Public Disclosure Authorized

Public Disclosure Authorized

64385

The World Bank

PREMnotes

MARCH 2011
NUMBER 8

Special Series on The Nuts & Bolts of M&E Systems

The Australian Government's M&E System

Keith Mackay

Countries from all over the world have shown an interest in Australia's experience in creating a monitoring and evaluation (M&E) system that supports evidence-based decision making and performance-based budgeting. The Australian M&E system in existence from 1987–97 was generally considered to be one of the most successful and was driven by the federal Department of Finance (DoF). This note discusses the genesis, characteristics, and success of this particular system and briefly considers the Australian government's approach to M&E after the system was abolished. The contrast between these two periods provides many valuable insights into success factors and challenges facing successful M&E systems, and into implementing evidence-based decision making more broadly.

By the early 1980s, Australia was facing a very difficult economic situation—not unlike the current difficulties faced by many developed and developing countries. A reformist Labor government was elected in 1983, and it was committed to a substantial reorientation of public spending toward the poor, while also significantly reducing budget spending. By 1989–90, the Labor government had reduced the budget share of gross domestic product (GDP) from 30 to 23 percent—by international standards, this is a major reduction. The crisis situation provided powerful incentives for fiscal discipline and for a series of ambitious microeconomic reforms, such as a floating exchange rate, tariff reductions, labor market flexibility, and the privatization of government business enterprises.

The fiscal pain of budgetary cuts was shared by the federal, state, and territorial governments; the federal government collects the bulk of taxes in Australia, and it provides considerable funding to other levels of government. Most public services are the responsibility of the states and territories.

The federal government did not want to simply reduce spending, it also wanted to significantly improve its efficiency and effectiveness and thereby receive greater value for money from government expenditures. A number of public sector reforms were implemented to encourage this new approach, including a substantial devolution of powers and responsibilities to government departments known as "letting the managers manage." The reforms also included a series of changes to streamline the budget system and enable the budget process to focus on "big picture" policy issues rather than on smaller details. Thus, Australia led the world in introducing a medium-term expenditure framework, a system of budget forward estimates, and consolidated running costs.

DoF was a major driver of these public sector reforms and had the active support of other central departments and a number of reform-oriented government ministers. The senior leadership of DoF wanted to move the departmental focus from relatively minor spending issues to higher-level policy issues, as spelled out in its policy analysis and briefings prepared for the annual budget process.

While DoF leadership had hoped that sector departments and agencies would conduct M&E to help them manage their own performance, it was evident that they were not prepared to do so. DoF therefore asked the Cabinet to agree to a formal evaluation strategy to force departments (and agencies) to plan and conduct evaluations on a systematic basis (box 1). Cabinet agreement was formalized in a *cabinet decision*—in Westminster systems of government, such decisions virtually have the force of law, and few public servants would dare to defy them. An advantage of such decisions is that they can be taken quickly; a disadvantage is that the lack of a legislative basis means they can easily be reversed when there is a change in government.[1]

Main Characteristics of the M&E System

Evaluation was the most important component of the M&E system. Evaluation was stressed because it was considered to be much more valuable than performance information because of the detailed insights it could provide into issues of causality. Some effort was paid to the collection and publication of monitoring information, but mainly from 1995 onwards.

The evaluation strategy had three main objectives. The first, and arguably the most important, objective was to provide fundamental information about program performance to aid Cabinet decision making and prioritization, particularly in the annual budget process, when a large number of competing proposals are advocated by individual ministers. The second objective was to encourage program managers within departments to use evaluations to improve their programs' performance. Lastly, the strategy aimed to strengthen accountability in a devolved environment by providing formal evidence of program managers' oversight and management of program resources. The strategy thus had the important element of "making the managers manage."

It was the responsibility of individual departments to conduct their own evaluations; they decided which programs to evaluate, what the focus of each evaluation would be, and the evaluation techniques that would be used. Many different types of evaluations were conducted, including rapid evaluations, which can take at least three to four months to complete; rigorous impact evaluations, taking 12–18 months; and cost-benefit analyses, which were conducted mostly for infrastructure investments. A sample of these evaluations showed that their cost ranged from about $56,000 to $560,000 (in 1993 prices). Departments could decide not to evaluate a program if it had recently been subject to a performance audit by the national audit office; these audits were also regarded as a form of evaluation.

