance-based budgeting in Australia will require not simply more and better evaluation findings, but also a renovated policy decision-making process. This is discussed below.

5. LESSONS FROM THE AUSTRALIAN EXPERIENCE

The Australian evaluation system lasted for a decade — from 1987 to 1997. During this time, Australia became a model of evidence-based decision-making and performance-based budgeting. By the early to mid-1990s, the evaluation findings that the system was producing were having a significant influence on the contents of departments' policy advice and on the Cabinet's budget decisions. Ministers frequently stated the high value that they attached to having evaluation findings available to better inform their decision making. Evaluation findings were also being used quite intensively within line departments in support of their ongoing management.

Considerable effort was needed to reach this situation. It involved the progressive refinement of the evaluation strategy, which required some trial and error. This involved ongoing monitoring, review and adjustment to the evaluation strategy itself. More fundamentally, there were a number of success factors that were crucial to the success of the strategy; their importance was highlighted starkly by changes that occurred during the subsequent Howard era, from 1996 to 2007. Reform champions at the most senior levels of DoF, and reformist ministers in the Hawke/Keating government (from 1983 to 1996), were a key success factor. Related to this was the important role of the department of finance. As the central budget office, it was powerful and highly influential, and it was the architect and implementer of the evaluation strategy. Its advocacy and use of evaluations required a substantial evolution of its staff and their skill sets. Conversely, if DoF had remained a more traditional budget office concerned only with scrutiny of line item costs, or if it had taken a more passive approach to evaluation, then it might have become a serious roadblock standing in the way of performance-oriented government.

Once senior officials had secured the agreement of key ministers, it was relatively easy to create the evaluation system in 1987, and to progressively fine-tune it in the following years. This flexibility resulted from the Westminster system of government in Australia, which does not require a legislative basis for central requirements such as an evaluation system. A disadvantage — compared, for example, with countries that have a Napoleonic system of government — is that such decisions can be easily reversed when there is a change in government.

The evaluation strategy was far from perfect. Even after it had been in existence for a decade, over one-third of evaluations suffered from methodological weaknesses of one kind or another, thus reducing their reliability and value. Related to this, there were insufficient evaluation skills within the public service — most line departments relied on program areas to conduct the evaluations of their own programs, and a number of these apparently lacked the necessary skills and were thus ill-equipped to cope with this additional, unwanted workload. This perceived burden might also have helped to maximize the pushback from line departments against the evaluation 'red tape' when the opportunity arose with the change in government in 1996.

Some line departments avoided the problem of ill-equipped program areas by creating large, specialist evaluation units; it is interesting to note that it is these same departments that constituted the 'islands' of evaluation good practice that remained in existence after the evaluation strategy had been abolished. In these departments at least, an evaluation culture has persisted. With the benefit of hindsight, it might have been desirable for DoF to have mandated the creation of such units in each line department, to address the problems of evaluation quality.

These units could either have conducted all major, significant evaluations, or alternatively could have provided close support and quality assurance to their program areas. However, at the time it would have been considered too intrusive to mandate such a requirement.

When the evaluation strategy was created in 1987, there would have been a number of poor-performing programs, and subsequent evaluations could be expected to have revealed their poor efficiency, effectiveness or appropriateness; this would usually have led to the programs being improved (especially if they were an important government priority) or else cut or even abolished. But by the time that the strategy was abolished in 1997, most programs would have been evaluated more than once, and it might be expected that the marginal returns to M&E would have declined considerably. There is no information on the changing returns to M&E over time, so this possibility can not be investigated; in any event, this was never stated as a reason for the abolition of the strategy.

The abolition of the evaluation strategy in 1997 revealed a number of risk factors facing government systems for monitoring and evaluation. One factor is that a change in government can lead to fundamental change in public sector management. Another risk factor is the departure of key champions of M&E and of a centralized performance orientation within government. Related to this is their replacement by M&E ‘opponents’ — key officials who are skeptical of the benefits of M&E activities or who, in Australia’s case, pursued both a ‘let the managers manage’ philosophy that had not worked in the past and a simplistic performance framework that was probably doomed to failure. The performance framework that was developed during the Howard government era, and the incentives that were meant to underlie it, were ideologically driven; there is ample evidence to show that they constitute a failed experiment in public sector management. Of course, the evaluation strategy was only one casualty of this mindset. DoF had traditionally been the main guardian of fiscal rectitude, but its budget estimates role and its policy advising role were significantly degraded during this period, as indeed was the entire policy decision-making process of government.

