



PUBLIC GOVERNANCE AND TERRITORIAL DEVELOPMENT
PUBLIC MANAGEMENT COMMITTEE

Working Party of Senior Budget Officials

HAND-OUT

THE PERFORMANCE FRAMEWORK OF THE AUSTRALIAN GOVERNMENT, 1987 TO 2011

BY KEITH MACKAY

OECD Journal on BUDGETING, Volume 2011/3 -- FORTHCOMING

7th Annual Meeting on
PERFORMANCE & RESULTS

OECD Conference Centre, Paris
9-10 November 2011

(English Text Only)

**THE PERFORMANCE FRAMEWORK OF THE AUSTRALIAN GOVERNMENT,
1987 TO 2011
BY
KEITH MACKAY***

This article analyses the performance frameworks adopted by successive Australian governments since 1987. One highlight was a formal evaluation strategy managed by the Department of Finance to support high-quality policy advice and budget decision making. The replacement of this strategy by a decentralised system of performance indicators is also discussed, together with the quality of government decision making. Lessons for other countries are also highlighted.

JEL classification: H110, H500, H700

Keywords: performance indicators, audit, evaluation strategy, government monitoring and review, budgetary decision making, government service delivery, public sector management, role of finance ministry, accountability, outcomes and outputs, intergovernmental relations, Australia.

* Keith Mackay is a consultant to the World Bank, where he was a Lead Evaluation Officer in its Independent Evaluation Group until his retirement in 2008. Valuable comments on an earlier draft were received from Manuel Fernando Castro, Nidhi Khattri, Philipp Krause, Steve O'Loughlin and Ximena Fernandez Ordonez.

1. Context

Australia has a population of 22.3 million people in a land area of 7.7 million square kilometres. The country has enjoyed continuous economic growth over the past 19 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 of these are based on the Westminster model of government. And lastly, there are local governments. Each of these governments is elected separately. Federal spending in 2010-11 is projected to be AUD 355 billion (currently about USD 375 billion), or 25.2% of 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, 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 in June 2010. There are 20 government departments, and the three central co-ordinating departments are Finance and Deregulation (DoF), Prime Minister and 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

The Department of Finance and Deregulation (DoF) is essentially the central budget office; it co-ordinates the expenditure side of the budget. It oversees budget accounting, including the financial framework. The 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). In addition, the DoF prepares savings options for achieving budget savings; line departments may prepare their own savings options. The Department of the Prime Minister and Cabinet (PM&C) focuses on the governments' policy objectives and on whole-of-government policy issues. The 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 the DoF, perform a "challenge" function, in terms of analysing and questioning the new policy spending proposals of line ministers.

2. First period: 1987 to 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% in 1984-85 to 23% in 1989-90; by international standards, this is a very significant reduction. The government achieved this by reducing its own spending as well as the grants it paid to the other levels of government. At the same time, the government was committed

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

to implementing a substantial reorientation of public spending towards 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 market-place to a much more flexible, open economy. The Australian dollar was floated, high tariffs were reduced, the financial sector was deregulated, and the flexibility of the labour market was increased. Many government business enterprises were privatised 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. There was also a major reduction in the number of government departments through amalgamation, with the objectives of achieving less balkanised policy advice and of encouraging the internal reallocation of resources through portfolio budgeting.

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. The estimates 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; Blöndal *et al.*, 2008). The forward estimates also provided departments with greater surety about future resource availability. As Thomas has noted:

The [Australian] government could be considered a leader among countries in terms of the early adoption and the range of reforms made to the design of public organisations, the delivery of programs and services and the introduction of new mechanisms of internal and external accountability... (Thomas, 2009, p. 374).

The government advocated the principles of programme management and budgeting, with a focus on the efficiency and effectiveness of government programmes – through sound management practices, the collection of performance information, and the regular conduct of programme 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 programme effectiveness reviews and joint management reviews of programmes.

The departmental secretary of the DoF, Michael Keating, was a major architect of many of the government’s public sector reforms.² The role of the DoF as budget co-ordinator and overseer of the spending of other departments also helped to ensure its influence on the reform agenda; the DoF enjoyed the powerful support of the other central departments for this agenda. The 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. The 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. The department’s concern

² In 1991, Michael Keating was replaced as departmental secretary of the DoF by Steve Sedgwick, another strong supporter of public sector reform and of the formal evaluation strategy that had been proposed by the DoF. Mike Keating then became secretary of PM&C, the most senior public servant.

with budget spending encompassed not simply a priority on cutting government outlays, but also on finding ways to make spending more efficient and effective.

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

By 1988, it had become evident to the 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. The review found that:

- there was a lack of integration of evaluation into corporate and financial decision making;
- evaluations tended to focus on efficiency and process issues rather than on the more fundamental question of overall programme effectiveness – *i.e.* whether or not programmes were actually meeting their objectives;
- there was a poor level of evaluation skills and analytical capacity; and
- the role of central departments in evaluation, especially the DoF, was unclear.

2.2. The government's evaluation strategy³

The 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 line departments would have the primary responsibility for determining evaluation priorities, preparing evaluation plans and conducting evaluations.

The strategy had three main objectives. The first, and arguably the most important, was that it provide fundamental information about programme performance to aid the Cabinet's decision making and prioritisation, particularly in the annual budget process when a large number of competing proposals are advocated by individual ministers. Second, it encouraged programme managers within departments to use evaluation for the improvement of their programmes' performance. Lastly, the strategy aimed to strengthen accountability in a devolved environment by providing formal evidence of programme managers' oversight and management of programme 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 the Department of Finance.

The evaluation strategy to which the Cabinet agreed had four formal requirements for departments (Box 2).

³ This evaluation system is described in greater detail by Mackay, 1998.

Box 2. Evaluation strategy: formal requirements

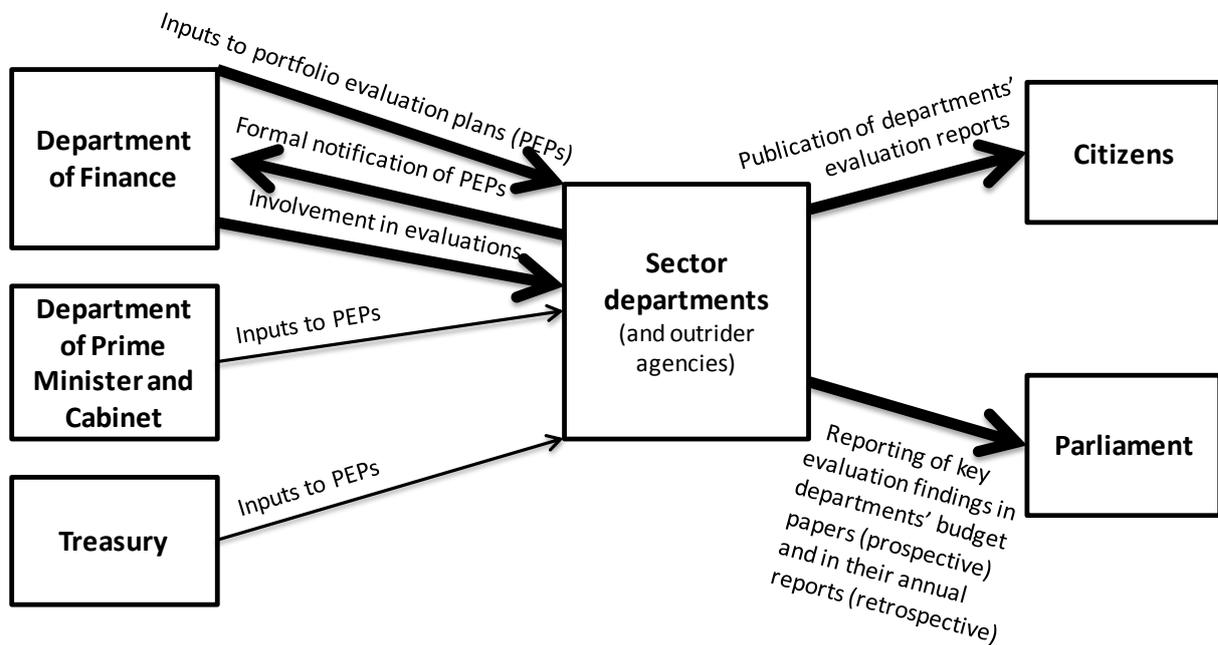
- That every programme be evaluated every 3-5 years.
- That each portfolio (*i.e.* comprising a line department plus outrider agencies) prepare an annual portfolio evaluation plan (PEP), with a three-year forward coverage, and submit it to the DoF. These plans were to comprise major programme evaluations with substantial resource or policy implications.
- That ministers' new policy proposals include a statement of proposed arrangements for future evaluation.
- 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 present in parliament each year should also report major evaluation findings.

The Cabinet also agreed that the DoF would have the opportunity to make an input to portfolio evaluation plans (PEPs) and to the terms of reference of individual evaluations to ensure their consistency with government-wide policies and priorities, and that the DoF would be available to participate directly in selected evaluations, subject to negotiation between the 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 in 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 Figure 1.

Line departments had expressed serious concerns with the planned role for the 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 be easily reversed when there is a change in government.

While the evaluation strategy had three stated objectives, from the perspective of the 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 DoF management 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 programme's objectives and the realities of its operating environment.

Figure 1. Evaluation planning and reporting flows



Achieving the necessary cultural change in the 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 programmes. 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 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 immediately after the Cabinet's agreement to the evaluation strategy, two government reports (from a parliamentary committee and from the Australian National Audit Office) noted the continuing unevenness in the coverage and scope of evaluation activity in line departments (Parliament of Australia, 1990; ANAO, 1991a). Both reports argued that the DoF should take a more active approach in encouraging departments to plan and undertake evaluations. The DoF then created a separate branch responsible for the provision of evaluation advice, support, training and encouragement to other departments and also within the DoF itself. This branch had nine evaluators capable of providing 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.

Australia's monitoring and evaluation system (M&E) essentially stressed evaluation, which was viewed as providing the necessary in-depth, reliable information on the efficiency and effectiveness of government programmes. 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. As discussed below, some of these audits focused on the evaluation activities of individual departments as well as on the government's overall evaluation strategy. The audit office's strong support for evaluation, which continues to the present day, has helped to highlight and provide additional legitimisation to evaluation. Another set of activities related to evaluation was undertaken 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). However, it had been hoped that evaluation findings would lead to the improvement of performance indicators and the setting of performance targets. By the mid-1990s, the DoF was concerned about departments' poor progress in articulating clear and achievable objectives for their programmes, and in collecting and reporting meaningful performance information on a regular basis. These concerns were confirmed by two reviews which the 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 programme objectives and the collection of regular performance information are often conceptually easier, and they were being done poorly. One explanation is that evaluation had been mandated, while the collection of performance information had not.

