

# Appendix B

## Evaluation in the Australian Public Service: current state of play, some issues and future directions

An ANZSOG research paper for the Australian Public Service Review Panel

March 2019

Matthew Gray and J. Rob Bray



This research paper represents the opinion of the authors and the authors alone. It has been commissioned through ANZSOG by the Independent Review of the Australian Public Service, to inform the Review's deliberations. However, it does not represent official views of the Review Panel, of the Review Secretariat, of any member of the expert reference panels, or of ANZSOG.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise, without the prior permission of the publisher.

© 2019 The Australia and New Zealand School of Government (ANZSOG) and the authors.

[ANZSOG website](https://www.anzsog.edu.au) <<https://www.anzsog.edu.au>>

## Appendix B

# **Evaluation in the Australian Public Service: current state of play, some issues and future directions**

**An ANZSOG research paper for the Australian Public Service Review Panel**

March 2019

Matthew Gray and J. Rob Bray

ANU Centre for Social Research and Methods

Australian National University

---

## ANZOG Expert Panel

This paper was prepared with the advice of an expert panel comprising.

- Chancellor Professor Tom Calma AO, University of Canberra
- Dr Nicholas Gruen, CEO, Lateral Economics
- Robyn Kruk AO
- Professor Allan McConnell, Associate Dean Research, Faculty of Arts and Social Sciences, The University of Sydney
- Professor Patricia Rogers, Director, Evidence and Evaluation Hub, The Australia and New Zealand School of Government
- Chancellor Professor Peter Shergold AC, Western Sydney University
- Professor Helen Sullivan, Director Crawford School of Public Policy, Australian National University
- David Tune AO PSM.

Our thanks for the useful contributions made by the panel members.

We also thank the current and former senior public servants who met with us and provided us with their well-informed insights and frank views on the current state of evaluation in the APS, and their perspectives on how this could be developed.

# CONTENTS

<b>Executive summary</b> .....	<b>7</b>
<b>Introduction</b> .....	<b>8</b>
<b>The role of evaluation</b> .....	<b>9</b>
<b>Evaluation in the Commonwealth – history and current practice</b> .....	<b>11</b>
A brief history .....	11
Current practice .....	12
Current state of evaluation in the APS .....	13
<b>Selected examples of international practice</b> .....	<b>14</b>
Netherlands .....	14
United States (US).....	14
United Kingdom .....	15
Canada .....	15
<b>Challenges in framing and conducting evaluation</b> .....	<b>16</b>
The focus of evaluation .....	16
Challenges in undertaking evaluations.....	17
<b>Modes of evaluation</b> .....	<b>20</b>
Internal versus external evaluation.....	20
Trials and other forms of experimentation.....	20
Role of big and linked data.....	21
Evaluation or review .....	21
<b>The evaluation framework and the public service</b> .....	<b>23</b>
Culture .....	23
Organisation .....	24
Capacity and resources.....	25
Possible approaches .....	25
<b>Forward looking prodding questions</b> .....	<b>27</b>
Q1: How should the evaluation function and the enhancement of an evaluation culture be organised in the Australian Public Service - is there a role for greater centralisation?.....	27
Q2: Should evaluation be focused on accountability within government, or should there be a stronger focus on external accountability and transparency? .....	27
Q3: To what extent are improvements in the efficiency and effectiveness of monitoring and evaluation contingent upon reforms in investment in and access to data, including big data? .....	27

Q4: Is there too much emphasis on large-scale outcome evaluation to the detriment of more timely and operationally useful evaluation activity? ..... 28

Q5: What is the realistic scope of policy and program experimentation and how can experimental design feed into better evaluation? ..... 28

Q6: Are there other simple changes to policy and program development processes which would enhance the capacity for their effective evaluation? ..... 28

Q7: How best to use evaluation capacity and capability, and is there a need for the public service to invest? ..... 29

**References..... 30**

## EXECUTIVE SUMMARY

This paper has been prepared as input to the work of the Panel undertaking the Independent Review of the Australian Public Service (APS). It is concerned with the conduct and use of evaluation in the APS.

We find across the Australian and international literature, including a range of reviews of the activities of the APS from the 1970s on, and from our own discussions with senior public servants, a near universal endorsement of the need for evaluation, across all stages of the policy and program cycle. We find also, notwithstanding this, and the repeated calls from enquiries and reports for a need to improve the conduct and utilisation of evaluation, a general view that the state of evaluation within the APS is poor.

While there are many technical issues which are important to evaluation, including questions of program design to improve the ability for effective evaluation, the need for improved data collection and dissemination, and APS skills, in the end the principal challenge is that of the way in which evaluation activity is organised and managed. We also note a need for greater clarity in understanding the role of the APS in evaluation's accountability role, in particular the balance between internal and external accountability.

On the basis of our review of the literature and discussions with senior public servants, we consider there is a need for a stronger centralised role to identify the priorities for evaluation, to oversight evaluation activity, including rigorous reporting on this; and to provide the impetus and support to build and consolidate an evidence-based, and accountability-focused culture which values, effectively manages, and uses evaluation.

We propose seven questions for the panel which are directed at these issues:

- Q1: How should the evaluation function and the enhancement of an evaluation culture be organised in the Australian Public Service - is there a role for greater centralisation?
- Q2: Should evaluation be focused on accountability within government, or should there be a stronger focus on external accountability and transparency?
- Q3: To what extent are improvements in the efficiency and effectiveness of monitoring and evaluation contingent upon reforms in investment in and access to data, including big data?
- Q4: Is there too much emphasis on large-scale outcome evaluation to the detriment of more timely and operationally useful evaluation activity?
- Q5: What is the realistic scope of policy and program experimentation and how can experimental design feed into better evaluation?
- Q6: Are there other simple changes to policy and program development processes which would enhance the capacity for their effective evaluation?
- Q7: How best to use evaluation capacity and capability, and is there a need for the public service to invest?

## INTRODUCTION

Evaluation of government policies and programs is recognised as being critical to the efficiency and effectiveness of public policy and for accountability. This paper provides an overview of perspectives on the role and management of evaluation and its place in public sector approaches to accountability and performance management in Australia and overseas. It concludes by raising a series of questions focused on potential reforms to these processes in Australia. It was commissioned as an input into the work of the Panel undertaking the Independent Review of the Australian Public Service (APS).