The changed decision-making processes were related to another risk factor that Australia faced, which was, ironically, economic prosperity and large budget surpluses. Conversely, large budget deficits and a macroeconomic crisis can and had acted as very powerful motivators for governments to cut expenditure by identifying inefficient or ineffective spending while searching for much better value for money from existing spending; this kind of situation provides fertile ground for monitoring information and especially evaluation findings to show which policies and programs are cost-effective, and which are not. The large budget surpluses that Australia enjoyed from 1998-99 and onward undermined the priority for budget discipline and eliminated a main driver of a performance orientation by government.

The demise of the evaluation strategy appears to have been followed soon after by a substantial decline in the quantity and quality of evaluations conducted by departments and agencies. It was hoped by DoF that, after a decade of having a mandatory approach to the planning and conduct of evaluation, an evaluation culture would have grown up within the public service, and that this would have persisted over time. That this appears not to have happened may be due to a number of reasons. One is the sudden replacement of a number of departmental secretaries after the change in government in 1996. Another is the substantial cuts in the public service, and these would have made it a lot more difficult to continue to fund what are often viewed as being more ‘discretionary,’ long-term activities such as evaluation and research. Another reason may be that many departmental secretaries and their ministers are naturally disinclined to conduct evaluations — while positive evaluation findings that reveal good performance are always welcome, adverse findings can pose significant political and reputational risks. The current disinclination of

departments to publish their evaluations and their reported willingness to frustrate freedom-of-information legislation would appear to be consistent with this interpretation. That said, some departmental secretaries remain strong advocates of monitoring and evaluation; they apparently view M&E as important to help their internal management and policy development — their support of evaluation islands attests to this.

The evaluation strategy that existed from 1987 to 1997 can be judged to have been a fairly successful evaluation system. This is because it achieved a high level of utilization, with a generally reasonable quality of evaluations; however, it proved not to be a sustainable system because of significant and adverse risk factors. The system took considerable effort and time to build — as did DoF's budget estimates and policy advising expertise. But it took much less time to degrade these functions. Current efforts to renovate government evaluation can also be expected to require significant effort and to be time-consuming.

Australia's experience is also notable because of its reliance from 1997 to 2007 on a performance framework that was based on a system of performance indicators. An advantage of indicators is that they are cheaper, simpler and quicker than evaluations. But while they can be used to highlight examples of good or bad performance, a major limitation is that they fail to explain the reasons for this performance; thus it is difficult to apply the performance lessons elsewhere. Australia's performance framework during this period provides a strong example of how *not* to go about constructing a system of performance indicators. The framework encountered many conceptual and data difficulties. It also suffered from severe problems of implementation by departments and agencies, and from a lack of effective oversight by the DoF.

There are interesting choices that face any government that wishes to achieve a more evidence-based approach to its decision making — ironically, these same choices again face the Australian government and the department of finance. One choice is how centralist or devolutionary to be. The answer to this must surely lie in what are the intended purposes of monitoring and evaluation information. If the intended use is to aid the internal management and decision making of line departments and agencies, this would seem to favor a more devolutionary approach — unless it could be convincingly shown that these agencies would tend to significantly under-invest in M&E if they were left to themselves. The reality that most departments and agencies currently engage in little evaluation activity indicates that a devolutionary approach to evaluation is insufficient to support sound program management.