Although features of the evaluation strategy had a devolved nature, there were also key aspects that were centralized. The strategy itself was designed and managed by the DoF. By overseeing the strategy, DoF played a strong quality assurance role—a policing role—and it could amend the strategy when it deemed an element to be ineffective. DoF also used its significant power and influence as the central budget agency (and as the overseer of departments' financial estimates) to promote the benefits of evaluations, and to point out to departments if they were not investing sufficient effort into evaluation. On a more frequent basis, the DoF desk officers (and also the desk officers of the other key central agencies—the Treasury and the department of the Prime Minister and Cabinet)—who "shadowed" each line department and agency would attempt to influence the evaluation plans of the line departments. The key aspects of evaluation plans were the decisions about which programs would be evaluated, when, and which particular issues would be addressed in each evaluation. DoF involvement was intended to ensure that evaluations addressed the difficult issues of

Box 1: The Australian Government's Evaluation Strategy

There were four formal requirements:

1. Every program should be evaluated every 3–5 years.
2. Each portfolio (that is, comprising a line department plus outrider agencies) prepares an annual portfolio evaluation plan (PEP), with a 3-year forward coverage, and submits it to DoF. These plans included major program evaluations with substantial resource or policy implications.
3. Ministers' new policy proposals should include a statement of proposed arrangements for future evaluation.
4. Completed evaluation reports should be regularly published, except in cases of policy sensitivity, national security or commercial-in-confidence considerations, and that the budget documentation that departments table in parliament each year should also report major evaluation findings.

Source: Mackay (2011).

program performance (rather than bland, uncontroversial issues) in an honest and objective manner. Major disagreements concerning evaluation planning priorities would often be escalated to the ministerial level. DoF desk officers would also seek to become involved in major evaluations, usually through participation in evaluation steering committees and by commenting on draft evaluation reports. Figure 1 represents the evaluation planning and reporting relationships among the key stakeholders inside and outside government.

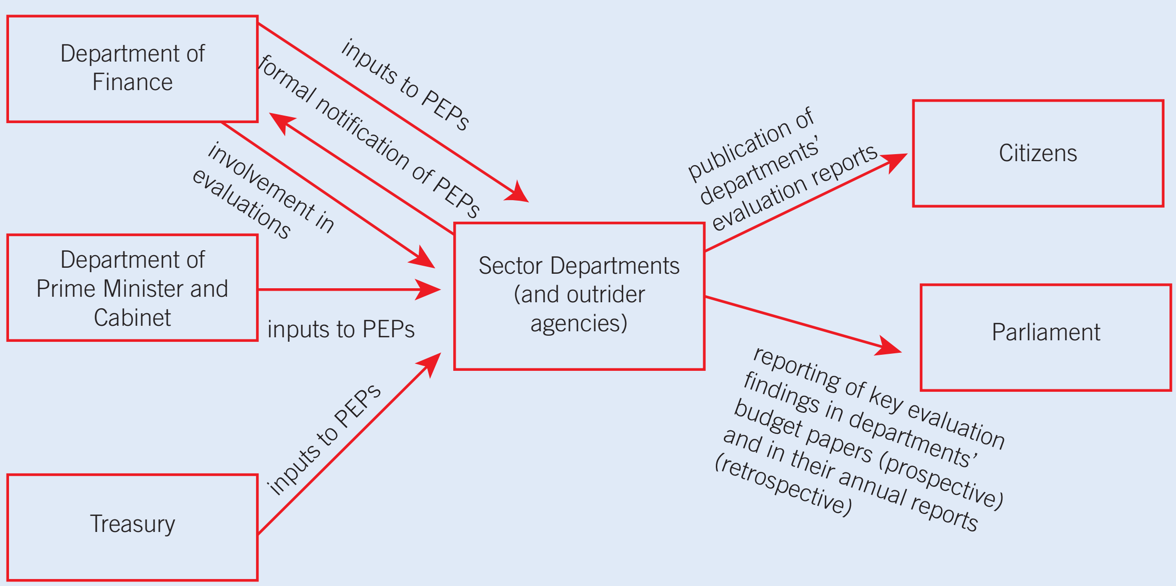
The main component of the M&E system—the evaluation strategy—lasted from 1987 to 1997, and continued to evolve over this entire period. It required effort to create the M&E system, and it was also found to be important to monitor the M&E system itself, so that DoF and other stakeholders could understand what was working well and what was not, and identify areas for improvement. A lot of trial and error was needed to progressively refine the M&E system.