This paper considers evaluation activities ranging from program monitoring, outcome evaluation, and economic analysis to policy reviews and public inquiries undertaken within the broad ambit of the public service. Parliamentary reviews and inquiries are outside the scope of this paper.

The paper has largely been prepared on the basis of a review of the literature, including policy documents and guidelines, supplemented by discussions with a number of current and former senior public servants on the current state of evaluation in the APS, and contributions from members of an expert panel.



## THE ROLE OF EVALUATION

The scope of evaluation activities can be quite broad and there are many conceptualisations of its role. A useful definition is provided by the United Kingdom (UK) Treasury Magenta Book Guidance for Evaluation (2011, 7)

*Evaluation examines the actual implementation and impacts of a policy to assess whether the anticipated effects, costs and benefits were in fact realised. Evaluation findings can identify ‘what works’, where problems arise, highlight good practice, identify unintended consequences or unanticipated results and demonstrate value for money, and hence can be fed back into the appraisal process to improve future decision-making.*

Volker (1996, 157) describes the evaluation task as involving ‘how well has the program been implemented (efficiency); to what extent has the program met its objectives (effectiveness); and how consistent are those objectives with the needs of the client group and wider government policy considerations (appropriateness)’.

Baird (1998, 10) suggests three key roles for evaluation:

- **Accountability:** making sure that public institutions, and their staff, are held accountable for their performance.
- **Allocation:** making sure that resources are allocated to those activities which contribute most effectively to achieving the basic objectives of the institution.
- **Learning:** making sure we learn from our successes and failures, to do things better in future.

Hellstern (1991, 271), focusing on the learning role, describes evaluation as having ‘become part of an institutionalized policy-learning cycle and a major intellectual activity to generate information and stimulate learning about the causal impacts of government actions and programs’. Althaus, Bridgman and Davis (2007), while recognising the limits of the formulation of the ‘policy cycle’, describe this as a process which ‘ends – and restarts – with evaluation’ (p.179). The UK Treasury (2018, 1) succinctly summarises the role of evaluation as ‘to provide objective analysis to support decision making’.

These perspectives are all relevant to the APS, although there is a debate about their relative priority.

A range of terms are used to describe different phases of evaluation. Typically three are identified:

- **Pre-emptive evaluation:** identifying the need for policy or program interventions, and evaluating the options available to address these. (This is sometimes termed *ex ante* evaluation.)
- **Formative or process evaluation:** assessing the implementation of a program or policy, variously to assess the need for fine tuning, or more significant changes, to monitor the extent to which it is implemented with regard to its intent or stated processes, or to examine the efficiency of implementation.
- **Outcome evaluation:** a focus on the impact of the policy or program. This includes the extent to which it achieved its stated aims, the degree to which it met the needs of the target group, and the incidence of any adverse outcomes. It also includes questions of cost-effectiveness, either in absolute terms of a positive benefit cost outcome, relative to alternative options, or within the parameters of the net social investment seen as appropriate.

The use of the terminology 'evaluation' across all of these phases, and covering functions otherwise described as monitoring, program auditing, and performance management, including concepts such as adaptive management, and the role of meta-analyses, can be considered appropriate. While policy formulation and program development are often not described in these terms, it generally involves evaluation of the initial evidence for intervention, and of the options available for responding. Similarly while the second phase is at times simply seen as monitoring the effective and efficient delivery of policy and programs, without the evaluative function of assessing this progress, monitoring becomes an empty exercise. Wollmann (2003, 4) notes that if one 'looks beneath the "surface" of varied terminology, it becomes apparent that the different terms "cover more or less the same grounds"'.

# EVALUATION IN THE COMMONWEALTH – HISTORY AND CURRENT PRACTICE

## A brief history

Questions around the role of public sector evaluation in Australia have been long-standing. While tracing some earlier roots Sharp (2003, 6) considers it emerged as a focus after two 'watershed enquiries' – Coombs and Baume. The Coombs' Royal Commission on Australian Government Administration (1974-76) recognising the challenges of measuring program effectiveness called for the establishment of a unit within the Department of Prime Minister and Cabinet 'to evaluate the effectiveness of ongoing programs' (1976, 385). The Senate Baume Committee (1976-1979) considered the role of evaluation in the human services seeing it at the time as a tool 'far too little applied and seldom funded' (Baume, 1979, 1) and arguing a need for greater focus so that the 'present system ... can be made more equitable, more rational, more efficient and more effective' (p.2).

The 1980s saw a series of changes to the operation of the public service directed at improving efficiency and focusing on outcomes. This was part of what is now known as the New Public Management (NPM) approach (Hood 1991, Osborne and Gaebler 1992). The key changes included the introduction of program budgeting and subsequent financial and personnel management reforms giving line agencies and program managers greater control and responsibility for program performance. As part of this there was focus on the collection of performance information and program evaluations (see Mackay 2011). In broad terms these changes have been seen as a shift in focus from 'due process' to 'results', or as put by Keating (2001, 145) a supplanting of the 'old paradigm which was largely process and rules driven with an emphasis on hierarchical decision making and control' with a 'focus on results in terms of efficiency, effectiveness, quality of service and whether the intended beneficiaries actually gain [and] a decentralised management environment which better matches authority and responsibility so that decisions on resource allocation and service delivery are made closer to the point of delivery'.

In 1987 a requirement was introduced for all new policy proposals to include a statement of objectives, identified performance measures, and evaluation plan. In 1988 Cabinet agreed to a formal strategy for evaluation. While departments remained primarily responsible for determining evaluation priorities and conducting evaluations, there were also formal requirements. These included: every program was to be evaluated every three to five years; each portfolio was to prepare an annual portfolio evaluation plan with these being submitted to the Department of Finance; new policy proposals were to include a statement of proposed future evaluation; and evaluation reports should be normally published, with budget documentation reporting major evaluation findings. The Department of Finance was also given an active role with the establishment of a branch responsible for providing advice, support and training in evaluation, as well as developing detailed evaluation handbooks (Mackay 2011).