If regular M&E information on the performance of all government programs is needed, and on a systematic, whole-of-government basis, to aid government decision making — i.e., to assist in deciding new programs or which existing programs to cut — or for accountability purposes, then this argues for a form of centralist approach with leadership from the central budget office (i.e., DoF). Only a centralist approach could achieve sufficient coverage and quality of the M&E information that would be produced. Australia's experience in the years leading up to the creation of the evaluation strategy, and in the years after its abolition, demonstrates that a principles-based approach that encourages M&E, but does not mandate it and does not monitor and ensure compliance, is highly unlikely to work.

Of course, only some aspects of Australia's evaluation system from 1987 to 1997 were centralized: the evaluation policy itself, and the central oversight and evaluation participation roles of DoF. Evaluations themselves were planned and conducted by line departments, with DoF and other central agencies investing considerable effort in attempting to influence departments' evaluation agendas as well as the focus and conduct of individual evaluations. Evaluation was therefore a collaborative effort.

Thus DoF, as the system's architect, was trying to have the best of both worlds — to bring together DoF's (and other central agencies') objective scrutiny and willingness to ask tough evaluation questions, with the program expertise of line departments and agencies. This approach had the potential to meet the evidentiary needs of budget decision making, as well as to maximize the 'ownership' of evaluation findings by line departments and agencies. These needs were substantially met, albeit at the cost of some important reduction in evaluation quality. The quality reduction could have been avoided if DoF had mandated a requirement for each line department to create a large, specialist evaluation unit. Of course, this would have entailed a budgetary cost.

Some observers (such as DoF's current secretary) have argued that another disadvantage of the evaluation system was an onerous burden on DoF, and on other departments; this viewpoint is debatable, however, and it was certainly not identified by the ANAO in its 1997 performance audit of the evaluation strategy as a whole. Moreover, the partially centralized approach to evaluation from 1987 to 1997 was framed as the quid pro quo for a devolutionary approach to other public sector reforms — because most line departments and agencies preferred not to devote much effort to measuring their performance. This observation is equally relevant today.

The nature of a centralized or partially centralized M&E system need not consist simply of measures to force line departments to comply. Indeed, there are reasons why an authoritarian approach might even be counterproductive in some respects: it might discourage the cooperation of line departments, whose program expertise is invaluable in any evaluations of them. And it may well also reduce the 'ownership' by line departments of evaluations that are forced on them, and thus reduce their willingness to use the evaluation findings.¹¹ With the evaluation strategy, 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. The incentives that it adopted for evaluation comprised a mix of 'carrots, sticks and sermons,' rather than simply relying on a 'sticks' approach.¹² The carrots included the advisory support provided by DoF, as well as possible access to resource agreements for line departments to aid their management of programs revealed by evaluations to be under-performing. 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.¹³ The sermons included persistent advocacy by the secretary of DoF and his senior executives, as well as the explicit support for evaluation from some powerful ministers.

Of course, the real challenge currently facing the government is not simply how best to renovate monitoring and evaluation, and the broader performance framework. While important, these are essentially technocratic or supply-side issues. More significant is the challenge of renovating the entire policy advising and government decision-making process (of which the budget is the central part); this decision-making process constitutes a principal demand-side use of M&E information. There appear to be problems with many of the stages of this entire process: obtaining reliable monitoring information and evaluation evidence (this requires adequate data, skilled evaluators inside and outside the public service, and the commitment to conduct reliable M&E); making full use of this information by DoF, other central departments, line departments, external advisers and others in their policy advising roles (this also necessitates having a sufficient number

¹¹ This has happened with Chile's centralized M&E system, for example, which was designed by and is managed for the budget needs of the finance ministry (Rojas et al. 2005).

¹² A detailed taxonomy of carrots, sticks and sermons is provided by Mackay (2007, Chapter 11).

¹³ DoF never released these rankings publicly, but the DoF secretary did inform his fellow secretaries when they asked to know their department's relative performance.

of highly skilled policy analysts); having budget processes that allow and demand high-quality policy advice from all key stakeholders (especially ministers and their departments); and having a government that places a high value on having evidence available to inform its decisions.