By the mid-1990s, about 160 major evaluations were underway at any point in time. Evaluations were defined as "major" if they related to programs that involved large expenditures or were of major policy significance. This level of evaluation activity involved considerable effort on the part of DoF, as well as on the part of departments and agencies.

While Australia's overall M&E system stressed evaluation, it became clearer as the 1990s progressed that insufficient attention had been paid to the regular collection, reporting, and use of performance information. DoF had hoped that by stressing the planning and conduct of evaluations, which can be technically difficult, departments would be induced to improve their collection of performance information. However, DoF reviews found evidence that program performance was not being well measured nor reported using performance information. Thus in 1995, DoF secured the Cabinet's agreement to a rolling series of in-depth reviews of the objectives and key performance information for every government program. These reviews were conducted jointly by DoF and each line department. They provided the new government that was elected in 1996 some of the groundwork for shifting the emphasis to performance information.

Another initiative related to performance information was the publication, from 1995 onwards, of annual reports on service delivery by the federal, state, and territorial governments.[2] These reports included public hospitals; schools

Figure 1: Evaluation Planning and Reporting Flows



Source: Mackay (2011).
Note: PEP = portfolio evaluation plan.

and vocational training; public housing; and police, court administration, and prisons. Initially these reports covered $38 billion in annual expenditures, or about 9 percent of GDP, and their coverage has since expanded.[3] The purpose of these reports was to provide greater transparency of and accountability for government programs. In addition, it was hoped that the reports would both support and spur improved performance by enabling comparison across different jurisdictions—described as "yardstick competition"—and thus help to identify best practice.

Success Factors and Obstacles

The fact that the M&E information—especially evaluation findings—was frequently used (see below) was due to a number of success factors. The strong support of reform champions at the most senior levels of DoF, and of reformist ministers in the government, was particularly important. DoF was highly active in managing the M&E system and in both coercing and supporting departments to undertake M&E. DoF's incentives included carrots, sticks, and sermons. The carrots included the evaluation advisory support provided by DoF such as evaluation handbooks, introductory M&E training, and identification of good practice departmental M&E arrangements, as well as access to resource agreements to help line departments improve management of their underperforming programs. The sticks included DoF's ability to influence line departments' budget allocations and to escalate department disputes to the ministerial level. DoF also had the option of embarrassing departments by releasing the comparative rankings it prepared concerning departments' approaches to the planning and conduct of evaluations. The sermons included persistent advocacy by the secretary and senior executives of DoF, as well as explicit evaluation support from some powerful ministers.

Another success factor was cultural change within DoF. Its budget analysts were capable, tough minded, and very conservative. Thus it was a challenge to change their mindset from focusing on detailed line item costings to focusing on high-level policy concerned with the performance of government programs. This required strong leadership and advocacy by successive DoF secretaries. Staff turnover was also required, with more emphasis on research and policy skills and less emphasis put on accounting skills. There was also some focused recruitment, so that evaluation experience became one of the selection criteria in the annual recruitment rounds for section heads in DoF.

The availability of evaluation findings helped DoF and line departments improve the quality of their policy advice to ministers: advice became more evidence-based. It became understood that the budget process was a "marketplace of ideas," where evaluation findings could provide a competitive advantage. As the government's chief microeconomic adviser recently stated:

- Policy decisions will typically be influenced by much more than objective evidence, or rational analysis. Values, interests, personalities—in short, democracy—determine what actually happens.
- But evidence and analysis can nevertheless play a useful, even decisive, role in informing policy makers' judgments. Importantly, they can also condition the political environment in which those judgments need to be made.
- Without evidence, policy makers must fall back on intuition, ideology, or conventional wisdom—or, at best, theory alone (Banks 2009, 3).

Of course, having plentiful M&E information and having good quality policy advice are not sufficient. If governments are not influenced by this information and advice, then there might be little point in creating an M&E system. But a powerful success factor in Australia was having reformist ministers who were not only receptive to M&E information, but who actively demanded it—as seen in a number of Cabinet policy and budget debates where evaluation findings were explicitly discussed by ministers.

The M&E system was not perfect, and it ran into several obstacles. One obstacle was the lack of available training in advanced evaluation techniques, as well as a shortage of trained evaluators in departments and agencies. These entities often responded to the formal evaluation requirements by relying on their operational program managers and their staff to be responsible for conducting the evaluations; however, the staff usually possessed few evaluation skills and also lacked experience in outsourcing evaluations. On the other hand, this devolutionary approach helped ensure that the evaluations drew on the staff program expertise and that there was a high level of ownership of the evaluation findings—both of these may be difficult to achieve with externally conducted evaluations.