Thus, in 1995 the DoF secured the Cabinet's agreement to a rolling series of comprehensive reviews, staggered over three years, of the programme objectives and performance information of all programmes in all departments (see, for example, DoF, 1996). These reviews were conducted jointly by the DoF and each line department, with the results reported to their respective ministers and to the Cabinet as a whole. The reviews laid the basis for a much greater focus on performance information after 1997 (discussed in Section 3 below).

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 AUD 38 billion in annual expenditure, or about 9% 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 below.

2.4. How successful was the evaluation strategy?

2.4.1. Evaluation planning

Since 1987-88, all government departments had prepared annual portfolio evaluation plans (PEPs), and these were intended to comprise the major evaluations of the department and its outrider

agencies. By the mid-1990s, about 160 of these evaluations were under way at any given time. The bulk of these evaluations were major, in that the programmes had significant policy or spending implications; however, a significant minority of these evaluations, particularly for the smaller departments, concerned only minor programmes or only the efficiency aspects of large programmes. Line departments themselves decided which programmes should be included in their PEPs and also which issues the evaluation terms of reference would cover. However, the DoF would usually endeavour to influence the departments' choice of evaluation priorities by making direct suggestions to them. In making these suggestions, the DoF would attempt both to anticipate and to help create the information needs of the Cabinet. Where the 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 Cabinet-endorsed, formal requirement under the evaluation strategy that portfolio evaluation plans be prepared and submitted to the DoF certainly provided a powerful incentive to line departments to prepare plans and to take them seriously. Another influential factor was the issuance by the DoF of formal guidelines to departments concerning the desirable content of these plans, together with follow-up monitoring and reminders to departments about the need for the plans. The evaluation branch of the DoF conducted internal reviews of the content and coverage of these evaluation plans, and provided feedback and prompting to departments as well as by identifying examples of good practice. The DoF secretary also used this information to informally pressure line departments to improve their evaluation activities. The DoF supported the creation of the Canberra Evaluation Forum which involves monthly meetings of the evaluation community to discuss topical evaluation issues. The meetings were organised by a steering group of departments and other interested parties; this forum still exists, and it attracts large audiences at its meetings.

In a number of performance audits and two "better practice" guides on programme evaluation and performance information, the Australian 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/DoF, 1996).

2.4.2. Conduct of evaluation

The formal requirement that all programmes be evaluated every 3-5 years was influential in creating a climate of expectation that evaluation is the norm rather than the exception. The concept of regular, comprehensive coverage of programmes also encouraged a planned, staged approach to evaluation. This formal requirement should not be accepted at face value, however. It is very seldom the case that all aspects of a programme are included in any single evaluation. Instead, it is usual that an evaluation will focus only on particular aspects of a programme or one of its sub-programmes. The challenge is to ensure that the evaluation addresses the main objectives or problem issues, and this is a role in which the DoF played an active part via persuasion concerning evaluation terms of reference and via direct involvement in individual evaluations.⁴ The 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 programme being evaluated, the maturity of the programme, and the funds available to conduct the evaluation.

⁴ There was only modest success with the requirement that ministers' new policy proposals include an evaluation plan of action that would be undertaken if the proposal was accepted. Feedback from portfolios indicated that this requirement was onerous for portfolio managers during the busy budget period. For example, only about 30% of proposals broadly met this requirement in the 1993-94 budget, although an additional 50% of proposals included a clear undertaking to evaluate the proposal if accepted (DoF, 1994b). These percentages were only achieved after considerable prodding by line areas within the DoF.

The evaluations were managed and conducted (or contracted out) by line departments. The rigour, depth and types of evaluation conducted varied considerably, as did their cost. At one end of the spectrum, the evaluations comprised rapid reviews of programme performance, using whatever evidence was available and investigating specific issues such as programme 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, it is probably fair to say that the majority involved relatively rapid review. Of course, evaluations of major programmes would often be major undertakings involving a range of evaluation tools, methods and approaches. A sample of evaluations analysed by DoF ranged in cost (in 1993 prices) from AUD 56 000 to AUD 560 000 (DoF, 1993).⁵

Most departments chose to set up evaluation units to co-ordinate their formal evaluation planning. At their smallest, these units comprised two or three individuals who provided some advice and 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 labour market programmes.

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

A National Audit Office survey found that, for evaluations conducted over the period 1995-97, about half examined the delivery of products or services to external clients, and a further 30% were associated with matters internal to the department. One-third of the evaluations examined the appropriateness of new or established programmes, and 15% were directed towards the development of policy advice for the government (ANAO, 1997).

The large number of evaluations under way 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. The result was a large and rapidly growing “library” of evaluation findings. The DoF also published a register of published evaluation reports, and this also provided

⁵ The annualised cost of these evaluations was equivalent, on average, to less than one per cent of the government’s spending on these evaluated programmes. Thus, if the evaluation findings had led to only a very modest improvement in the efficiency or the effectiveness of the programmes, the evaluations would have been cost effective.

some quality assurance because the public availability of these reports exposed them to peer scrutiny. The ANAO survey found that 75% of evaluations conducted in 1995 and 1996 were either released to the public or were available on request.

2.4.3. Evaluation quality

The quality of evaluation reports is a much more difficult dimension to measure. The rigour of programme 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 weaknesses 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 programme.

The perspective of the DoF was that the quality of evaluations can be expected to vary enormously. This variation 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, the 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 programme managers and their staff for the actual conduct of evaluations. This devolutionary approach helped ensure that the evaluations drew on the programme expertise of staff and that there was a high level of “ownership” of the evaluation findings; both of these elements may be difficult to achieve with external evaluations. The DoF 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 rigour were achieved.

Some disadvantages of this devolved approach were a lack of evaluation skills in many programme areas and a 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 the DoF in particular – and the DoF and departments also prepared guidance material such as evaluation handbooks (for example, DoF, 1991, 1994a, 1996). There is also a fairly large community of evaluation consultants in Canberra, including numerous academics with either subject area knowledge (*e.g.* health issues) or with specialist research and analysis skills. Nevertheless, the 1997 ANAO study also revealed that 20% of departments were concerned about the lack of available training in advanced evaluation techniques, and this 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 that technical advisory panels were available to help assess the work of consultants; having steering groups available to help manage

consultants; and ensuring that 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.

2.4.4. 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 ministerial or Cabinet decision making. Banks 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' judgements. 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. (Banks, 2009a, p. 3)⁶

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. The 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). The survey respondents were 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 the 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 AUD 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 1994-95 (the last year for which estimates were available), the amount involved had risen to AUD 2.3 billion.⁷ Measured in dollar terms, the proportion of new policy proposals influenced by evaluation rose from 23% to 77% over that period; and for most of these proposals, 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 ministers 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 programmes or issues.

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

⁷ In that year, the government's total outlays were AUD 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 scrutinised comprised new policy proposals plus savings options.

It is also the case that evaluation can have a significant influence on the “savings options” put forward by the 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 AUD 500 million of savings options (or 65% of the total) were 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 “market-place 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% of the AUD 3.74 billion of proposals considered (new policy proposals plus savings options).⁸ The corresponding proportion for the 1993-94 budget, however, was only 19% of proposals. One important reason for this difference was the substantial revision of labour market, industry, regional and aboriginal policies in the 1994-95 budget; the major policy review on which these decisions were based had been heavily influenced by a number of evaluations commissioned specifically to help guide the policy review (DoF, 1994b).

Campbell has observed that the DoF “functioned as the nerve center for Cabinet on how programs actually functioned. It spearheaded public service reform throughout the 1980s. Its signature initiatives were devolution of authority over running costs to departments and the promotion of an evaluation culture that allowed ministers to monitor and gauge the effectiveness with which departments employed such discretionary spending power” (Campbell, 2001, p. 275).

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). And in comparing the relative success of Australia’s and Canada’s approaches to incorporating evaluation into the budget process, Schick has stated that:

Canada organised a vast [evaluation] effort around the Comptroller General in the 1970s; Australia adopted an ambitious evaluation strategy in the late 1980s. Canada’s effort bore little fruit; Australia’s produced significant reallocation of budget resources. Canada is thought to have failed because it centralised evaluation, thereby dampening co-operation by spending departments which may have been adversely affected by the findings. Australia is thought to have succeeded because it gave affected departments a big stake in designing and using evaluations. (Schick, 2002, pp. 23-24)

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 utilisation 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

⁸ Moreover, where policy proposals were supported by evaluation, the evaluation was judged to have also influenced the Cabinet’s decision in 89% of cases.

improvements in operational efficiency, and to a lesser extent with respect to resource allocation decisions and the design of service quality improvements for the benefit of clients. This high level of utilisation reflected a strength of the Australian evaluation system: evaluation was essentially a collaborative effort involving the 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 to 2007

3.1. Context

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 emphasised 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%.

3.2. Changes in public sector management

Over the eleven 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 analysed in this article. This framework embodied a mix of principles, expectations and formal requirements. Overall, the number and nature of formal requirements were simplified considerably from those of the previous government (Russell, 2003).

3.2.1. 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 and, by 2001, had put all departmental secretaries on three-year contracts.

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 CEO actions by excessive administrative controls. Performance pay was introduced for departmental secretaries in 1999; the performance contract was between them and their minister.⁹

⁹ Performance pay had been introduced for the senior executive service and less senior public servants in the late 1980s.