From 1996 a series of changes were made to public sector management and a new performance framework was developed. The initial changes include: the simplification of formal requirements; deregulation and devolution of evaluation with Finance no longer being involved in quality control; and evaluations ceasing to be systematically used in the budget process. There is some evidence that these led to fewer evaluations being conducted and fewer being published. In 1999 a new 'Outcomes and Outputs Framework' was implemented, replacing program budgeting. The intention was to place more emphasis on impact and the combined effect of related activities, while also strengthening financial discipline over the price of outputs. The aggregation of programs into 'outcomes' was the subject of considerable debate over subsequent years including whether the outcomes were overly broad and not conducive to practical accountability oversight by the Parliament, and whether the connections between inputs

and outcomes were too uncertain. A series of studies including Project Sunlight (see for example Tanner 2008) and the CFAR (see below) led to a focus on a combination of outcomes and programs ('sub-outcomes') and discussion of 'intermediate outcomes' to better connect inputs to outcomes.

From 2007 there was some recentralisation of public sector management but there was no systematic change in the use of evaluation in budget processes, and no consistent approach to the publication of evaluations. There were however a significant number of policy and related reviews.

## Current practice

Two of these reviews, the Review into Australian Government Administration and the Commonwealth Financial Administration Review, led to the *Public Governance, Performance and Accountability (PGPA) Act 2013* which provides the overarching arrangements of current practice. The Explanatory Memorandum for the PGPA Bill (Wong 2013, 7–8) stated that 'The key dimensions of resource management (the cycle of planning, budgeting, implementing, evaluating and being held accountable) are not well reflected in the FMA and CAC Acts.' It went on to declare that the bill would ensure 'a clear cycle of planning, measuring, evaluating and reporting of results to the Parliament, Ministers and the public ... introducing a framework for measuring and assessing performance, including requiring effective monitoring and evaluation [and that] ... the Department of Finance and Deregulation will play a stronger role in encouraging a more systematic approach to performance monitoring and evaluation.'

The *PGPA Act 2013* emphasises the importance of performance monitoring, evaluation, whole of government coherence and cooperation and organisational capability, with the JCPAA noting (2015) that it allowed for more flexibility in how performance reporting is undertaken. One manifestation of the focus on organisational capability was a program of capability reviews of all government departments.

The *PGPA Act 2013* requires that Commonwealth entities prepare and include in their annual reports, annual performance statements reporting on:

*... the results actually achieved against the measures set out in their corporate plan and Portfolio Budget Statements, and on the basis of any review and evaluations they have committed to undertaking to demonstrate their performance in achieving their purposes. The statements will present the performance of the significant activities for which the entity is responsible at the end of each reporting period, by reporting against the targets, goals and measures that the entity established at the beginning of a reporting year.*  
(Department of Finance 2017, 9)

The *PGPA Act 2013* is currently subject to an Independent Review. The Consultation Draft concludes that 'the PGPA Act and Rule provide a sound framework for the measurement and reporting of the performance of entities across the Commonwealth, but the quality of performance reporting needs to improve.' (Alexander and Thodey 2018, 2).

Alexander and Thodey further recommend that 'the PGPA Rule should be amended to raise the minimum standard for performance reporting by including a requirement that performance information must be relevant, reliable and complete. This will require entities to improve the quality of their performance measures.' (p.2) It also recommends that 'The Secretaries Board should take initiatives to improve the quality of performance reporting, including through the greater use of evaluation, focussing on strategies to improve the way entities measure the impact of government programs' (p.2).

## Current state of evaluation in the APS

Concerns have been expressed in recent years about both the quality and extent of evaluation. Particular criticism has been forthcoming from the JCPAA and the ANAO (e.g., JCPAA 2015; Auditor General 2016a, 2016b; 2018b). The JCPAA has recommended that the Department of Finance ‘commit to an ongoing monitoring, reporting and evaluation initiative for the Commonwealth Performance Framework, performance information in Portfolio Budget Statements’ (2015, p.xi). It further reported Department of Finance evidence that while there is annual reporting of Key Performance Indicators (KPIs) it is harder to find longer-term evaluations (p.20). The Productivity Commission has also expressed concerns about a lack of an evidence base to inform key areas of policy (e.g., Productivity Commission 2016). Submissions to the Independent Review of the APS also show similar concerns (e.g., Australasian Evaluation Society, Business Council of Australia, and the Institute for Public Administration Australia).

The concerns about the quality of outcome evaluation, the extent of evaluation and whether evaluation effort was being appropriately targeted emerged very strongly in our consultations with senior current and former public servants. Indeed across most of the discussions a highly negative assessment of the current state of evaluation within the APS emerged. Common themes were that many evaluations did not add value because they were too narrowly focused and did not ask the right questions (at times further complicated by poor statements of policy and program objectives), poor methodology, limited data, and a lack of independence. Sometimes there has been a bias towards showing program ‘success’ and evaluations seem restrained by the fear of causing embarrassment to government. More generally there was a concern as to which programs and policies were evaluated, and which were not, with a strong sentiment that there should be some whole of government priorities established about the evaluations to be undertaken. There was also some concern about the adequacy of data available for evaluation and the need to establish an overall evidence base.

Although strongly supporting the conduct of robust evaluations (with some favouring a stronger focus on cost benefit analyses), the need for more timely findings to support the refinement of policies and programs was also stressed. There was a need for a stronger focus on formative evaluation, and a more supportive approach by senior public servants and Ministers. The importance of ongoing program monitoring was also emphasised. Several respondents also mentioned the particular challenges that arise when evaluating programs that were delivered through third parties, including state governments which had their own lines of accountability.

While there were differing opinions on how these limitations could be addressed, there was, on balance, strong support for greater centralisation both within Departments (with a number already having taken this approach) and across government. The three major coordinating and economic departments (Treasury Finance and Prime Minister and Cabinet) were all identified as being possible locations. While the Department of Finance was most frequently proposed as the body best suited to take on the across government coordination, it was emphasised by all that there was a need for approaches to be developed that truly fostered evaluation, rather than leading to box ticking. This also gives rise to issues of resources to be applied to evaluation and the development and nourishment of the necessary skills within the APS, including within whichever agency took on the central role.

The continued existence of these significant concerns, after many decades of APS reform, and the recurrent calls for a stronger focus on evaluation, highlights the challenges of achieving change, and a need for greater reflection on why this had failed in the past.