Some departments successfully addressed the need for more advanced skills and experience by setting up a central evaluation unit to provide advice on methodology and training in research and evaluation and participate in evaluation steering committees. Although the national audit office and DoF both advocated such central units as examples of good practice, most line departments chose not to follow this approach. Therefore the lack of evaluation skills resulted in a reduction in the quality of some evaluations—the national audit office found that over one-third of a sample of evaluation reports suffered from methodological weaknesses of one kind or another. Of course, if line departments relied on poor quality evaluations in the new policy proposals that their ministers sent to the Cabinet, DoF would certainly advise the Cabinet that the evaluation was unreliable.

There was considerable opposition on the part of line department secretaries to the creation of the evaluation strategy in 1987, mainly on the grounds that it as an intrusion into their areas of responsibility. However, once the strategy was established, there was little opposition to it during the following decade. This changed with the change in government in 1996. The new conservative government distrusted the public service, which it downsized by 20 percent, and it also wanted to significantly simplify public administration. It was at this point that line departments took the opportunity to highlight the burden of planning and conducting evaluations. These arguments immediately found a receptive audience with the new government, which had decided to devolve many central functions to line departments and agencies, thus freeing them from most central controls and requirements. Evaluation was only one of many central requirements that were abolished or significantly weakened; while evaluation was officially encouraged, the decision to conduct evaluations was essentially left to line department secretaries. The government placed much more emphasis on the collection and reporting of performance information, mainly for accountability purposes. However, there was little central oversight of this information by DoF, and no real attempt at quality control, with the result that the published performance information was severely criticized by a number of parliamentary committees, the national audit office, senior officials, academics, and others.

Other functions that were devolved included the setting of pay and conditions for public servants (individual employment contracts were also introduced) and responsibility for financial estimates. DoF was downsized considerably, most of its policy analysis areas were abolished, and its responsibility for maintaining and overseeing the financial estimates of line entities, which is a traditional, core function of any central budget agency, was removed. At the same time, the new government decided to rely heavily on sources outside the civil service for advice on policy issues, such as business consulting firms in the private sector.

Extent of Utilization of the M&E System

There is clear evidence that evaluations were used intensively in the budget process during the 1987–96 period. DoF conducted several surveys on the extent to which evaluation findings influenced budget proposals prepared by line department officials for submission to Cabinet. The survey participants were DoF officers, who carefully analyzed all new policy proposals and who typically attended all Cabinet meetings concerned with budget issues. Because of their involvement, these DoF officers had firsthand knowledge of the influence of evaluation findings on the budget proposals of line ministers and on the meetings of Cabinet ministers who made the final decisions on budget allocations. The close familiarity of DoF officers with these proposals and with any evaluations or reviews on which they were drawn, and also their participation in Cabinet's budget meetings, gave them an insider's perspective on the influence of evaluation findings.

In the 1990–91 budget, some $230 million of new policy proposals submitted by line ministers were judged to have been directly or indirectly influenced by the findings of an evaluation. By 1994–95, the last year for which estimates were available, this had risen to $2.3 billion. Measured in dollar terms, the proportion of new policy proposals influenced by evaluation findings rose from 23 to 77 percent over that period; and for most of these, the influence of evaluation was judged by DoF officers to be both direct and major. These results indicate that line department staff and ministers felt that it was important to use evaluation findings to strengthen their new policy proposals. Ministers often expressed that it was valuable to them to have evaluation findings available for their Cabinet debates. Overall, it was very helpful to have had the active support of key Cabinet ministers to encourage portfolios to plan and conduct high-quality evaluations. This support was also reflected by the many Cabinet decisions that called for evaluations of specific programs or issues.

Evaluations can have a significant influence on the "savings options"[4] put forward by DoF or by portfolios for Cabinet consideration in the budget process. In 1994–95, about $500 million of savings options—or 65 percent of the total—were influenced by evaluation findings; the influence of evaluation findings 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 that made the costs of competing policies transparent and encouraged debate and consultation among stakeholders within government.