Table 1. Australia's performance framework: key aspects

| | 1987-96 | 1996-2007 | 2007-11 |
|--|---|--|---|
| Australian Public Service (APS) | Cohesive public service; central rules and standards, e.g. pay, classifications, terms of employment. | Public service downsized and balkanised; individual employment contracts; heavy use of business consultants; departmental secretaries usually employed on three-year contracts. | Efforts to renovate public service, e.g. for policy skills; moves to recentralise some functions, e.g. procurement, pay grades. |
| Philosophy underlying public sector management | Substantial devolution to departments; central requirements, e.g. evaluation, to "make the managers manage". | Very high level of devolution ("let the managers manage"); reduction in red tape; much greater reliance on private sector. | Some recentralisation, with heavy emphasis on encouragement; "let the managers manage"; further reduction in red tape. |
| Policy cycle | Formalised, disciplined; heavy reliance on analysis by public service; Expenditure Review Committee (ERC) at centre of budget process. | Much less disciplined; greater reliance on non-APS policy advice; many policy/expenditure decisions taken in Prime Minister's Office; ERC relatively weak. | Decision making initially in hands of four key ministers; now greater reliance on budget/ERC processes; APS policy skills to be strengthened. |
| Role of the Department of Finance (DoF) | Powerful, respected, high level of policy skills; heavily involved in scrutinising new policy proposals – the "challenge" function; responsible for budget estimates; heavily involved in evaluation. | 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 the DoF (from 2006). | 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 | Formal strategy and requirements (from 1987); enforcement by the DoF; heavy utilisation in policy advice and by the ERC; evaluation use by line departments. | Evaluation deregulated; only a few remaining evaluation islands among departments/agencies; small number of strategic reviews (from 2006); no systematic use of evaluation in the budget process. | 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), programme objectives, accountability | Programme budgeting (from 1986 on); evaluations usually published; only late attention to performance indicators via reviews of PI, programme objectives (from 1994); federal/state reporting of service delivery performance (from 1995); formal reporting requirements (annual reports, portfolio budget statements). | Programme budgeting abolished (from 1999); new "Outcomes and Outputs Framework" for formal reporting, based on performance indicators (1999); principles-based, no quality control by the DoF; accrual accounting (1999); evaluations rarely published; federal/state reporting of service delivery performance. | "Outcomes and Programmes Reporting Framework", based on performance indicators, and now including programme budgeting; evaluations rarely published; federal/state reporting of service delivery performance; citizen surveys planned. |

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.¹⁰ Accrual budgeting was introduced in 1999, and the government viewed it as “heralding a new ‘businesslike’ era in government ... [because it] would show the full cost of all programmes, not just the immediate cash outlays, and therefore make it easier to price and compare them with alternative private sector provision” (Blöndal *et al.*, 2008, p. 152).

To facilitate this *modus operandi*, departmental secretaries were given considerable autonomy over human resource decisions, including the pay of public servants in their departments; this led to a less centralised, less unified, and more balkanised public service. The government also chose to rely much more on alternative sources of policy advice, particularly business consultants as distinct from academics. These consulting companies have not only provided policy inputs but have even been asked to prepare detailed policies (Banks, 2009a). Thus, by the final year of the Howard government in 2007, the total expenditure on consultants was AUD 484 million, or about the same amount as the entire senior executive service in government which comprised some 2 700 staff.

These changes reflected what was, essentially, a “let the managers manage” philosophy. It was analogous to the one adopted in the early 1980s.

3.2.2. *The policy cycle and the role of the DoF*

The election of the Howard government in 1996 led to an environment in which there was greater contestability in policy advice and policy debates. The government diversified its sources of policy advice and policy preparation, including business consultants, think-tanks and academics. Indeed, the government displayed an ideological preference for advice from outside the public sector. 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 policy process. One respected commentator argued that:

Howard, like most Prime Ministers, wants research and advice to realise his aims. He would not be swayed, for example, by research suggesting that [his favourite policy options were] ineffective. In such a highly political environment the risk is that policy is more divorced from evidence-based research. (Kelly, 2006, pp. 14-15)¹¹

For almost all of this period, Australia enjoyed considerable budget surpluses, and this was substantially due to a continuing strong economy. The need for budget discipline was thus reduced significantly and, as a result, the government often decided a large proportion of budget spending right at the end of the budget process (Blöndal *et al.*, 2008). These decisions were made outside of the formal budget (*i.e.* policy) process, and “controlling expenditure has become ever more difficult” (Blöndal *et al.*, 2008, p. 161). The lack of budget discipline meant that monitoring information and

¹⁰ This particular New Zealand reform is often viewed as unsuccessful in that country; see, for example, Schick, 1996, 1998; Campbell, 2001. The notion of a minister purchasing the outputs of a government department implies an imaginary market-place for what is essentially, in most cases, a bilateral monopoly situation. The absence of substantive incentives to take seriously this kind of notional purchaser-provider relationship undermines any value in this approach. The Australian approach went beyond that of New Zealand in that ministers (and the parliament via its control over appropriations) were expected to purchase outcomes rather than outputs. Given the often fuzzy links between outputs and outcomes, this Australian version is even more fanciful.

¹¹ This perspective is confirmed by a number of interviews of senior public servants (Howard, 2005).

evaluation findings about the performance of government programmes became essentially irrelevant. It can thus be argued that the large budget surpluses from 1998-99 onwards eliminated a main driver of a performance orientation by government.

The DoF was an early casualty in this changed environment, and it went through a “traumatic” period in the late 1990s and early 2000s (Blöndal *et al.*, 2008). The government had appointed a new departmental secretary, Peter Boxall, to head the 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 more fully on accrual budgeting and the contracting out of government activities. Under its new leadership, the DoF wanted departments and agencies to take the main responsibility for budget estimates, including their forward estimates, as well as for reporting their performance under 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. (Blöndal *et al.*, 2008, p. 143)

Campbell (2001) has provided evidence that the weakening of the policy formulation process started during the early part of this period as a result of the irregularity of Cabinet meetings and poor preparation for them, and that the DoF no longer served as a “nerve centre” for the Cabinet concerning the performance of government programmes. This evidence is based on a number of interviews of senior executives that Campbell conducted in the latter part of the 1990s.¹² Based on this evidence, he concluded that the policy process had been undermined due to “... the current government’s tendencies to operate less collegially and collectively than did the Labor government, to trust less and not include in decisions top officials, and to introduce a [DoF] budgeting technology which appears to focus on outputs – often even inputs – at the expense of attention to outcomes” (Campbell, 2001, pp. 278-279).

This situation led the normally diplomatic OECD to conclude that:

In Australia, the deregulation of the public service and the adoption of an arm’s-length posture by the central agencies allowed management freedom but is currently considered to have deprived the Finance Ministry of the information necessary for it to adequately advise the minister. (OECD, 2002, p. 5)

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

Peter Boxall was replaced as DoF departmental secretary by Ian Watt 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 the DoF who had the financial skills necessary to monitor agencies’ financial performance, cash

¹² This was also confirmed by a departmental secretary upon his retirement in 2009 (Borthwick, quoted by Burgess, 2009). He stated that “during my time [in the Australian public service], some of the most important decisions were made by the Prime Minister. ...Cabinet was little more than a rubber stamp.” He advised that “departments should more rigorously analyse and evaluate different choices”.

flows, financial management compliance, and ability to estimate expenditures (Hawke, 2007). These functions are normally the most basic ones 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 towards 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,¹³ the central budget office had shifted its focus from a substantial emphasis on government results and performance to a focus instead on financial inputs. This shift represents a much more old-fashioned conceptualisation of the role of a department of finance.

Policy capacities in the wider public service were also significantly degraded during this period. This led the Public Service Commissioner in 2005 to state her concern that: “For some time now I have been concerned that there has been an erosion of the capacity for sound research, evaluation and analysis in some areas of the [public] service” (Briggs, 2005, p. 10). She also noted that a number of senior public servants shared these concerns.

Similarly, the head of the Productivity Commission (the government’s principal adviser on microeconomic reform, including matters relating to Australia’s productivity and living standards) has repeatedly noted that, from 1996 until the present, continued funding cuts have resulted in a large decline in the number of public servants who are skilled in quantitative methods and other analytical tools, which he identified as exactly the types of skill needed for evidence-based policy making (Banks, 2009a).

3.2.3. 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 stated 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, the DoF. Line departments 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 (PEPs). 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 the 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 programme 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 four formal requirements:

¹³ Advocates of this period have argued that one of its strengths was the explicit focus on results – *i.e.* on outputs and outcomes (and impacts). Yet the evaluation strategy and other public sector reforms in the earlier period stressed the achievement of efficiency and effectiveness. Efficiency is the achievement of results at least cost, while effectiveness is the achievement of programme objectives – which are defined as the outcomes that the government wishes to achieve. Thus the focus in the two periods was not so different, with the exception that in the later period there seemed to be much less explicit focus on the cost of government programmes.

- That every programme be evaluated every 3-5 years.
- That portfolios prepare annual portfolio evaluation plans.
- That new policy proposals indicate how they would be evaluated.
- That evaluation reports should normally be published.

The current DoF departmental secretary has recently argued that the decision to abolish the evaluation strategy reflected a view that it was too cumbersome and 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 – had provided to line departments was essentially discontinued. The DoF still encouraged line departments to conduct evaluations of key policies and programmes, 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 presented in parliament, were meant to report the implications of any evaluations that have been conducted in the past year; however, a report by the Australian National Audit Office (ANAO, 2003) concluded that departments’ annual reports often do not report evaluations or their findings.¹⁴ Portfolio budget statements (PBSs) were also meant to indicate any planned evaluations in the coming year. In addition, new policy proposals were required to report evaluation evidence, if any 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; this terminology is indicative of the business-oriented mindset of the Howard government.

There was a new requirement to review “terminating programmes” (those with a defined end date) and “lapsing programmes” (those that were expected to continue but where funding would have to be reauthorised) and to report the findings to the Expenditure Review Committee of Cabinet (McPhee, 2008).¹⁵ These reviews were focused on asking if and how the government should continue to be involved in a programme. However, a team of OECD reviewers has 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” (Blöndal *et al.*, 2008, p. 151). This requirement has now been abolished.

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

A former Public Service Commissioner and departmental secretary has summarised the situation thus:

¹⁴ It is not known if such omitted findings reflected adversely on the performance of the departments.

¹⁵ This requirement excludes most transfer payments.

One of the strengths of the Hawke/Keating governments was the more systematic gathering of performance information and evaluation of programs. Cabinet submissions were required to identify research and evaluation evidence in support of policy proposals and to set out how the proposed new policy would be evaluated. In a short-sighted attempt to reduce the size of submissions, the requirement was dropped in the late 1990s, and evaluation was no longer a mandated requirement. While performance reporting continues to contribute to “evidence-based” policy advising, loss of internal capacity and systematic evaluation, and the increasing reliance on chosen external consultants, has affected the capacity of departments to offer high-quality policy advice. (Podger, 2009, pp. 79-80)

Towards the end of the Howard government era (1996-2007), another DoF departmental secretary, Ian Watt, 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 the 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 Australian government’s 62 climate change programmes (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 4-6 staff seconded from various departments. Review teams are located in the DoF where there is a secretariat to manage the reviews. The current secretary of the DoF considers that some of the reviews have been of excellent quality, although other reviews have been hampered by lack of agency involvement and co-operation (DoF, 2010).