## SELECTED EXAMPLES OF INTERNATIONAL PRACTICE

### Netherlands

Performance based budgeting was widely adopted in the Netherlands from about 2000. This had a focus on policy outcomes and budget documentation was required on the general goals, operational objectives, and performance indicators for the program or policy. Recent developments in the Netherlands have seen a reduction in the number of performance indicators (to the most important ones), along with an increase in the level of policy evaluation and the provision of performance statements for each program in annual reports (Podger and de Jong, forthcoming). The evidence is that line departments are shifting from using performance indicators primarily for compliance assessment towards using the information to improve policies and programs. These changes have not been without criticism with the parliament expressing concerns about the omission of some outcome indicators (Podger and de Jong, forthcoming).

### United States (US)

Performance based budgeting, and the conduct of detailed program (outcome) evaluations has a long history in the US. Implementation has however been inconsistent and variable with the US Government Accountability Office (GAO) concluding that across the government as a whole 'few appear to conduct in-depth program evaluations regularly to assess their program's impact or inform policymakers how to improve results' (GAO 2011, 1). Recent decades have seen an increased focus on outcomes-based performance management and evaluation.<sup>1</sup> These include legislative changes have placed a greater emphasis on transparency, and under the Obama administration a central requirement that agencies undertake program evaluations. In the Federal government, while the evaluation function is highly decentralised, the Office of Management and Budget (OMB)<sup>2</sup> plays a key role at the Executive level including whole of government program monitoring and evaluation initiatives and in taking a leadership role, and is complemented by the GAO which serves the Legislative Branch. Initiatives under the last administration included a greater emphasis on program evaluation (outcomes evaluation), reducing administrative burden on agencies, better communication of performance results and focusing on comparative analysis of the effectiveness of programs. In November 2017 Congress passed *the Foundations for Evidence-Based Policymaking Act of 2017*, which seeks to reinforce a focus on evidence-based policy, including a requirement for agency heads to appoint Chief Evaluation Officers and submit annual plans to the OMB, and Congress. This legislation is currently before the Senate.

A feature of the US is also that there is a very active set of not-for-profit organisations, think tanks and academic groups who undertake high quality evaluations of government programs and policies, both through outsourcing, but also at times as a result of their own interests in the policy area.

---

<sup>1</sup> This section relies heavily upon Mark and Pfeiffer 2011.

<sup>2</sup> The highly decentralised nature of the US public delivery of programs and services results in a great deal of state level variation, making it difficult to provide an overall assessment of what is happening in the US.

## United Kingdom

The UK has long been a leader in embedding performance measurement and management into the public sector with virtually all public sector organisations publishing performance data (Talbot 2010). The UK has a sophisticated system of performance and evaluation measurement and monitoring, which has become increasingly outcome focused. The system has a substantial degree of centralisation with public sector organisations required to publish performance data. Although the level of outcome evaluation in the UK was assessed in 1997 as being relatively low when compared to many other OECD countries (Talbot 2010), since that time there has been an increasing focus on this form of activity including high quality resources to support evaluation. (This includes the Treasury Green and Magenta Books. Of note in the 2011 revision of the Magenta Book was a reorientation of the material to address both policy makers and evaluators – highlighting the importance of evaluation being embodied in policy decisions.) There however continues to be ongoing criticism of gaps in the coverage of evaluation evidence, poor quality evaluations, and difficulties for researchers and evaluators outside of government being able to access administrative and other government data (National Audit Office 2013).

There has also been the establishment of 'What Works Centres' which are responsible for synthesising evidence on the effectiveness of policy.

## Canada

It appears that formalised and centralised evaluation practices were introduced in Canada from 1969 and in 1977 the first government wide evaluation policy was established. This was, at least in part, based upon the idea of letting managers manage, with agencies being given greater responsibility for their programs while being held accountable for the performance of their programs. In the 1990s there was a move to performance monitoring and high-level reporting. Unlike the situation in many other OECD countries the increased focus on performance monitoring was not accompanied by a reduction in the role of evaluation with the view being that 'monitoring and evaluation are complementary, and not competing, tools to measure and manage performance' (Lahey and Nielsen 2013, 46). The 2000s saw the introduction of results-based management and the improvement of management practices.

In 2016 there were major changes made in Canada with the *Directive on Results* and the *Policy on Results* replacing several Treasury Board policies and standards. The effect of the changes seems to be a more devolved approach to evaluation with less prescriptive standards. The Treasury Board of Canada Secretariat is responsible for providing leadership for performance measurement and evaluation functions; performing and overseeing resource alignment reviews and centrally led evaluations of departmental spending; making available reports and summaries of centrally led evaluations publicly available in a timely manner; raising any compliance issues that arise; and working with departments to ensure that the data collection and reporting process is efficient.<sup>3</sup>

An interesting feature of the Canadian system is the Centre of Excellence for Evaluation which is designed to be the Canadian government's evaluation policy centre. There are some suggestions in the literature that the Canadian monitoring system is less concerned with supporting operations and more with meeting formal obligations.

<sup>3</sup> [Government of Canada Policy on Results](https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300) <<https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300>>.

## CHALLENGES IN FRAMING AND CONDUCTING EVALUATION

There are many challenges to the conduct of evaluations. This section considers a number of these broadly grouped into aspects involved in its scope and focus, and into aspects related to its roles at varying points of the policy and program cycle. Methodological aspects are considered in the section 'Modes of Evaluation'. The separate consideration of methodological aspects reflects the point made by Rossi and Freeman (1989, 29) that 'evaluation research is more than the application of methods. It is also a political and managerial activity, an input into the complex mosaic from which emerge policy decisions and allocations for the planning, design, implementation, and continuance of programs'.

### The focus of evaluation

Three critical issues concern the focus of the evaluation: whether this should be on policies as a whole, or on components including programs and projects; the focus of accountability; and the intensity of the evaluation process.

### Policies, programs and projects

Evaluation can focus on policies (including functions such as regulation), programs, or, within these, specific projects. The choice of focus involves a set of trade-offs. For example, attention on specific projects while often enabling a clear focus for an evaluative activity, can ignore wider questions of the 'network' effects of multiple projects. On the other hand it can be frustrated by seeking to disentangle the effects of multiple policy and program interventions. However, attempting an outcome evaluation across a suite of implemented measures can, as was seen with the aggregation of programs into outcomes, often be inadequate in informing about the actual causal relationships, and which specific measures actually work.