Changes currently underway within the Australian government provide hope that some of these stages are being or will be renovated. These stages can be viewed as a ‘results chain’ — similar to the results chains that underlie each government program, linking its expenditure via its processes to its outputs, outcomes and impacts. The danger is that a persistent weakness at any point along the policy results chain could tend to frustrate the achievement of evidence-based decision making. Put another way, if a government pays little attention to M&E information, then the substantial effort needed to build up an M&E system is likely to be partly or largely wasted. That said, there are certainly other uses of M&E information — such as for program management, departmental decision making, and for accountability purposes — and these other uses might make an M&E system a worthwhile effort.

Australia is at an interesting juncture. It once had the best evaluation system in the world, but now it lags considerably behind developed and developing countries such as Canada, Chile, Colombia, Mexico and the United States.¹⁴ Only time will tell if current efforts to renovate monitoring and evaluation in the Australian government are successful.

¹⁴ Chile, Colombia and the United States are discussed by Mackay (2007). For an analysis of Canada, see Lahey (2010). Mexico is discussed by Castro et al. (2009).

BIBLIOGRAPHY

AGRAGA (Advisory Group on Reform of Australian Government Administration). 2010. *Ahead of the Game: Blueprint of Reform of Australian Government Administration*. Canberra: Department of the Prime Minister & Cabinet.

ANAO (Australian National Audit Office). 1991a. *Implementation of Program Evaluation -- Stage I*. Efficiency Audit Report No. 23, 1990-91. Canberra: ANAO.

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

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

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

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

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

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

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

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

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

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

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

———. 2007. *Application of the Outcomes and Outputs Framework*. Performance Audit Report No. 23, 2006-06. Canberra: ANAO.

———. 2010. *Conduct by Infrastructure Australia of the First National Infrastructure Audit and Development of the Infrastructure Priority List*. Canberra: ANAO.

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

ANAO/DoF (Department of Finance and Administration). 2004. *Better Practice Guide: Better Practice in Annual Performance Reporting*. Canberra : ANAO.

Banks, Gary. 2009a. *Challenges of Evidence-Based Policy-Making*. Canberra: Productivity Commission and Australian Public Service Commission.

- . 2009b. *Evidence-Based Policy-Making: What is it? How do we get it?* ANU Public Lecture Series, presented by ANZSOG, 4 February 2010. Canberra: Productivity Commission. Quoted by Alan Mitchell, 'Hard data paves way to reform.' *Australian Financial Review* 16 February 2009, p.20.
- 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.
- . 2003. 'Australian Public Sector Financial Management Reforms'. Presentation at a World Bank seminar, 25 June 2003.
- Blondal, Jon R., Daniel Bergvall, Ian Hawkesworth and Rex Deighton-Smith. 2008. 'Budgeting in Australia.' *OECD Journal of Budgeting* 8(2): 1-64.
- Borthwick, David, quoted by Verona Burgess. 2009. 'Everyone on the same page'. *Australian Financial Review* 13 March 2009.
- Campbell, Colin. 2001. 'Juggling inputs, outputs, and outcomes in the search for policy competence: recent experience in Australia'. *Governance: an International Journal of Policy and Administration* 14(2): 253-282.
- Castro, Manuel Fernando, Gladys Lopez-Acevedo, Gita Beker Busjeet and Ximena Fernandez Ordonez. 2009. *Mexico's M&E System: Scaling Up from the Sectoral to the National Level*. ECD Working Paper No. 20. Washington D.C.: Independent Evaluation Group, The World Bank.
- Chan, Matthew, Mark Nizette, Lisa La Rance, Charles Broughton and Derek Russell. 2002. 'Australia'. *OECD Journal of Budgeting* 1(4): 35-69.
- Conroy, Stephen. 2010. 'NBN cost-benefit analysis a waste of time and money: Stephen Conroy'. *The Australian* 5 November 2010.
- 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: Australian Government Publishing Service.
- DoF (Department of Finance). 1991. *Handbook of Cost-Benefit Analysis*. Canberra: DoF.
- . 1993. *The Cost of Evaluations: The Findings of a Pilot Study*. 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.
- DoF (Department of Finance and Administration). 2000. *The Outcomes & Outputs Framework: Guidance Document*. Canberra: DoF.
- . 2007. *2006-07 Annual Report*. Canberra: DoF.
- DoF (Department of Finance and Deregulation). 2010. *Incoming Government Brief*. Canberra: DoF.
- Ergas, Henry. 2010. 'Faulkner and Tanner fell short of their lofty aims'. *The Australian* 13 July 2010