3.2.4. *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 3). This framework was intended to further strengthen the shift in focus from inputs to results. It required departments to agree with their ministers on the outcomes – the government objectives – towards which they were working, and to also agree the departmental outputs – such as service delivery for specific target groups or for the community at large – that would be produced to help achieve these outcomes (DoF, 2000; Chan *et al.*, 2002; Russell, 2003). By 2004, for example, the 145 departments and agencies collectively had 199 outcomes, and usually between one and ten outputs each¹⁶ (McPhee, 2005; Blöndal *et al.*, 2008). CEOs were given the authority to allocate appropriations across different types of output to achieve the desired outcomes. The DoF provided guidance regarding performance management and performance reporting. However, the 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 Figure 2.

¹⁶ The ANAO (2007) found that one agency had 52 outputs.

Box 3. The Outcomes and Outputs Framework: objectives and potential benefits

The Outcomes and Outputs Framework had three principal objectives; these mirror the objectives of the earlier evaluation strategy:

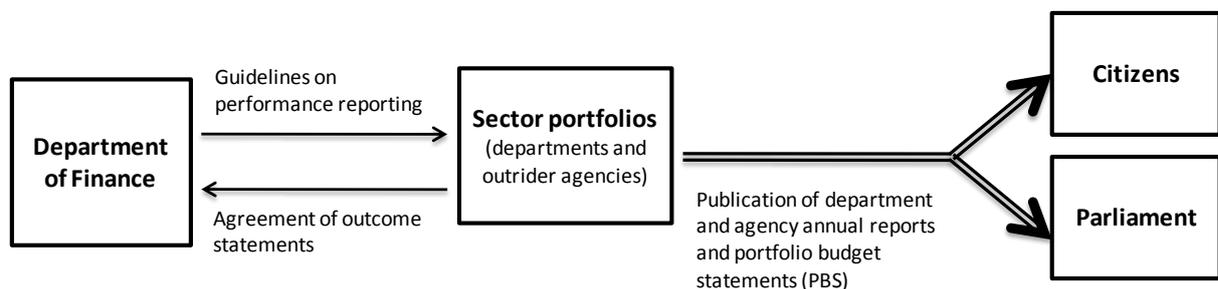
- To support the government's policy development.
- To support and strengthen departments' internal management, including staff learning.
- To strengthen external reporting, for accountability purposes.

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

- Clarify and ensure agreement among key stakeholders concerning objectives and expected standards of performance. This shared understanding is also crucial to ensure that purchaser-provider arrangements work well.
- Encourage departments to understand well the results chains underlying their work activities – *i.e.* the logical (and desirably evidence-based) links between departmental spending, their activities, their outputs, and their contribution to desired outcomes.
- Facilitate benchmarking comparisons across departments, and between the public and private sectors. This in turn would make it easier for the government to select least-cost providers, thus yielding efficiency gains. Accrual budgeting could be expected to facilitate considerably such an approach.
- Help departments to cope better with management complexity, particularly when the work of several departments jointly impinges on achievement of a particular outcome objective. This relates to the concept of "joined-up government".*

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

Figure 2. The 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. Some examples of outcome statements and performance indicators under the Outcomes and Outputs Framework are shown in Table 2.

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 programme administration as measured by: <i>i</i>) number (and trend) of funded grants and grants administered; <i>ii</i>) per cent and number of applications processed within agreed time frame; <i>iii</i>) per cent 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 care services outside school hours 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 the 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) finalised. • Performance against service standards (median processing times). |
| Transport and Regional Services | 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. |

Source: Portfolio budget statements (2006-07 budget) for each portfolio.

Ex post reporting is achieved via departmental annual reports that discuss actual spending and actual performance. *Ex ante* reporting is achieved via portfolio budget statements that relate planned spending to promised performance. Both types of report are presented in parliament. The annual appropriations provided by the parliament are intended to allow the government to achieve its specified outcomes; indeed, the stated outcome objectives constitute the legal basis for appropriations. Each department and agency is able to specify its own outcomes and outputs, subject to the endorsement of its minister. With the introduction of this new framework in 1999, programme budgeting – which had been introduced in 1986 – was abolished.

3.3. The 1996-2007 performance framework: extent of success

3.3.1. The Outcomes and Outputs Framework

The core part of the government's performance framework was the formal Outcomes and Outputs Framework. This framework relied on the collection and reporting of performance information in order to aid policy development and departments' internal management, and for 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 programme and agency performance, it requires careful analysis for this to be achieved. Simple reporting of data is insufficient; it is typically necessary to supplement the data with qualitative information and especially sound analysis and interpretation. 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 above. Fortunately, these published comparisons continued uninterrupted during the Howard era.

Of course, while such analyses 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 programme'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.

Specific 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. Similar problems were highlighted frequently in subsequent years but were never satisfactorily addressed during the Howard government (1996-2007). The role of the DoF in this situation is briefly discussed below.

The many problems with the framework have been identified by ANAO performance audits, speeches of the Auditor-General, reports of parliamentary enquiries, the OECD, and a range of academic observers and senior public servants; see, for example, ANAO, 2001a, 2003, 2007; Blöndal *et al.*, 2008; McPhee, 2005; Mulgan, 2008; Murray, 2008; Parliament of Australia, 2004, 2007; Podger, 2009; Thomas, 2009; and Webb, 2010. The problems that have been identified with the Outcomes and Outputs Framework are serious and fundamental (Box 4).¹⁷

¹⁷ Notwithstanding this long list of serious deficiencies, some agencies have been identified by the Auditor-General as providing good-practice reporting. They include the Department of Family and Community Services, the National Library of Australia, and the Australian Customs Service (McPhee, 2005).

Box 4. The 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; Parliament of Australia, 2007).
 - One media commentator noted that “in the past, the ‘outcomes’ have been defined as vaguely being in line with the minister’s expectations, rather than clearly defining the target group of the spending, what activities would be undertaken with the funding, and the intended result” (Tingle, 2009, p. 17).
 - Andrew Podger, a former Public Service Commissioner and departmental secretary, stated that “it is easy to be sucked into unrealistic outcomes approaches with meaningless outcomes statements and limited connections between the program activities and the claimed outcomes” (Podger, 2009, p. 81).
- There were about 200 outcomes for the government – and, for example, only three for the entire defence department – and so they had a high level of aggregation. In addition, 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; Parliament of Australia, 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 (Blöndal *et al.*, 2008).
- Definitions continued to change over time, even a decade after the framework was introduced (Blöndal *et al.*, 2008; Webb, 2010).
- Portfolio budget statements reported information on 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, which has responsibility for approving the government’s proposed budget (see Section 3.3.2), found the outcomes structure of the budget confusing, and strongly preferred a programmatic approach to performance reporting and budgeting (Parliament of Australia, 2007, quoted by Mulgan, 2008).

Given the severity of these problems, it is not surprising that the framework was often not integrated with agencies’ internal performance management processes (ANAO, 2007; Murray, 2008). Blöndal *et al.* (2008) argued that the outcomes focus was not integrated into budget decision making nor in agencies’ daily operations. Instead, the focus of departments and agencies, of Cabinet, and of

parliamentarians, remained on programmes.¹⁸ 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 Auditor-General noted the dissatisfaction of parliamentary committees with the performance information provided by departments (Barrett, 2003). Particular difficulties included definitional changes from year to year and different definitions adopted by different departments; these made it difficult to analyse departmental performance over time, or to make efficiency comparisons between departments. These difficulties reinforce the conclusion that the government failed to achieve one of its reform objectives, *i.e.* greater transparency via benchmarking comparisons of unit costs for all departments and agencies. It can be argued that one advantage of a systemic and centralised approach to performance measurement is that it enables exactly such comparisons to be made. A devolved approach, in contrast, is inherently unlikely to achieve this.

A subsequent Labor government finance minister concluded in 2008 that:

Some outcomes are so broad and general as to be virtually meaningless for budget accounting purposes leading taxpayers to only guess what billions of their dollars are being spent on...

The Outcomes and Outputs Framework was intended to shift the focus of financial reporting from inputs... to outputs and outcomes, *i.e.* actual results. While this is worthy in theory, it has not worked. Basic information on inputs was lost in the changeover, and reporting of outcomes is seriously inadequate. (Tanner, 2008, p. 4)

The extent to which the various objectives of the Outcomes and Outputs Framework were achieved is summarised in Table 3.

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.¹⁹

¹⁸ Programmes 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 programmes have been defined in a much more detailed manner. Programmes in the Australian government are also organisation-based.

¹⁹ A 2003 ANAO performance audit on annual performance reporting concluded that agencies did not possess the necessary data standards and procedures to assure data quality and coherence (ANAO, 2003). In contrast, the 2007 ANAO performance audit on the Outcomes and Outputs Framework relied heavily on self-reporting by agencies in its conclusion that agencies have sound systems for capturing and reporting performance information, and that they provide accurate and reliable information (ANAO, 2007). This later performance audit conclusion was contradicted by the Murray inquiry that was commissioned by the subsequent Labor government (Murray, 2008).

Table 3. The 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 |

However, there do appear to have been at least two fundamental problems with the framework. The first is its conceptual – some would say 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 given by a number of departments to external reporting indicates the lack of strong incentives for them to take it sufficiently seriously.²⁰

The DoF itself 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/DoF, 2004), but beyond that 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.²¹ The DoF reportedly shrugged off repeated criticisms, such as those made

²⁰ The Auditor-General has observed that: “In my experience, organisations that have reliable performance information for both external requirements and internal management have achieved good performance reporting through strategies such as: establishing a robust performance culture based on public sector values; maintaining strong links between reporting, planning and management; and ensuring strong links between external and internal reporting” (McPhee, 2008, p. 16).

²¹ It is interesting that the current DoF departmental secretary has noted that the DoF remains the main generator of formal “red tape” requirements, generating nearly half of the 2 000 compliance requirements (especially those relating to financial compliance) on government agencies (Tune, 2010).

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 oversight by the DoF had been “lackadaisical”, *i.e.* unenthusiastic (Murray, 2008, p. 88).

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. It can be argued that these reforms in the 1996-2007 period “threw out the baby but kept the bathwater”.

3.3.2. 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 in each state and territory, 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 will often accept these changes. There are several opportunities throughout the year for the Senate to scrutinise government performance.²² The most important involves the scrutiny and passage of the main budget proposals in May; the portfolio budget statements (PBSs) are scrutinised at the same time. 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 and they are prepared under the authority of the portfolio minister. The Senate’s scrutiny of annual reports takes place in November and is much less detailed than that of PBSs; annual reports are prepared by departments on their own authority, and they are reportedly viewed by many parliamentarians as “public relations vehicles” (Thomas, 2009).