### Transparency and accountability to whom

There is a strong consensus that one of the core roles of evaluation is providing transparency and accountability. The question of accountability to whom is more controversial. Two central questions are:

- To what extent should evaluation be undertaken simply with respect to the goals and objectives formulated by government for the initiative, or to what extent should it take into account the perspectives and priorities of the groups impacted by the actions, broader stakeholders, or the community as a whole? Should the evaluation take account of the wider set of needs or circumstances, rather than those identified as being the program or policy focus?
- Should the results of evaluation be made available to decision makers at the government level to allow them to take their legitimate decisions on the policies and programs they determine to be appropriate? Or should they be more widely available to allow these decisions, and indeed these decision makers, to be critically assessed?

Although as seen above in Volker's language of 'appropriateness' with respect to the needs of client groups, and as noted later in the statement of Alexander and Thodey on accountability to citizens as well as 'Blueprint for Reform' generated by the Moran Committee (2010, ix), which proposed 'a systematic evaluation of how services are delivered from the citizen's perspective ... [and] enabling citizens to collaborate on policy and service design', these wider perspectives are less apparent in contemporary Australian practice.



One dimension of this was also highlighted in consultation where the very narrow focus of many evaluations on particular programs was seen as failing to answer broader questions as to whether the intervention was the most effective and efficient response, the potential effects of other interventions, and whether the governments' activities responded to the needs, circumstances and priorities of the target populations.

One potential reason for this discrepancy may be seen in the observation by O'Faircheallaigh, Wanna and Weller (1999, 193) who write 'evaluation is not an objectively technical, nor neutral process. It is inherently and intensely political, because it offers judgements regarding the appropriateness of utilising public resources in specific ways, of employing public servants in particular tasks, of allocating funds to programs, of providing clients and groups with public benefits.'

This observation reflects both the importance of these issues, as well as the significant questions raised as to the role of evaluation.

### **Quality versus fit for purpose, or apparently sufficient to be compliant**

Evaluation in itself can be resource intensive and this can reduce the level of resources, including time and management attention directed at the program or activity, particularly for relatively small-scale programs. As a consequence decisions about the appropriate intensity and depth of evaluation are required. Relevant factors in this decision include the novelty of the initiative, and the extent to which there may be a pre-existing knowledge base which allows the evaluation to be tailored to a more specific sub-set of concerns. Another factor is timeliness of evaluation findings. The full impacts of a policy or program can take a significant period of time (years and in some case many years) to occur. Examples include the role of innovations in education or early childhood development, on future employment, crime and other outcomes. Such timescales are however not compatible with policy cycles, let alone the need to take decisions on whether such innovations should be funded.

Other pressures include a desire of evaluators to implement 'gold standard' evaluations, driven variously by concerns that lesser quality may result in their evaluation failing to fully account for impacts, for reasons of professional prestige, or simple empire building. In contrast administrators and others may seek to minimise the scrutiny of their activities, or the demands that evaluation will place on them. This latter can lead to minimalist approaches more directed at ensuring compliance with requirements, than judgements based on the actual costs and benefits of the evaluation.

The pragmatics of decision making on this is discussed by Rossi and Freeman (1989, 40) in terms of 'in some cases ... evaluations may justifiably be undertaken that are "good enough" for answering relevant policy and program questions even though from a scientific standpoint they are not the best designs'.

## **Challenges in undertaking evaluations**

### **Policy formulation/pre-emptive**

Good policy making draws upon both theory and practice, with evaluation playing a critical role in both. Theoretical precepts from a range of fields (e.g. economics, psychology, epidemiology, public health, criminology) are all rooted in scientific methodologies which evaluate competing theories against the evidence. The evaluation of previous program and other experience is a vital contribution to knowing what might work, and what can be practically implemented in any particular society within its institutional framework.

Within the policy and program formulation stage a central challenge is that of being able to pre-emptively evaluate the best policy option. Specific resources for this include an evidence base which as well as addressing the immediate concern, and past experience, goes beyond this. Specific contributions include the role of what is often conceived of as academic or pure research which goes beyond immediately identified problems and responses,

and analysis of questions such as scalability. An important contribution to this capacity is that of policy modelling. The availability of adequate data underlies the ability to undertake these tasks. In addition to access to administrative and related data, both internally and as a resource for the 'academic' research noted above, a particular contribution is made by longitudinal data collections, both those constructed from administrative sources (e.g. Multi-Agency Data Integration Data Project (MADIP) and the Business Longitudinal Analysis Data Environment (BLADE)) and general survey-based collections (e.g., Household, Income and Labour Dynamics in Australia (HILDA) Study and the Longitudinal Study of Australian Children (LSAC)).

Critical also to subsequent phases of evaluation is the extent to which an evaluation strategy is incorporated into the decision making, including the early assignment of responsibility for/commissioning of the evaluation and the establishment of baseline data, as well as ensuring that evaluation specific, as well as program necessary, data collections are established.

### **Formative/process**

Hall (2009, 35) describes formative evaluation as answering two evaluation questions: 'How is the program being delivered and can delivery be improved; [and] How can the program be improved to achieve better outcomes?' A somewhat different emphasis is given by Rossi and Freeman (1989, 167) who pose three questions '(1) the extent to which a program is reaching the appropriate target population, (2) whether or not its delivery of services is consistent with program design specifications, and (3) what resources are being or have been expended in the conduct of the program'. The tension between these two approaches has been described by Lahey and Nielsen (2013, 49–51) as the balance in formative evaluation between learning and accountability. They suggest that within the Canadian system there was a transition in focus from learning to accountability in the 2000s. In practical terms both elements are important to effective program implementation.

While formative evaluation can take various forms, common elements include undertaking Post-Implementation Reviews (PIRs)<sup>4</sup>, and ongoing program monitoring and reporting, frequently against a set of KPIs. While KPIs, at least at a high level of program outcome, have been firmly embedded in Departmental reporting, including in budget documents, practice with regard to PIRs is less consistent. This latter can be seen in the Auditor General's Report on the cashless debit card which noted 'however, Social Services ... did not undertake a post-implementation review of the CDCT despite its own guidance and its advice to the Minister that it would do a review' (2018a, 9).