- Ergas, Henry, and Alex Robson. 2009. 'Evaluating major infrastructure projects: how robust are our processes?' In Productivity Commission. *Strengthening Evidence Based Policy in the Australian Federation, Volume 1: Proceedings*: 127-167. Canberra: Productivity Commission.
- FaCS (Department of Family and Community Services). 2001. *Research and Evaluation Framework*. Canberra: FaCS.
- . 2002a. *Annual Report 2001-02*. Canberra: FaCS.
- . 2002b. *Portfolio Budget Statements 2002-03. Family and Community Services Portfolio*. Canberra: Commonwealth of Australia.
- . 2002c. *Research and Evaluation Digest*. Canberra: FaCS.
- Harris, Tony. 2009. 'Eradicate this dodgy data'. *Australian Financial Review* 14 April 2009.
- Hawke, Lewis. 2007. 'Performance budgeting in Australia.' *OECD Journal of Budgeting* 7(3): 133-147.
- Hawke, Lewis, and John Wanna. 'Australia after budget reform: a lapsed pioneer or decorative architect?' In John Wanna, Lotte Jensen and Jouke de Vries (eds.). *The Reality of Budget Reform in OECD Countries: Trajectories and Consequences*, pp. 65-90. Cheltenham: Edward Elgar Publishing.
- Howard, Cosmo. 2005. 'The policy cycle: a model of post-Machiavellian policy making?' *Australian Journal of Public Administration* (64)3: 3-13.
- JCPA (Joint Committee of Public Accounts and Audit). 2004. *Review of Auditor-General's Reports 2003-04: First and Second Quarters*. Canberra: Parliament of Australia.
- 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.
- Kelly, Paul. 2006. 'Re-thinking Australian governance: the Howard legacy'. *Australian Journal of Public Administration* 65(1): 7-24.
- Lahey, Robert. 2010. *Monitoring and Evaluation in the Government of Canada; 'Lessons Learned' from 30 Years of M&E Development*. ECD Working Paper forthcoming. Washington D.C.: Independent Evaluation Group, The World Bank.
- Mackay, Keith. 1998. *The Development of Australia's Evaluation System*. ECD Working Paper No.4. Washington D.C.: Operations Evaluation Department, The World Bank.
- . 2004. *Two Generations of Performance Evaluation and Management System in Australia*. ECD Working Paper No. 11. Washington D.C.: Operations Evaluation Department, The World Bank.
- . 2007. *How to Build M&E Systems to Support Better Government*. Washington D.C.: Independent Evaluation Group, The World Bank.
- McClintock, Paul. 2010. *Statistics, Public Accountability and the Federation*. Presentation to the NatStats 2010 Conference. Sydney: COAG Reform Council.
- McPhee, Ian, Auditor-General of Australia. 2005. *Outcomes and Outputs: Are We Managing Better as a Result?* Presentation to the CPA National Public Sector Convention, 20 May 2005. Canberra: ANAO.
- . 2008. *The ANAO's Contribution to Improving Public Sector Administration*. Presentation to the Canberra Evaluation Forum, 21 February 2008. Canberra: ANAO.