The Senate has eight committees that scrutinise the budget bills for portfolios. In addition to examining the information content of PBSs, 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 over 4 000 witnesses (Thomas, 2009). Parliamentary committees also consider ANAO performance audits on specific issues (Box 5).

²² Other opportunities include the Senate’s consideration of the additional estimates bills in February, its consideration of ANAO efficiency and financial audits, and the Senate’s own inquiries into issues relating to government performance and other issues.

Box 5. Performance audits by the Australian National Audit Office (ANAO)

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 AUD 87 million (in 2010-11), the ANAO conducts performance audits and financial audits. Performance audits 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 audits. 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 programmes. About 90% of the ANAO 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:

- aid given by AusAID 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.

Thomas has observed that:

...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. Podger has argued that:

Along with most [departmental] secretaries, I was frustrated by the style of most of the Senate committees I attended, which rarely addressed the major policy or performance issues and mostly pursued more superficial and immediate political matters. I believe these committees need more professional support from the Parliamentary Service, rather than from political staff, to probe more deeply and over time substantial policy and performance matters. (Podger, 2009, p. 188)

Hawke (2007) has 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 *versus* promised performance; other committees have few such questions. Mulgan (2008) conducted a contents analysis of Senate committee budget hearings and found that – comparing

1986 to 1992 to 2003 – there was a growing focus on government results, especially on outputs (rather than outcomes), and less attention was being paid to inputs such as spending and personnel. Mulgan concluded that this reflects the maturing accountability priorities of parliamentarians. Even so, the extent to which Senators referred to information contained in portfolio budget statements and annual reports was not high even in 1986, and fell over the period examined; it was particularly low by 2003. He argued that:

...the fact that Senators have largely ignored the formal documentation placed before them, especially in the case of the much-vaunted Outcomes and Outputs Framework, reveals how far the finance [DoF] bureaucracy fails to meet the accountability needs of parliamentarians. (Mulgan, 2008, pp. 467-468)

Perhaps more positively, Thomas concluded that:

Even if parliamentarians do not take full advantage of the mountainous volumes of performance information made available to them, the language of outputs and outcomes has become [since the 1980s] part of the culture of parliament and more attention is being paid to the effectiveness of policies and programs. (Thomas, 2009, p. 391)

Thus Thomas believes that Australia now displays a “performance-based model of accountability”.²³ 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.

As discussed above, the performance framework of the Howard government emphasised the collection and reporting of performance indicators, although the formal reporting requirements were not enforced by the Department of Finance. The evaluation strategy was abolished, and evaluation was deregulated. 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. It could be expected that this, in turn, would depend on the corporate culture of each individual department and, in particular, on the priorities and commitment of each departmental secretary.

The poor quality of performance information reported in portfolio budget statements and in annual reports does not engender confidence in the totality of performance information collected by departments and agencies. And, *prima facie*, it might also suggest poor internal utilisation of this information by these bodies. This conclusion might be too harsh, however. A small number of agencies reportedly rely on reasonably advanced reporting techniques, such as balanced scorecards, performance dashboards and “traffic light” reports, for their internal reporting and management, in addition to the provision of regular performance information (Hawke, 2007). A possible explanation for this discrepancy is that external reporting to parliament is often not taken particularly seriously, whereas internal reporting for the use of agency senior management would be.

In addition, the voluminous and comprehensive performance information on government service delivery at the federal, state and territory levels that has been published annually since 1995 (for

²³ It is interesting to note in passing that Bovens has developed a taxonomy of the ways in which accountability can improve performance: *i*) preventive, by setting standards to hold institutions to account; *ii*) remedial, by encouraging responsibility to fix problems and to prevent recurrence; and *iii*) educative, by tracing connections between past, present and future policies (Bovens, 2006, p. 26, quoted by O’Loughlin, 2010).

example, SCRCSSP, 1995, and SCRGSP, 2010) is generally regarded as sound. Most of these data relate to state and territory governments, and some data quality assurance is provided by the Productivity Commission.²⁴ Stakeholder surveys have revealed that central and line ministries (in the federal and state governments) find these data helpful in policy analysis and formulation, in assessing the resource needs of agencies, and in briefings for ministers and parliaments.

The lack of close central oversight of departments' performance monitoring activities makes it difficult to reach a definitive conclusion about the collection, quality and internal utilisation of monitoring information. The same can be said about the scale, nature, quality and utilisation of departmental evaluation. Thus, in the Howard era, there was not a single evaluation system but instead the possibility of 20 – one for each department or more, given that major agencies themselves might choose to conduct their own evaluations.

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 Department of Family and Community Services and the Department of Employment and Workplace Relations (Box 6). 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 monitoring and evaluation (M&E). Another reason 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.

Of course, these observations about the supply of monitoring and evaluation information tell us nothing about the extent of its utilisation for purposes of internal departmental management or to aid government decision making. Nevertheless, the fact that at least some departments choose to continue to take evaluation seriously suggests that they derive significant benefits from doing so. Neither the DoF nor the ANAO have conducted in-depth reviews of utilisation, so it is not possible to draw definitive conclusions. There is some fragmentary evidence, however.

Soon after the election of the Howard government in 1996, when the government was still facing a substantial budget deficit, the government embarked on a major effort to cut and to reprioritise government spending. In the area of labour market programmes, there was heavy utilisation of detailed evaluation findings – that had been commissioned under the previous government – concerning the relative cost-effectiveness of various types of labour market programme, such as wage subsidies and direct job creation in the public sector. These findings were used to cut total spending on these programmes by 20%, or about AUD 1.5 billion, over two years. Within the lower budget available for labour market programmes, much more emphasis was placed on wage subsidies (Mackay, 1998, p. 11).

²⁴ The Productivity Commission may devote some additional effort to data quality issues in coming years (Independent Reference Group, 2010).

Box 6. 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 AUD 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, 2002b). The department regularly prepared a research and evaluation plan, including 145 research/evaluation projects. The total budget for this work was AUD 26 million per annum, equivalent to about 1.4% of the department's operational budget (FaCS, 2001, 2002c). Funding was provided to six university research institutes, and for three longitudinal data sets. The research emphasis was on behavioural 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

This department 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 programme participants, with the data interpreted using econometric models to estimate the net effects of programme 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 the 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).

There is some evidence that most departments were not using performance information sufficiently to aid their policy development and internal resource allocation (ANAO, 2003).

The head of the Productivity Commission (the government's main advisory body on microeconomic reform) has criticised a number of policy decisions, including some major infrastructure investments, taken by the Howard government because of doubts about whether they could be justified by evidence and analysis (Banks, 2009a,b).

4. Third period: 2007 to 2011

4.1. Context

A Labor government was elected in November 2007, and Kevin Rudd became the new Prime Minister. He was committed to making substantial changes to many of the policies of the previous

government, and was less favourably disposed towards 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, Prime Minister Rudd commissioned at least 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.

4.2. The policy cycle and government decision making

One feature of this period has been the 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, and about the lack of evidence-based decision making by the government.²⁵ 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 did initiate a number of actions 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 the hands of the Prime Minister and with several close colleagues, rather than in the hands of 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 criticised both implicitly and explicitly. One example is an ANAO performance audit relating to a government body, Infrastructure Australia, that the Labor government created in April 2008 in order to improve the quality of infrastructure planning and the government’s investment strategy. A main function of Infrastructure Australia is to develop a priority list of major infrastructure projects that the government should undertake; the government had stated that “only public infrastructure projects that meet a minimum benchmark social rate of return – determined through rigorous cost-benefit analysis, including evaluation and review – should be funded” (ANAO, 2010, p. 14). However, the ANAO found that, of nine projects that had met the Infrastructure Australia criteria, only seven were funded by the government (at a cost of AUD 6.3 billion), and that an additional ten projects that had either not been subject to cost-benefit analysis or that had failed it were also funded by the government in the May 2009 budget (the cost of these projects is unclear). This *modus operandi* casts doubt on the economic efficiency – as well as the evidentiary basis – of the government’s AUD 20 billion Nation Building Fund, which is meant to be allocated on the basis of impartial assessments by the Infrastructure Fund.

The most criticised 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 AUD 43 billion – *i.e.* about 3.5% of GDP.²⁶ 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 AUD 14 billion and AUD 20 billion (Ergas and Robson, 2010). However, despite repeated recommendations by the Productivity Commission, the OECD, the Business Council

²⁵ In contrast, senior officials during the Howard era usually voiced their concerns only in their retirement speeches or after they had retired.

²⁶ More recent costings have ranged from AUD 35.7 billion to AUD 55 billion.

of Australia and the 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, p. 5).

Henry Ergas, who has been a senior adviser to the government, has concluded that:

... it is difficult to find any period in Australian history when there have been more badly flawed spending programs... What gives rise to the greatest concerns is the weakening of the policy and program evaluation disciplines that, at least in principle, had been endorsed by successive governments.

How could assurances of the importance the government attaches to project appraisal be taken at face value once public spending on an unprecedented scale had been committed on the basis of analyses of unprecedented shallowness? (Ergas, 2010, p. 14)

Ultimately, the quality of evaluation depends on the value governments place upon it. Governments that view project evaluation as merely a nuisance that stands in the way of the decisions they want to take, and that believe they can get away with no evaluation or poor quality evaluation will, over time, invariably succeed in devaluing the evaluation process. This has, we believe, occurred in Australia in recent years. (Ergas and Robson, 2010, p. 154)

In a similar vein, the head of the Productivity Commission has criticised 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 shall be seen below, the government has announced several initiatives that should ameliorate these weaknesses.

4.3. Changes in public sector management

The government has taken several steps to improve the policy-advising capacities of the public service. One step is the provision in early 2009 of AUD 70 million in additional funding over four years to the three central departments – Prime Minister and Cabinet (PM&C), Treasury and Finance (DoF) – to enable them to strengthen their policy capacities. Another step has been the government’s acceptance of a blueprint for the reform of government administration.

4.3.1. “Ahead of the Game”

The reform blueprint, entitled “Ahead of the Game”, was prepared by an advisory group under the leadership of the secretary of the department of PM&C (AGRAGA, 2010). The blueprint contains a number of initiatives to improve the performance of the federal public sector (summarised in Box 7).