Central to undertaking this form of evaluation are information systems which accurately identify what actually is being implemented, the resources being utilised, and to the degree possible information on the outputs and outcomes of the program.

In his independent review of government program implementation, with relevance both to pre-emptive and process evaluation, Shergold (2015, vi) argues a need to move 'to proactive, performance-focused risk engagement. Too often there remains a tendency to focus on compliance ... rather than on performance ... too much focus on looking backwards, relying on evaluation and audit to identify problems after the event. There is not enough looking forward to prevent mistakes occurring'.

<sup>4</sup> In addition to the use of PIRs more generally a specific form of PIR is required where regulatory decisions have been implemented without a compliant Regulation Impact Statement. (See Department of Prime Minister and Cabinet 2016).

## Outcomes

Outcome or impact evaluations are perhaps the most commonly recognised form of evaluation, the UK Treasury describes it as ‘Impact evaluation involves an objective test of what changes have occurred, the extent of those changes, an assessment of whether they can be attributed to the intervention and comparison of the benefits to costs. It supports understanding of the intended and unintended effects of outputs, as well as how well ... objectives were achieved’ (2018, 52).

Underpinning the capacity to undertake such evaluations are:

- Policies and programs to have clear articulated goals, and a clear program logic relating the intervention to these.
- Adequate data both on the activities associated with the policy or program, and also on the outcomes of the target population, encompassing both outcomes directly related to objectives, and potentially impacted by the initiative.
- An underpinning methodological capacity directed not just at measuring impact, but also in building the counterfactual baseline against which the change is to be measured, being able to allocate outcomes against a range of interventions and in attaching value to the outcomes, including monetary value either for classical cost benefit analysis, or variants such as Social Returns on Investment (SROI).

Each of these bring with them a set of challenges. Program objectives are frequently ill-defined, expressed in terms of generalised (although often important) aspirations; or specific objectives are either expressed in a way (or are not mentioned) to make policies politically or socially palatable. The adequacy of data is frequently problematic. This arises from overall underinvestment in data collections, the extent to which program data is often limited to the needs of implementation rather than evaluation, limited or non-collection of baseline data and issues of access and use for evaluation purposes. This latter may involve complex protocols in accessing data across departments, limited accessibility of data where programs may be delivered by third parties, and problems of data provision when evaluation is conducted externally.

As noted in the discussion of formative evaluations, a critique of outcome evaluations is that findings are often too late to enable corrective action to be taken in policy and program implementation. Approaches to address this include concepts such as ‘rapid appraisal’ (Rogers and Fraser 2014, 11–13), as well as the balance of activity in the formative and outcome phases. A converse issue concerns the timescale over which some impacts may be manifest. It is increasingly recognised that some policies, such as child development interventions, have benefits which flow on through adulthood. Incorporating this impact into evaluation is problematic.

Finally, conducting such evaluations leads to the question of how the findings are used. While one element concerns the potential for changes to the policy or program, other dimensions concern the embedding of the learnings into future policies, and the role of the findings in terms of both bureaucratic and political accountability.

## MODES OF EVALUATION

This section considers several dimensions of the way in which evaluations are conducted which are pertinent to the conduct of outcome evaluation within the APS. These are whether evaluations should be conducted internally or externally; the role of trials and other forms of experimentation; the potential role of big data, and finally the question of the use of much wider policy reviews within government.

### Internal versus external evaluation

While most frequently posed in terms of whether evaluation should be conducted by the agency responsible for the implementation of a policy or program, or by an independent third party, this question in fact covers a spectrum of arrangements. These range from evaluation being conducted by those directly responsible for implementation to centralised within agencies, undertaken by a different agency within government, through to the appointment of external, usually private sector or academic evaluators. There are a spectrum of relationships from co-production to total separation between the evaluator and the agency. Each point across these arrangements has advantages and disadvantages. Advantages of undertaking an evaluation internally include potentially closer understanding of the policy/program questions and therefore a better informed evaluation and alignment of the evaluation findings with the needs of policy formulation. It also facilitates learning from the evaluation findings by those involved in developing or delivering the policy or program, including potentially building critical analysis skills, and usually facilitates data access. Disadvantages of undertaking evaluations internally can include a vested interest in positive (or in some cases negative) findings or political pressures to avoid any implied criticism of policies or political decisions; competing time pressures from operational demands; potentially less robust expertise in evaluation methods; and perceptions of a lack of transparency, and hence credibility of evaluation findings. Externally undertaken evaluations can be a mechanism for providing discipline in constructing and providing the data necessary to undertake effective and appropriate evaluations.

However issues of independence, expertise and quality are not categorically associated with internal versus external evaluation. Internal evaluations can be objective and independent, especially where they are undertaken in a supportive APS culture, while external evaluators may be influenced by factors such as wanting to keep departments and ministers satisfied so as to influence possible future contracts. These elements may also be sensitive to the extent to which evaluations are subject to wider scrutiny, including through the discipline of external review and requirements, for example, for evaluation data to be made available to allow replication and alternative analyses.

### Trials and other forms of experimentation

A key challenge for many evaluations is that of constructing a robust counterfactual against which the policy and program outcomes can be measured. In many areas of science this is addressed through experimental design. Randomisation and randomised control trials, behavioural experiments and other forms of policy experimentation can provide valuable evidence in assessing the impacts of various policy options or programs. These methods, if implemented properly, can allow robust estimates of the causal impacts of the trial or experimentation to be estimated. Randomised control trials, while widely applied in medicine, have been much less commonly used in policy development. A key reason for this is that public policy interventions are generally not administered to the target population on a random basis without regard to perceived need (Treasury (UK) 2011). In this regard randomisation can be seen as raising significant ethical questions, as well as being administratively complex. In

addition there is often a policy imperative which requires the implementation of a response, rather than the testing of what may be the best response. It can also be the case that a randomised control trial [and indeed any form of pilot project] is found to have a positive impact, but these impacts are not found following a wider roll out of the program. This can be because the trial had more than usual effort put into its implementation, but this is not scalable; substitution or displacement effects which only show up when the program is implemented on a large scale; or because the pilot was successful due to increased oversight (a Hawthorne effect) (Treasury (UK) 2011).