DoF officers were also surveyed for their opinions on the extent to which evaluation findings had influenced the Cabinet's final decisions—separate from the influence of evaluation on the proposals drafted by officials and submitted by sector ministers to the Cabinet—in the 1993–94 and 1994–95 budgets. While the evidence is mixed, it indicates that evaluation findings played a substantive role. In 1994–95, evaluation findings were assessed to have influenced the Cabinet's decision in 68 percent of the $3.74 billion of proposals considered (new policy proposals plus savings options). The corresponding proportion for the 1993–94 budget, however, was only 19 percent of proposals. One important reason for this difference was the substantial revision of labor market, industry, regional, and aboriginal policies in the 1994–95 budget—the major policy review on which these decisions were based had been heavily influenced by a number of evaluations commissioned specifically to help guide the policy review.

The observation of the Auditor-General is worth noting:

> 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, 13).

It is interesting to speculate on whether the returns to M&E declined over the decade in which the evaluation strategy was in existence. When the evaluation strategy was created in 1987, there would have been a number of underperforming programs, and subsequent evaluations could be expected to have revealed their poor efficiency, effectiveness, or appropriateness; this would usually have led to the programs being improved (especially if they were an important government priority) or else cut or even abolished. But by the time that the strategy was abolished in 1997, most programs would have been evaluated more than once, and it might be expected that the marginal returns to M&E would have declined. However, the heavy reliance on M&E information in the 1994–95 budget does not lend support to this possibility. Because there is no information on the changing returns to M&E over time, this issue cannot be investigated further.

A survey conducted in 1997 by the national audit office found that evaluation findings were highly utilized by line departments for ongoing operations and internal management. The survey also found that the impact or use of evaluation findings by line departments was most significant for improvements in operational efficiency and, to a lesser extent, for resource allocation decisions and the design of service quality improvements to benefit clients. This high level of utilization was a strength of the Australian evaluation system: evaluation was essentially a collaborative effort involving DoF, other central departments, and line departments. Although responsibility for evaluation was largely devolved to line departments, the involvement of the central departments in the planning and oversight of major evaluations helped achieve broad ownership of the evaluations themselves and of their findings.

The Sustainability of the M&E System

While the future sustainability of any existing M&E system can be difficult to predict, that question has been answered for the M&E system that was created in Australia in 1987: it was abolished in 1997 following the change in government in 1996. By far the most significant component of the M&E system was evaluation, and the evaluation strategy that DoF had centrally managed was abolished. Instead, the new government created a new Outcomes and Outputs Framework that required all departments and ministers to specify the government outcomes (that is, objectives) toward which they were working, and also to specify the departmental outputs, such as service delivery for specific target groups, that would be produced to help achieve these outcomes. By 2004, departments and agencies collectively had 199 outcomes and usually between 1 and 10 outputs each. These outcomes and outputs (including their quantity, quality, and costs) were required to be reported to the Parliament in each department's annual report and in their budget documentation. However, this framework suffered from a number of fundamental, conceptual, and implementation difficulties (discussed by Mackay [2011]), including, for example:

- Poor specification of outcomes, using only broad, aspirational terms;
- Poor logical links between many outputs and outcomes;
- Lack of reporting on unmet targets, or on areas where performance was poor;
- Different specifications of outputs and outcomes by different departments, making it very difficult to make comparisons; and
- Continuing changes in definitions over time, making it very difficult to conduct time series comparisons.

This performance framework is widely judged to have been a failure, and it was itself abolished by a later government in 2009.

Although the M&E system was abolished in 1997, it had been hoped that an evaluation culture would have taken hold within the departments and agencies, and that these entities would have been more than willing to continue to conduct evaluations for their own purposes—to aid their own policy development, prioritization of activities, ongoing program management, and for accountability purposes. 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, including the departments of family and community services; employment; and health. These departments can be considered to be islands of good practice.[5] However, it seems that even these departments tended to conduct evaluation less frequently, to address only particular issues on a selective basis. It not clear whether any departments continued to conduct evaluation as regularly or as systematically as they did under the previous evaluation strategy. There has been no analysis of the reasons why these islands of good practice evaluation persisted after the abolition of the government's evaluation strategy. However, one likely reason includes the personal commitment of some key individuals in these departments—that is, champions of M&E. Another reason for these islands of good practice could be the corporate culture and mindset of professional staff in the areas of health and education, whose professional training emphasized the value of research, evaluation, monitoring, and statistics. The reality that most departments and agencies appear to engage in little evaluation activity would suggest that a wholly devolutionary approach to evaluation is insufficient—both for purposes of internal management and to support evidence-based decision making.