Box 7. Ahead of the Game: 2010 public sector reforms

- Surveys are to be conducted of citizen satisfaction with federal government service delivery, to identify the drivers of citizen satisfaction, and to enable benchmarking comparisons with other jurisdictions and internationally. Survey results are to be presented in annual reports. The surveys may eventually be extended to other levels of government.
- Reviews of departmental and agency performance are to be conducted, on a five-yearly basis, to assess their strategy, programme delivery and organisational effectiveness. The reviews will be conducted by eminent external reviewers, and will lead to agency capability plans. The structure of these reviews was influenced by the capability reviews of Canada and New Zealand.
- Departments and agencies are encouraged to strengthen their policy capabilities and to do more research and evaluation. To support these efforts, the Australian Public Service Commission (PSC) 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 Department of Finance (which was renamed the Department of Finance and Deregulation in 2007) is to streamline red tape – *i.e.* administrative and legislative requirements placed on departments and agencies – to reduce the compliance burden, especially with respect to financial management and corporate functions such as human resources. The DoF has identified over 2 000 such requirements.
- The DoF is to identify agencies that could be amalgamated or subsumed into departments. The DoF is to assist agencies to review their corporate functions with a view to more centralised provision.

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

These reforms reflected the government's acceptance of the need to strengthen the policy capabilities of the public service: to undertake rigorous research, gather and analyse data, conduct evaluations of policies and programmes, and provide the highest-quality policy advice. Departments are encouraged to improve their policy capabilities. This principles-based approach means that they have the autonomy to decide how much effort, if any, to devote to this reform. They are also encouraged to develop partnerships with academic and research institutions.

The government also accepted that Australia lags behind other countries in systematically measuring departmental and agency performance, in areas such as strategy development, programme delivery and organisational 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 and Finance, the Public Service Commission, the agency itself, and other agencies as appropriate. The lead agency for these reviews will be PM&C; it will develop review methodologies, including the specific issues to be reviewed for each agency, and will manage the reviews. 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. Once reviews have been conducted, the head of PM&C and the agency will jointly develop a capability improvement plan, and each agency head will be held accountable for its implementation.

The government expects that it will take a number of years for these reforms to be fully implemented. A number of the planned reforms are intended to create a more unified public service –

a partial reversion to the situation before the Howard era. However, the government, and perhaps PM&C in particular, is cautious in how these revisions are presented. It has argued that: “There are elements of central planning in the reforms, but this should not be read as a centralist agenda” (AGRAGA, 2010, p. 74). For a public sector and public service that regard autonomy for departments and agencies as a central tenet, any suggestion of a reversal in this philosophy appears to be regarded as an anathema.

One way of viewing these reforms is that they constitute an attempt to fix a number of the problems caused by the public service changes introduced by the Howard government over the 1996-2007 period.²⁷ The changes to the Outcomes and Outputs Framework, discussed below, also belong in this remediation category.

4.3.2. The “Outcomes and Programmes Reporting Framework”

Lindsay Tanner, who became the Labor government’s finance minister after the 2007 election, had long criticised the quality and lack of transparency of budget documentation, including portfolio budget statements. As finance minister, he implemented “Operation Sunlight” which had the objectives of improving the transparency of government budgetary and financial management, and promoting good governance practices (Tanner, 2008). This initiative included a number of substantive changes to the government’s performance reporting:

- Replacement in 2009 of the previous government’s Outcomes and Outputs Framework with a new “Outcomes and Programmes Reporting Framework”.
- Budget reporting in portfolio budget statements to be provided on a programmatic basis – *i.e.* a reintroduction of programme budgeting which had been abolished by the Howard government in 1999 – rather than on an outcomes basis. This reform was introduced in the 2009-10 budget documentation.
- Improved specification of outcomes, to make them more specific and tangible. The Department of Finance played a leading advisory role *vis-à-vis* line departments and agencies; outcomes/programmes structures were agreed with each department and agency. It was recognised, however, 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 bridge between the information provided by the outcomes/outputs framework and that of the outcomes/programmes framework.
- The introduction of cross-portfolio outcomes, with shared responsibility across portfolios and with performance monitored by the Department of the Prime Minister and Cabinet.
- Better reporting in annual reports of targets and comparisons with actual performance, with effect from FY2010. Targets are provided not only for the budget year but also for the three out-years of the forward estimates period. The finance department is to publish an omnibus report comparing the portfolios’ actual performance with their targets.

²⁷ Other areas of public sector management where some recentralisation is occurring include government procurement and public sector pay (McPhee, 2009).

The ANAO is currently conducting a performance audit of the new Outcomes and Programmes Reporting Framework.

4.4. Government monitoring and evaluation

The secretary of the DoF recently briefed the government, noting that:²⁸

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, pp. 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 favourable). But it is not being done to support the government's overall decision making nor for accountability purposes.

The government's formal M&E requirements remain focused on monitoring²⁹ but there are some developments that suggest that considerably more attention could soon be paid to evaluation, with the DoF once again taking a leading role.

As discussed above, the renovated Outcomes and Programmes Reporting Framework remains the centrepiece of the government's approach to performance monitoring. Only time will tell if the technical problems of its predecessor framework have been satisfactorily addressed and, more importantly, if departments and agencies choose to take this newer framework more seriously. The Labor government's commitment to transparency is a significant step in the right direction. Agencies might choose to be more receptive to this government priority, and the DoF may – or may not – choose to take a more pro-active role in policing the behaviour of agencies. However, given that the current climate in the public service remains focused on agency autonomy, and given the DoF mandate (reflected in its new title, the Department of Finance and Deregulation) to cut rather than add to government red tape, there do not seem to be strong grounds for optimism that the new framework will be markedly more successful than its predecessor.

4.4.1. Performance reporting on government service delivery

One continuing strong feature of performance monitoring and reporting in Australia is the annual reports on government service delivery. As discussed in Section 2 above, these reports involve the publication of detailed performance information on services provided by the federal, state and territory governments. At least as important as the performance information is the discussion of the reasons for performance differences among the different jurisdictions. This discussion also refers to contextual

²⁸ This brief, which was initially confidential, was prepared for the Labor government re-elected in September 2010. Although the comments applied to the state of evaluation at that time, they would be equally applicable to the situation at the time of the Howard government.

²⁹ From 2009, budget rules were changed to require new policy proposals (NPPs) to outline the review and performance measurement framework under which each proposal would operate if agreed by the Cabinet. The proportion of NPPs meeting this requirement, and the quality of the performance frameworks, are not known at this stage. As noted in footnote 4, a comparable requirement during the 1987-97 evaluation strategy had only limited success.

differences that help explain the differences in performance; these include issues such as population size, remoteness, and so on.

These reports relate to government service delivery in 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, the disabled and children (SCRGSP, 2010). These services collectively involve annual expenditure of AUD 136 billion, covering 13.1% of GDP; most of these services are delivered by the state and territory governments. The voluminous 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. The reports also provide information on trends over time. As noted earlier, the main purpose of these reports is to provide transparency and accountability regarding government service provision, and it is hoped that this transparency will drive improvements in performance (O'Loughlin, 2010).

User surveys reveal the continuing usefulness of these reports on service delivery (Productivity Commission, 2007). Almost 80% of respondents are satisfied or very satisfied with the reports. The main uses of the reports are: to brief management and ministers (69%); as a research tool (68%); to assess performance (58%); and to develop policy (43%). The reports are also used by parliaments and state audit offices. However, the surveys also highlight that improved comparability of data, better data quality and more timely data are all areas that are judged to be needing improvement.

The reports on government service delivery rely completely on the 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 (McClintock, 2010), this appears to be the main potential weakness of these reports. For example, the health minister of New South Wales has publicly repudiated his government's data on hospital performance: they 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 also stated that the governments of New South Wales and of Victoria have falsified their hospital performance data (Harris, 2009); these data are always highly politically sensitive. He noted that:

Notwithstanding the significance of performance information, most Australian jurisdictions do not require annual auditing of the government's published indicators. Only the West Australian parliament compels state agencies to report performance indicators annually and only that state requires the auditor-general to issue an opinion each year on the relevance and appropriateness of those indicators. (Harris, 2009, p. 36)

That the reliability of these data on government service provision is uncertain has to be of significant concern. On the other hand, if the visibility of these data is encouraging some governments – or at least some departments in some governments – to falsify their data, then that at least tells us that data transparency is providing some “bite”. In other words, the performance information is sufficiently sensitive and important that it is worth falsifying; conversely, when data are ignored and are viewed as unimportant, then it can be inferred that performance reporting is essentially a waste of time.

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” website (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 socio-economic characteristics of student households.³⁰

This website 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 nine million hits on the first day. The federal government has been a strong promoter of the website, notwithstanding strong opposition to it (on philosophical grounds) from teacher unions. The federal government launched a similar website for hospitals in late 2010 (www.myhospitals.gov.au).

One final observation on the data on government service provision is that, while Australia has made considerable progress in collecting comparable performance information on service delivery in the different jurisdictions of the country, the opportunity for learning across these jurisdictions – *i.e.* knowing what does or does not work well in different jurisdictions, and the reasons why – has not yet been fully exploited (Banks, 2009a). It is expected that the reports on government service delivery will in future place more emphasis on highlighting examples of good performance, to encourage the more widespread adoption of good service delivery practices and to encourage innovative reforms.

4.4.2. Performance information required in federal-state funding

The federal government provides about half (AUD 94.1 billion in FY2010-11) of the total funding for state and territory governments (Commonwealth of Australia, 2010). Of this amount, half (AUD 45.4 billion) is earmarked for specific purposes that include schools, skills training, health care, public housing and services for the disabled. The federal government sets broad policy parameters for these payments, in terms of the overall objectives, leaving the states and territories with full autonomy as to how they use these funds. Most of the remainder of the federal funding of the states and territories comprises general budget assistance, *i.e.* untied funding.

In March 2008, the Labor government reformed federal/state relations by reducing the number of special-purpose payments and by creating national agreements for each type of payment (O'Loughlin, 2010; McClintock, 2010). These agreements are performance-focused. They specify the roles and responsibilities of the different levels of government as well as the specific objectives of the funding, and they set performance benchmark targets in terms of measurable output and outcome performance indicators. However, there is generally no provision for funds to be withheld if a jurisdiction fails to meet a performance benchmark.

The statistics that are used to measure the performance of each jurisdiction are essentially the same as those used in the reports on government service delivery. Both these reports and the formal annual reports on each national agreement are prepared by the Productivity Commission. The data that are used are sourced from national statistics collections (especially the population census and the household surveys) and from governments' own administrative data. However, the data are not considered fully adequate – in terms of their coverage, quality, timeliness and comparability – partly because they were developed for other purposes (McClintock, 2010).