- . 2009. *The Business of Government: Why Public Sector Management Must Evolve*. Brookes Oration Presentation to the Deakin Business School, Deakin University. Canberra: ANAO.
- Mulgan, Richard. 2008. 'The accountability priorities of Australian parliamentarians'. *Australian Journal of Public Administration* 67(4): 457-469.
- Murray, Andrew, Senator. 2008. *Review of Operation Sunlight: Overhauling Budgetary Transparency*.
- OECD (Organisation for Economic Co-operation and Development). 2002. *Public Sector Modernisation: A New Agenda*. GOV/PUMA (2002), 2. Paris: OECD.
- 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.
- Podger, Andrew. 2005. 'Retirement speech'. 30 June 2005. (Referred to by Paul Kelly 2006.)
- . 2007. 'What really happens: departmental secretary appointments, contracts and performance pay in the Australian Public Service'. *Australian Journal of Public Administration* 66(2): 131-147.
- . 2009. *The Role of Departmental Secretaries: Personal Reflections on the Breadth of Responsibilities Today*. Canberra: Australian National University.
- Productivity Commission. 2007. *Productivity Commission Annual Report 2006-07*. Canberra: Productivity Commission.
- Rojas, Fernando, Keith Mackay, Yasuhiko Matsuda, Geoffrey Shepherd, Azul del Villar, Ariel Zaltsman, and Philipp Krause. 2005. *Chile: Study of Evaluation Program—Impact Evaluation and Evaluations of Government Programs*. Washington, DC: World Bank.
- 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.
- SCRCSSP (Steering Committee for the Review of Commonwealth/State Service Provision). 1995. *Report on Government Service Provision*. Melbourne: Industry Commission.
- SCRGSP (Steering Committee for the Review of Government Service Provision). 2010. *Report on Government Services 2010*. Canberra: Productivity Commission.
- SSCFPA. (Senate Standing Committee on Finance and Public Administration). 2007. *Transparency and Accountability of Commonwealth Public Funding and Expenditure*. Canberra: Parliament of Australia.
- Steketee, Mike. 2008. 'Hospital statistics 'worth nothing' '. *The Weekend Australian* 11-12 October 2008.
- Stutchbury, Michael. 2010. 'Imagine if Labor's broadband payoff is just hype.' *The Australian* 7 December 2010.

Symonds, Alexander. 2009. 'Reviews up for review'. *Australian Financial Review* 5 February 2009.

Tanner, Lindsay. 2008. *Operation Sunlight: Enhancing Budget Transparency*. Canberra: Australian Government.

Thomas, Paul G. 2009. 'Parliament scrutiny of government performance in Australia.' *Australian Journal of Public Administration* 68(4): 373-398.

Tune, David. 2010. *Evaluation: Renewed Strategic Emphasis*. Presentation to the Canberra Evaluation Forum. 19 August 2010. Canberra: DoF.

Wanna, John and Stephen Bartos. 2003. "Good practice: does it work in theory?' Australia's quest for better outcomes'. In John Wanna, Lotte Jensen and Jouke de Vries (eds.). *Controlling Public Expenditure: The Changing Roles of Central Budget Agencies*, pp. 1-29. Cheltenham: Edward Elgar Publishing.

Webb, Richard. 2010. *The Commonwealth Budget: Process and Presentation*. Parliament of Australia Research Paper. Canberra: Department of Parliamentary Services.

Wilkins, Roger. 2008. *Strategic Review of Australian Government Climate Change Programs*. Canberra: Department of Finance and Deregulation.

Other Titles in the ECD Working Paper Series

- #1: *Lessons from National Experience*, by Keith Mackay, 1998.
- #2: *Zimbabwe: Issues and Opportunities*, by Stephen Brushett, 1998.
- #3: *Indonesia's National Evaluation System*, by Alain Barberie, 1998.
- #4: *The Development of Australia's Evaluation System*, by Keith Mackay, 1998.
- #5: *Comparative Insights from Colombia, China and Indonesia*, by R. Pablo Guerrero O, 1999.
- #6: *Evaluation Capacity Development: A Diagnostic Guide and Action Framework*, by Keith Mackay, 1999.
- #7: *Sub-Saharan Africa: Lessons from Experience in Supporting Sound Governance*, by Mark Schacter, 2000.
- #8: *Strengthening Capacity for Monitoring and Evaluation in Uganda: A Results-Based Management Perspective*, by Arild Hauge, 2001.
- #9: *Guide to Conducting Reviews of Organizations Supplying M&E Training*, by Marie-Hélène Adrien, 2003.
- #10: *The Development of Monitoring and Evaluation Capacities to Improve Government Performance in Uganda*, by Arild Hauge, 2003.
- #11: *Two Generations of Performance Evaluation and Management System in Australia*, by Keith Mackay, 2004.
- #12: *An Assessment of the Impact of Bangalore Citizen Report Cards on the Performance of Public Agencies*, by Adikeshavalu Ravindra, 2004.
- #13: *Building Country Capacity for Monitoring and Evaluation in the Public Sector: Selected Lessons of International Experience*, by Salvatore Schiavo-Campo, 2005.
- #14: *Evaluation Capacity Development in the Republic of Ireland*, by Richard Boyle, 2005.
- #15: *Institutionalization of Monitoring and Evaluation Systems to Improve Public Sector Management*, by Keith Mackay, 2006.
- #16: *Experience with Institutionalizing Monitoring and Evaluation Systems in Five Latin American Countries: Argentina, Chile, Colombia, Costa Rica and Uruguay*, by Ariel Zaltsman, 2006.
- #17: *A Diagnosis of Colombia's National M&E System, SINERGIA*, by Keith Mackay, 2007.