Other forms of randomisation and experimentation may have greater potential for application. These include phased introduction, trialling or piloting in randomly selected geographic locations or population subgroups (but not randomisation at the individual level), or the use of program variants in initial introduction, refining to an optimal model based on performance. The establishment of the Behavioural Economics Team of the Australian Government (BETA) within PMC represents a strategic shift towards greater use of these approaches in Australia. State and Territory governments have established similar types of units.

## Role of big and linked data

The advent of increased computing power and the ubiquitous online nature of life has increased the amount, and potentially the availability of administrative by product data for evaluation. This data comes from information that the government records from its interactions with citizens, people's use of online technological devices, and their interactions with the private sector means. It is often possible to link individuals across data sets. This opens up new possibilities for performance and evaluation.

Access to increasing amounts of data has a number of advantages from an evaluation perspective, including the ability to obtain data very quickly; substantially reduced costs; more precise data than is available from individuals self-reporting (not subject to recall bias or biases due to non-response to surveys); and, if effectively linked, access to multiple indicators. It can also be very valuable to the pre-emptive role of evaluation in policy development, and provide opportunities for robust longitudinal series. At the same time administrative data has limitations. It cannot tell us how people experience different events, or how they felt, and items collected are focused on what is needed for administration, not evaluation. As a consequence administrative data is often a complement for other forms of data and evidence drawn upon by evaluations.

Notwithstanding this potential there are significant challenges to the better use of this data. These include ethical and legal issues – including the extent to which departments can collect and share data, problems in gaining access, especially where data is 'owned' by different departments, or where decisions are based on risk aversion. Creating analytical data sets from administrative data systems can be complex and there is often poor documentation. In addition, where programs or policies are delivered by third parties, data may not be collected, may be collected inconsistently across providers, or be deemed to be the intellectual property of the operator.

An initiative focused on maximising the value of the Australian government's data holdings is the Data Integration Partnership for Australia which aims, amongst other things, to improve technical data infrastructure and integration capabilities and to maximise the use of these assets whilst preserving the privacy of individuals.

## Evaluation or review

As noted above, policy reviews and public inquiries were identified as part of the spectrum of evaluation activities. The relative role of program and policy evaluation and the range of forms of review, raise a series of questions.

Program monitoring and outcome evaluation usually focus on the (often formal) assessment of implementation effectiveness and outcomes achieved, measured against the objectives set for the policy or initiative being evaluated. Policy reviews and public inquiries, while generally having strong evaluative elements, are usually

broader. They tend to commence from a perspective of considering questions such as the underlying purpose or objective that the policy or program needs to address, what is the best policy or program approach, community views and a range of other issues which are frequently more constrained in a traditional evaluation. The breadth of reviews allow them to take a wider perspective on policies and programs, and interactions between these, in contrast to evaluations which most frequently focus on just one, or a narrow suite of related initiatives.

Policy reviews and public inquiries can take a wide range of forms. Policy reviews may be undertaken internally within government (e.g., Harmer Pensions Review) or largely be external experts (e.g. National Commission of Audit (2013), Ministerial Taskforce on Child Support (2004-05); Defence First Principles Review (2014-15)).<sup>5</sup> Public reviews typically 'involve external experts, undertake research and evidence-gathering activities and are 'public' with public terms of reference, published reports and processes of public engagement. They are designed to serve the purposes of government and yet they operate (to varying degrees) independently of government' (Regan and Stanton 2018, 4). These authors note that reviews can provide 'earmarked space beyond the day-to-day demands of government (and beyond the limits of departmental or institutional boundaries' (p.22). The work of the Productivity Commission can also be seen as playing an important role as a more institutional approach to the conduct of reviews independently within government.

---

<sup>5</sup> However not all such policy reviews and inquiries can be considered as being either independent or evaluative, Rather some can be considered to be more political exercises with limited evidentiary input, directed at identifying sets of policies consistent with the ideological position of an elected government, or designed in a way as to use the form of an independent inquiry to prosecute a particular policy direction.

## THE EVALUATION FRAMEWORK AND THE PUBLIC SERVICE

Although this analysis has identified significant challenges in the framing and conduct of evaluation, these, we consider, as noted by Shand (1998, 15) are secondary:

*The major issues of evaluation are management rather than methodological ones; they are about how and where evaluation should be organized, located, planned and managed to best affect decision-making. They are about managing for performance, rather than simply measuring performance.*

Three specific challenges arise from this. First is establishing the cultural change which accepts this role in decision making and accountability; secondly the organisational mechanisms required to support and drive this culture; and thirdly building the resource base to enable this to occur.

### Culture

In large part the cultural values important to the conduct of evaluation are already espoused in the APS values, in particular impartiality, professionalism (service) and accountability. However, the evidence considered here suggests that there remains considerable scope for these to be more robustly reflected in organisations. Two specific aspects are the development of a critical approach to policy and program management which includes a willingness to accept that what is currently being done may have weaknesses or failings, and a more open approach to accountability. Both of these are challenges, especially for organisations which increasingly seek to emphasise commitment to the policies and programs they are implementing.

While improving both internal and external accountability are important, the second of these perhaps raises the larger challenge. The case however for this external accountability is forcefully stated by Alexander and Thodey (2018, 10):<sup>6</sup>

*Citizens have a right to know how their money is used and what difference that is making to their community and the nation – what outcomes are being achieved, how, and at what price. Insightful performance reporting goes beyond simply measuring activities. It talks in terms of outcomes and impacts (the value created by these activities).*

This focus is also provided, and indeed expanded, by Shergold (2015, 23) in his report *Learning From Failure* with a declaration that: 'the results of policy and program evaluations should be made publicly available as a matter of course. With the huge advances in digital technology, shared information can build a stronger relationship between government and citizens. Equally important, individuals can often find new methods of using public information in unexpected ways to create public benefit.' The question of data is considered further below.