There appears to be a consensus that the performance reporting framework for the national agreements is too cumbersome and needs to be simplified. For example, the national healthcare

³⁰ The index is based on measures of average income, level of education, and types of employment of households whose students attend the school. The index also measures remoteness and the proportion of Aboriginal students at the school. These socio-economic characteristics are correlated with student test results; it is the remaining variance in student average performance that a school can influence. The school outcomes data are the same as those used in the annual reports on government service delivery.

agreement specifies 70 performance indicators, 26 progress measures, 15 outputs and 11 outcomes, and the annual performance report is currently almost 700 pages long (SCRGSP, 2009). The chair of the Council of Australian Governments (COAG) Reform Council has noted that this complexity makes it much more difficult to achieve effective public accountability (McClintock, 2010). The performance frameworks for the national agreements are currently being reviewed.

Notwithstanding these early difficulties with the data and with the performance reporting frameworks, it can be concluded that the early identification of them by senior stakeholders provides some confidence that they will be satisfactorily addressed.

A notable feature of the national agreements is that for some – currently minor – parts of them there is performance-based funding. The funding is termed “reward payments” to states and territories that are able to achieve ambitious performance benchmarks (COAG Reform Council, 2010). The amount of this funding is currently modest (around AUD 400 million per annum) but might grow in coming years if performance-based funding is judged to provide effective performance incentives.

4.4.3. A rejuvenated DoF evaluation strategy

The secretary of the DoF advised the newly re-elected government in 2010 that “procedures for evaluation and review ... need to be reinvigorated” (DoF, 2010, p. 1.1). He has noted the range of evaluations and reviews that are conducted by different entities within the government; these include the department’s own strategic reviews, DoF *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 the 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. As noted earlier, the DoF secretary stated that while some departments’ programmes have been subject to good-quality self-evaluation, the average quality appears to be low. Moreover, the evaluations conducted by line departments and agencies are done for their internal decision making and management; they are typically not conducted to support the government’s overall decision making nor for accountability purposes. In addition, departmental secretaries tend to be disinclined to make evaluation findings available – either publicly or to the DoF – and so the DoF is unable to use these findings to inform its policy advice. The DoF secretary believes that this partly reflects concerns that the DoF might use any adverse evaluation findings for savings options to cut programmes. Podger noted that, consistent with a secretive approach, a number of his fellow departmental secretaries take measures to frustrate freedom-of-information legislative requirements – for example, by ensuring that their senior officials keep few file notes and regularly destroy their diaries (Podger, 2005, quoted by Kelly, 2006, p. 12). Such approaches will tend to frustrate both formal and informal accountability requirements, in terms of accountability to the DoF and to other central agencies, to the elected government and to the parliament. Thus, a question for the government to consider is whether the fashion towards secretarial and departmental autonomy has become too permissive.

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, handout notes for slide 14), 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, pp. 3-5). The DoF is still considering options for such a framework, and one model that is attractive to the DoF secretary is the Canadian approach which involves the evaluation of

all government programmes on a four-year cycle in order to determine the efficiency, effectiveness and alignment with government priorities of all programmes.³¹ The DoF secretary 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 programme savings); *ii*) some centralisation of evaluation by means of DoF oversight of departmental and agency evaluations, in order to enhance evaluation quality; and *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 programme of reviews, and by provision of greater authority for the reviews, such as by ensuring that their terms of reference are explicitly endorsed by the Cabinet (DoF, 2010, pp. 3-5).

The DoF secretary is keen to avoid what he perceives as the problems with Australia's evaluation strategy in the 1987-97 period: that the requirements to prepare detailed portfolio evaluation plans and to evaluate every programme on a 3-5 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 to be faced 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 programmes.
- 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 programme 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 programmes are evaluated at the right time, the right issues are evaluated, and 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 evaluation as a core part of their work rather than as a “politically correct” distraction.

³¹ Income transfer programmes are explicitly excluded in the Canadian system. The DoF secretary has argued for the inclusion of such programmes in a rejuvenated evaluation strategy in Australia.

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 topic is discussed below.

5. Lessons from the Australian experience

The Australian evaluation system lasted for a decade, from 1987 to 1997. During this period, Australia became a model of evidence-based decision making and performance-based budgeting. Thus 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, and this in turn required an element of trial and error. There was also a need for ongoing monitoring, review and adjustment to the evaluation strategy itself. More fundamentally, there were a number of important 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. The reform champions at the most senior levels of the DoF, and reformist ministers in the Hawke/Keating governments (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 the 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, and this would have reduced their reliability and value. Related to this problem, there were evidently insufficient evaluation skills within the public service; most line departments relied on programme areas to conduct the evaluations of their own programmes, 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 maximise 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 programme 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 the 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 programme 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 programmes, and subsequent evaluations could be expected to have revealed their poor efficiency, effectiveness or appropriateness; this would usually have led to the programmes 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 programmes 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 cannot be investigated further; 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 other government systems for monitoring and evaluation. One such factor is obviously 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 centralised performance orientation within government. Related to this is their replacement by M&E “opponents” – key officials who are either sceptical 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 now to show that they constitute a failed experiment in public sector management. Of course, the evaluation strategy was only one casualty of this mindset. The 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 did act as a very powerful motivator 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 programmes are cost-effective and which are not. The large budget surpluses that Australia enjoyed from 1998-99 onwards 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. The DoF had hoped 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 reason 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 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 attitude.

The evaluation strategy that existed from 1987 to 1997 can be judged to have been a fairly successful evaluation system because it achieved a high level of utilisation, 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 the expertise of the DoF regarding budget estimates and policy advice. 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 or the strength of any relationship between government spending and subsequent outcomes; 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 facing any government that wishes to achieve a more evidence-based approach to its decision making. Perhaps ironically, these very same choices once again face the Australian government and the Department of Finance. One choice is how centralist or devolutionary to be. The answer 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, then this would seem to favour a more devolutionary approach – unless it could be convincingly shown that these line units would tend to significantly under-invest in M&E if they were left to their own devices. The reality that most departments and agencies currently appear to engage in little evaluation activity would suggest that a devolutionary approach to evaluation is insufficient to support sound programme management.

If regular M&E information on the performance of all government programmes is needed, and on a systematic, whole-of-government basis, to aid government decision making – *i.e.* to assist in deciding new programmes or which existing programmes to cut – then this would argue for some form of centralist approach with leadership from the central budget office (*i.e.* the DoF). Only a centralist approach is likely to be able to 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. Similar arguments would apply when a major purpose of M&E information is to support accountability for performance, such as accountability to the parliament.

Of course, only some aspects of Australia's evaluation system in the 1987-97 evaluation system were centralised: the evaluation policy itself, and the central oversight and quality assurance roles of the DoF. Evaluations themselves were planned and conducted by line departments, with the 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, the DoF, as the system's architect, was attempting to have the best of both worlds – to bring together its own (and other central agencies') objective scrutiny and willingness to ask tough evaluation questions with the programme expertise of line departments and agencies. This approach had the potential to meet the evidentiary needs of budget decision making as well as to maximise the "ownership" of evaluation findings by line departments and agencies. It can fairly be said that these needs were substantially met, albeit at the cost of some important reduction in evaluation quality. This reduction in quality could have been avoided if the 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 the current secretary of the DoF) have argued that another disadvantage of the evaluation system was that it imposed an overly onerous burden on the DoF and 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. It is also worth remembering that the partially centralised approach to evaluation from 1987 to 1997 was framed as the quid pro quo for a devolutionary approach to other public sector reforms. It reflected the observation that most line departments and agencies prefer not to devote much effort to measuring their performance. This observation is equally relevant today.

The nature of a centralised or partially centralised 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 co-operation of line departments, whose programme 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.³² With the evaluation strategy, the 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.³³ The carrots included the advisory support provided by the DoF, as well as possible access to resource agreements for line departments to aid their management of the programmes revealed by evaluations to be under-performing. The sticks were related to the ongoing relationship of the DoF with line departments and the finance department's ability to influence their budget allocations. The 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.³⁴ The sermons included persistent advocacy by the secretary of the 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

³² This has happened with Chile's centralised monitoring and evaluation system, for example, which was designed by and is managed for the budget needs of the finance ministry (Rojas *et al.*, 2005).

³³ A detailed taxonomy of carrots, sticks and sermons for encouraging M&E is provided by Mackay (2007, Chapter 11).

³⁴ The 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.

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 the DoF, other central departments, line departments, external advisers and others in their policy-advising roles (this also necessitates having a sufficient number 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 under way 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 programme, 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 programme management, for 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.³⁵ Only time will tell if current efforts to renovate monitoring and evaluation in the Australian government are successful.

BIBLIOGRAPHY

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

ANAO (Australian National Audit Office) (1991a), *Implementation of Program Evaluation – Stage 1*, Efficiency Audit Report No. 23, 1990-91, ANAO, Canberra.

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

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

³⁵ 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.

- ANAO (1992b), *Auditing Program Evaluation – ANAO Performance Auditing Guide*, ANAO, Canberra.
- ANAO (1992c), *Department of the Treasury – Procedures for Managing the Economic Policy Program*, Efficiency Audit Report No. 36, 1991-92, ANAO, Canberra.
- ANAO (1993), *Program Evaluation: Strategies, Practices and Impacts – Industry, Technology and Regional Development Portfolio*, Efficiency Audit Report No. 35, 1992-93, ANAO, Canberra.
- ANAO (1996), *Performance Information – Department of Employment, Education, Training and Youth Affairs*, Performance Audit Report No. 25, 1995-96, ANAO, Canberra.
- ANAO (1997), *Program Evaluation in the Australian Public Service*, Performance Audit Report No. 3, 1997-98, ANAO, Canberra.
- ANAO (2000), *Family Relationships Services Program – Department of Family and Community Services*, Performance Audit Report No. 24, 2000-01, ANAO, Canberra.
- ANAO (2001a), *Performance Information in Portfolio Budget Statements*, Performance Audit Report No. 18, 2001-02, ANAO, Canberra.
- ANAO (2001b), *Developing Policy Advice*, Performance Audit Report No. 21, 2001-02, ANAO, Canberra.
- ANAO (2003), *Annual Performance Reporting*, Performance Audit Report No. 11, 2003-04, ANAO, Canberra.
- ANAO (2007), *Application of the Outcomes and Outputs Framework*, Performance Audit Report No. 23, 2006-07, ANAO, Canberra.
- ANAO (2010), *Conduct by Infrastructure Australia of the First National Infrastructure Audit and Development of the Infrastructure Priority List*, Performance Audit Report No. 2, 2010-11, ANAO, Canberra.
- ANAO/DoF (Department of Finance) (1996), *Performance Information Principles*, ANAO, Canberra.
- ANAO/DoF (Department of Finance and Administration) (2004), *Better Practice Guide: Better Practice in Annual Performance Reporting*, ANAO, Canberra.
- Banks, Gary (2009a), *Challenges of Evidence-Based Policy-Making*, Productivity Commission and Australian Public Service Commission, Canberra.
- Banks, Gary (2009b), “Evidence-based policy-making: What is it? How do we get it?”, *Australian National University Public Lecture Series*, presented by ANZSOG (Australia and New Zealand School of Government), 4 February 2010, Productivity Commission, Canberra, quoted by Alan Mitchell, 2009.
- 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.