- #18: ***Insider Insights: Building a Results-Based Management and Evaluation System in Colombia***, by Manuel Fernando Castro, revised 2009.
- #18: ***Insider Insights: Building a Results-Based Management and Evaluation System in Colombia***, by Manuel Fernando Castro, revised 2009.
- #19: ***Implementing a Subnational Results-Oriented Management and Budgeting System: Lessons from Medellín***, Colombia, by Rafael Gómez, Mauricio Olivera, Mario A. Velasco, 2009.
- #20: ***Mexico's M&E System: Scaling Up from the Sectoral to the National Level***, by Manuel Fernando Castro, Gladys Lopez-Acevedo, Gita Beker Busjeet, and Ximena Fernandez Ordonez, 2009.
- #21: ***Evaluation of Government Performance and Public Policies in Spain***, by Osvaldo Feinstein Eduardo Zapico-Goñi, 2010.
- #22: ***Implementing a Government-wide Monitoring and Evaluation System in South Africa***, by Ronette Engela and Tania Ajam, 2010.
- #23: ***The Canadian M&E System: Lessons Learned from 30 Years of Development***, by Robert Lahey, 2010.
- #24: ***Performance in Government: The Evolving System of Performance and Evaluation Measurement, Monitoring, and Management in the United Kingdom***, by Colin Talbot, 2010.

Other Recommended Reading

Annual Report on Evaluation Capacity Development. OED, 2002.

Conducting Quality Impact Evaluations under Budget, Time, and Data Constraints. IEG, 2006.

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

Evaluation Capacity Development: OED Self-Evaluation. OED, 2004.

Evaluation Capacity in Africa: Selected Proceedings from a Seminar in Abidjan. African Development Bank and World Bank, 1998.

Evaluation Capacity in Asia: Selected Proceedings from the International Conference, Beijing, 1999. United Nations Development Programme, Chinese National Center for Science and Technology Evaluation, and World Bank, 2000.

Glossary of Key Terms in Evaluation and Results Based Management. Development Assistance Committee, OECD, 2002.

How to Build M&E Systems to Support Better Government, K. MacKay, 2007.

Impact Evaluation — The Experience of the Independent Evaluation Group of the World Bank. IEG, 2006.

Influential Evaluations: Detailed Case Studies. OED, 2005.

Influential Evaluations: Evaluations that Improved Performance and Impacts of Development Programs. OED, 2004.

Monitoring and Evaluation: Some Tools, Methods and Approaches, 2nd Edition. OED. 2004.

OED: The First 30 Years. P. Grasso, S. Westy, and R. Weaving (eds.), OED, 2003.

Public Sector Performance—the Critical Role of Evaluation. K. Mackay (ed.), OED, 1998.

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

Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean: Proceedings of a World Bank / Inter-American Development Bank Conference. E. May, D. Shand, K. Mackay, F. Rojas, and J. Saavedra (eds.), 2006.

More publications can be downloaded from the ECD website at
<http://www.worldbank.org/ieg/ecl>.