It is important to speculate on possible reasons why a broader evaluation culture did not persist. One 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. Another reason for the disappearance of the evaluation culture—an important risk factor—was the departure of key reform champions, and not just reformist ministers, but also the leading champions of M&E in the central departments, who were either advised to retire or were moved to less influential positions. It is well known from other countries that the existence of such champions is a key factor for the creation of a successful M&E system; however the converse also applies—their departure constitutes a serious risk for an M&E system.

The main factor that led to the abolition of the M&E system was the change in government in 1996. The new government believed the civil service to be caught up in red tape and inherently less efficient than the private sector; the government's public sector philosophy was, once again, to "let the managers manage." The new government also changed the entire policy process by: relying less on the civil service and much more on nongovernment sources of advice; substantially weakening the role of DoF, which had been the main guardian of fiscal rectitude; concentrating policy and budget decision making in the Prime Minister's Office; making many expenditure decisions after the end of the formal budget process; and basing many government decisions purely on ideological considerations, with relatively little attention paid to hard evidence such as M&E information. This approach can perhaps be regarded as the antithesis of evidence-based decision making. It was considerably facilitated by the economic good fortune that Australia enjoyed as a result of booming exports (largely due to a very strong resource sector), and especially by the continuing high levels of budget surplus. Thus, just as difficult macroeconomic circumstances can provide a powerful motivator for public sector reform and for greater effort to be devoted to getting the most value from government spending, abundant prosperity can have the opposite effect by undermining these reforms.

A new government, elected in 2007, has taken a number of initiatives that collectively are likely to increase the supply of and demand for monitoring information and evaluation findings. The initiatives include a fresh approach to monitoring that replaces the Outcomes and Outputs Framework and reintroduces program budgeting. There is also a renewed focus on evaluation and review, with substantive interest from DoF in creating a whole-of-government system that avoids the weaknesses of the earlier evaluation system. More emphasis is also being placed on developing the policy skills of the civil service in the context of government decision-making processes, which will provide greater scope for policy advice. Only time will tell if the current efforts to reinvigorate monitoring and evaluation in the Australian government are successful.

About the Author

Keith Mackay retired from the *World Bank* in 2008 and now works as a consultant. He was a *Lead Evaluation Officer* in the Bank's *Independent Evaluation Group*, providing technical advice and other support to countries working to strengthen their M&E systems as an integral part of sound governance. Before joining the Bank in 1997, he was the manager of Australia's whole-of-government evaluation system. He has written over 60 publications on M&E systems.

Acknowledgments

For their comments, the author thanks Jim Brumby (Manager, PRMPS), Philipp Krause (Consultant, PRMPR), as well as Gladys Lopez-Acevedo (Senior Economist, PRMPR), and Jaime Saavedra (Acting Sector Director PRMPR). The views expressed in this note are those of the author. To access other notes in this series, visit www.worldbank.org/poverty/nutsandbolts.

Endnotes

1. Note that this evaluation strategy related to the federal government's own spending, not to the services provided by lower levels of government.
2. www.pc.gov.au/gsp.
3. Coverage currently relates to $136 billion in annual expenditure, or 13 percent of GDP.
4. Savings options are areas of government expenditure that could be reduced or abolished entirely.
5. It is interesting to note that these were the same departments that had created specialist evaluation units during the period of the formal evaluation strategy.

Further Reading

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

Barrett, Pat. 2001. "Evaluation and Performance Auditing: Sharing the Common Ground: A Review of Developments." Address to the Australasian Evaluation Society, Canberra, October 10.

Mackay, Keith. 2007. *How to Build M&E Systems to Support Better Government.* Washington, DC: World Bank, Independent Evaluation Group.

——. 2011. "The Australian Government's Performance Framework." World Bank ECD Working Paper No. 25, Washington, DC.

This note series is intended to summarize good practices and key policy findings on PREM-related topics. The views expressed in the notes are those of the authors and do not necessarily reflect those of the World Bank. PREM*notes* are widely distributed to Bank staff and are also available on the PREM Web site (http://www.worldbank.org/prem). If you are interested in writing a PREM*note*, email your idea to Madjiguene Seck at mseck@worldbank.org. For additional copies of this PREM*note* please contact the PREM Advisory Service at x87736.

This series is for both external and internal dissemination