- Barrett, Pat, Auditor-General of Australia (2003), “Australian Public Sector Financial Management Reforms”, presentation at a World Bank seminar, 25 June.
- Blöndal, Jon R., Daniel Bergvall, Ian Hawkesworth and Rex Deighton-Smith (2008), “Budgeting in Australia”, *OECD Journal on Budgeting*, 8(2):133-196.
- Borthwick, David, quoted by Verona Burgess (2009), “Everyone on the same page”, *Australian Financial Review*, 13 March, p. 46.
- Bovens, M. (2006), “Analysing and Assessing Public Accountability: A Conceptual Framework”, *European Governance Papers C-06-02*, quoted by Mary Ann O’Loughlin, 2010.
- Briggs, Lynelle (2005), “A Passion for Policy” in John Wanna (ed.), *A Passion for Policy: Essays in Public Sector Reform*, pp. 1-18, *ANZSOG Lecture Series 2005-06*, Australian National University, Canberra.
- Burgess, Verona (2009), “Everyone on the same page”, *Australian Financial Review*, 13 March, p. 46.
- 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, Independent Evaluation Group, The World Bank, Washington DC.
- Chan, Matthew, Mark Nizette, Lisa La Rance, Charles Broughton and Derek Russell (2002), “Australia”, *OECD Journal on Budgeting*, 1(4):35-69.
- COAG (Council of Australian Governments) Reform Council (2010), *COAG Reform Agenda: Report on Progress 2010*, COAG Reform Council, Sydney.
- Commonwealth of Australia (2010), *2010-11 Australian Government Budget – Budget Paper No. 3*, Commonwealth of Australia, Canberra.
- Conroy, Stephen, Communications Minister (2010), “NBN cost-benefit analysis a waste of time and money: Stephen Conroy”, *The Australian*, 5 November, p. 5.
- Crossfield, Len and Anne Byrne (1994), *Review of the Evaluation Function in DEET*, Department of Employment, Education and Training, and Department of Finance, Canberra.
- DEWR (Department of Employment and Workplace Relations (2002), *2001-02 Annual Report*, Australian Government Publishing Service, Canberra.
- DoF (Department of Finance) (1991), *Handbook of Cost-Benefit Analysis*, DoF, Canberra.
- DoF (1993), *The Cost of Evaluations: The Findings of a Pilot Study*, DoF, Canberra.
- DoF (1994a), *Doing Evaluations: A Practical Guide*, DoF, Canberra.

- DoF (1994b), “The Use of Evaluation in the 1994-95 Budget”, Finance Discussion Paper, DoF, Canberra.
- DoF (1996), “Performance Information Review: Overview of the First Year of the Review”, DoF, Canberra.
- DoF (Department of Finance and Administration) (2000), *The Outcomes and Outputs Framework: Guidance Document*, DoF, Canberra.
- DoF (2007), *2006-07 Annual Report*, DoF, Canberra.
- DoF (Department of Finance and Deregulation) (2010), “Incoming Government Brief”, DoF, Canberra.
- Ergas, Henry (2010), “Faulkner and Tanner fell short of their lofty aims”, *The Australian*, 13 July.
- Ergas, Henry and Alex Robson (2010), “Evaluating Major Infrastructure Projects: How Robust are our Processes?” in Productivity Commission, *Strengthening Evidence-Based Policy in the Australian Federation, Volume 1: Proceedings*, pp. 127-167, Productivity Commission, Canberra.
- FaCS (Department of Family and Community Services) (2001), *Research and Evaluation Framework*, FaCS, Canberra.
- FaCS (2002a), *Annual Report 2001-02*, FaCS, Canberra.
- FaCS (2002b), *Portfolio Budget Statements 2002-03 – Family and Community Services Portfolio*, Commonwealth of Australia, Canberra.
- FaCS (2002c), *Research and Evaluation Digest*, FaCS, Canberra.
- Harris, Tony (2009), “Eradicate this dodgy data”, *Australian Financial Review*, 14 April.
- Hawke, Lewis (2007), “Performance Budgeting in Australia”, *OECD Journal on Budgeting*, 7(3):133-147.
- Hawke, Lewis and John Wanna (2010), “Australia after Budgetary Reform: A Lapsed Pioneer or Decorative Architect?” in John Wanna, Lotte Jensen and Jouke de Vries (eds.), *The Reality of Budgetary Reform in OECD Nations: Trajectories and Consequences*, pp. 65-90, Edward Elgar Publishing, Cheltenham, United Kingdom.
- Howard, Cosmo (2005), “The Policy Cycle: A Model of Post-Machiavellian Policy Making?”, *Australian Journal of Public Administration*, (64)3:3-13.
- Independent Reference Group (2010), *Review of the Report on Government Services’ Performance Indicator Framework. Report to the Steering Committee for the Review of Government Service Provision*, Productivity Commission, Canberra.
- 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), *The Canadian M&E System: Lessons Learned from 30 Years of Development*, ECD Working Paper No. 23, Independent Evaluation Group, The World Bank, Washington DC.
- Mackay, Keith (1998), *The Development of Australia’s Evaluation System*, ECD Working Paper No. 4, Operations Evaluation Department, The World Bank, Washington DC.
- Mackay, Keith (2007), *How to Build M&E Systems to Support Better Government*, Independent Evaluation Group, The World Bank, Washington DC.
- Mackay, Keith (2011), *The Australian Government’s Performance Framework*, ECD Working Paper No. 25, Independent Evaluation Group, The World Bank, Washington DC.
- McClintock, Paul (2010), “Statistics, public accountability and the federation”, presentation to the NatStats 2010 Conference, COAG Reform Council, Council of Australian Governments, Sydney.
- 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, Australian National Audit Office, Canberra.
- McPhee, Ian, Auditor-General of Australia (2008), “The ANAO’s Contribution to Improving Public Sector Administration”, presentation to the Canberra Evaluation Forum, 21 February, Australian National Audit Office, Canberra.
- McPhee, Ian, Auditor-General of Australia (2009), “The Business of Government: Why Public Sector Management Must Evolve”, Brookes Oration Presentation to the Deakin Business School, Deakin University, Australian National Audit Office, Canberra.
- Mitchell, Alan (2009), “Hard data paves way to reform”, *Australian Financial Review*, 16 February, p. 20.
- 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*, Canberra.
- OECD (2002), “Public Sector Modernisation: A New Agenda”, GOV/PUMA(2002)2, OECD, Paris.
- O’Loughlin, Mary Ann (2010), “Intelligent Federalism: Accountability Arrangements under COAG’s Reform of Federal Financial Relations” in Productivity Commission, *Strengthening Evidence-Based Policy in the Australian Federation, Volume 1: Proceedings*, pp. 259-274, Productivity Commission, Canberra.
- 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.

- Parliament of Australia, Joint Committee of Public Accounts and Audit (2004), *Review of Auditor-General's Reports 2003-04: First and Second Quarters*, Parliament of Australia, Canberra.
- Parliament of Australia, Senate Standing Committee on Finance and Public Administration (2007), *Transparency and Accountability of Commonwealth Public Funding and Expenditure*, Parliament of Australia, Canberra.
- Podger, Andrew (2005), "Retirement speech", 30 June (quoted by Paul Kelly, 2006, pp. 12-13).
- Podger, Andrew (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.
- Podger, Andrew (2009), *The Role of Departmental Secretaries: Personal Reflections on the Breadth of Responsibilities Today*, Australian National University, Canberra.
- Productivity Commission (2007), *Productivity Commission Annual Report 2006-07*, Productivity Commission, Canberra.
- 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*, The World Bank, Washington DC.
- Russell, Derek (2003), "Performance Management and Evaluation: The Australian Experience" in Jay-Hyung Kim (ed.), *Developing a Performance Evaluation System in Korea*, Korea Development Institute, Seoul.
- Schick, Allen (1996), *The Spirit of Reform: Managing the New Zealand State Sector in a Time of Change*, State Services Commission, Wellington.
- Schick, Allen (1998), "Why most developing countries should not try New Zealand's reforms", *World Bank Research Observer*, 13:123-31.
- Schick, Allen (2002), "Does Budgeting Have a Future?", *OECD Journal on Budgeting*, 2(2):7-48.
- SCRCSPP (Steering Committee for the Review of Commonwealth/State Service Provision) (1995), *Report on Government Service Provision*, Industry Commission, Melbourne.
- SCRGSP (Steering Committee for the Review of Government Service Provision) (2009), *National Agreement performance information 2008-09: National Healthcare Agreement*, Productivity Commission, Canberra.
- SCRGSP (2010), *Report on Government Services 2010*, Productivity Commission, Canberra.
- Steketee, Mike (2008), "Hospital statistics 'worth nothing'", *The Weekend Australian*, 11-12 October.
- Stutchbury, Michael (2010), "Imagine if Labor's broadband payoff is just hype", *The Australian*, 7 December.
- Symonds, Alexander (2009), "Reviews up for review", *Australian Financial Review*, 5 February.

- Tanner, Lindsay (2008), *Operation Sunlight: Enhancing Budget Transparency*, Australian Government, Canberra.
- Thomas, Paul G. (2009), "Parliament Scrutiny of Government Performance in Australia", *Australian Journal of Public Administration*, 68(4):373-398.
- Tingle, Laura (2009), "Spending to be spelled out", *Australian Financial Review*, 12 May.
- Tune, David (2010), "Evaluation: Renewed Strategic Emphasis", presentation to the Canberra Evaluation Forum, 19 August, Department of Finance, Canberra.
- 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 – Better Guardians?*, pp. 1-29, Edward Elgar Publishing, Cheltenham, United Kingdom.
- Webb, Richard (2010), *The Commonwealth Budget: Process and Presentation*, Parliament of Australia Research Paper, Department of Parliamentary Services, Canberra.
- Wilkins, Roger (2008), *Strategic Review of Australian Government Climate Change Programs*, Department of Finance and Deregulation, Canberra.