<sup>6</sup> This focus is however not new. The Baume Committee for example declared 'evaluation is also a check on the decision maker. It increases the total amount of accountability. Without it those who supply resources for programs, by either taxes or donations, cannot know whether those resources are being used effectively or efficiently. Without evaluation we cannot be sure whether those to whom we entrust much power and great resources are acting responsibly' (1979, 5).

However, as noted earlier, the strength of these sentiments raise some challenges for the APS with its focus on accountability to the Minister and responsiveness to government.<sup>7,8</sup>

## Organisation

The historical approach to the organisation of the evaluation function within the APS has been described above, along with the changes in it over time, as well as the approaches of other countries and the views of a number of commentators. While this material strongly suggests that there is no single, or simple, approach some factors can be identified:

- The existence of a centralised function across departments appears to provide an impetus and focus for a commitment to evaluation activity. The nature, role and placement of this can take a range of forms, and there are significant questions as to which is the most appropriate. These include:
  - The role of the function. Whether it is to provide leadership, support, oversight, or the imposition of formal requirements. Whether it should be involved in establishing priorities for evaluation, and indeed whether its role should extend to undertaking of evaluations;
  - The placement of the function. Should it be in a central agency, and if so which, or whether it should be placed in an independent or statutory body, or indeed should be constituted as a new body. As noted, Treasury, Finance and Prime Minister and Cabinet were all identified as being possible locations, the balance of views from our discussions tended towards the function being placed in the Department of Finance, on the other hand Gruen (2018) argues for an independent Evaluator General.<sup>9</sup>

To the extent the role is one of oversight a central challenge is ensuring that this does not develop into a ‘ticking the box’ function at the expense of true agency commitment.

- Within departments there appears to be merit in having a centralised evaluation function. This plays three roles: firstly in pooling, and more effectively building, expertise in both the conduct of evaluations and the management of externally commissioned studies within the department; secondly in both ensuring that both formative and outcome evaluations are undertaken, and that the results are held and available across the department for future pre-emptive policy formulation; and thirdly, as well as potentially enhancing the independence of the evaluation function, it allows for a concentration on the evaluation task, which is often neglected in program areas under the pressures of program implementation. Scope also exists for these areas to network across departments to share experiences, learnings and expertise.

<sup>7</sup> Hence in the APS values Accountability is couched in terms of ‘within the framework of Ministerial responsibility’ (*Public Service Act 1999* (Cth), pt 4 s10(4)).

<sup>8</sup> This question has a wider range of implications as to the role of more participatory approaches to policy formulation and program development and implementation, including questions of co-design, and in the use of concepts such as citizen juries to evaluate policy choices and the effectiveness and appropriateness of interventions.

<sup>9</sup> Under this proposal there would be a demarcation between program delivery, which would continue as usual, and the understanding of program performance. It is proposed that monitoring and evaluation would be designed and operated in the field by officers of the Evaluator General.

While the Evaluator General would cooperate with deliverers, it would report to the parliament, independently of them. Its outputs would be made available first and foremost to service deliverers to assist them optimise their performance, but ultimately to the public, together with appropriate comment and analysis. As well as promoting transparency the mechanism is seen as placing countervailing pressure on agencies to more fully embrace evidence based policy making while improving the public sector by strengthening the expertise, resourcing, independence and transparency of the evidence base on which it proceeds.



## Capacity and resources

While as noted by Shand (1998), the question of the technical capacity is secondary to the organisational challenge, it remains a challenge for the APS, in particular where emphasis is more frequently placed on management relative to technical skills. Enabling an effective evaluation function, even where this is more focused on the management of externally commissioned evaluation functions, requires staff with a suite of technical skills. These include in the area of modelling, critical to being able to undertake *ex ante* evaluation of policy options, and strong data and conceptual skills and a knowledge of evaluation approaches for the second and third stages.

In addition there is the vital role of data. As highlighted above by Shergold, this goes beyond simply providing a resource within government, to also include the wider public. Such users also include academic and other researchers. While some initiatives have been taken in this respect, including some progress in developing and in some cases implementing, more open data, there have also been significant retrenchments. Key areas for action include:

- Enabling easy and more comprehensive access to major administrative data collections (typically through accessing detailed (as opposed to summary) confidentialised unit record files); and
- Ensuring the ongoing commitment of resources to the data collections, including major economic social and health surveys, conducted both within the ABS and externally.

In both these areas a key priority is on the production of longitudinal data. As has been highlighted in the literature, the starting point for policies and programs is in the evaluation of data on both the need for interventions and the assessment of options for this. Both of these are critically based on an understanding of dynamics within the society.

## Possible approaches

From the literature and discussions with senior public servants a range of possible models and mechanisms to support and drive more effective evaluation in the APS can be identified. We note that while it is unlikely that there will ever be a single answer on how best to manage and undertake evaluation within the APS, on balance we consider that the evaluation pendulum has swung to the highly decentralised end of the spectrum, and that this is having negative consequences for evaluation, particularly outcome evaluations.

While many of the challenges for improved evaluation are organisational, an exception is that of accountability. A central issue relates to the extent evaluation is focused on reporting to government on policies and programs within the context of the goals and priorities they have set (as part of APS accountability to the elected government), or whether a wider set of accountabilities should be considered. While a broader accountability approach has generally been espoused, including in a number of recent reviews of the APS, our view is that this approach raises important questions and challenges to the APS in operating within the 'framework of Ministerial responsibility'.

Organisationally our assessment is that there is a need for a stronger centralisation of the management of evaluation across the APS. This would involve an agency taking an oversight role of evaluation, including the development of a strategic approach to ensure evaluation effort is more appropriately targeted and incorporated into policy proposals, and with departments being accountable to the agency for the evaluation they undertake and manage, including regular reporting, and where appropriate, peer review. The agency role should also include reporting on evaluation conduct and results to the government and parliament, along with the public release of evaluations. While potentially the agency could also undertake or manage evaluations, we are less convinced that this should be the usual approach. The agency would also carry a responsibility for, as discussed below, improving the quality of evaluation activity and the capacity of APS staff involved in evaluation. This function could be placed in either a central agency, or a statutory body. We note from our discussions that while Treasury, Finance and

Prime Minister were all identified as being possible locations, the balance of views tended towards Finance being the appropriate body, although some noted a need for the recruitment and development of appropriate staff to undertake this wider and more strategic policy related role.

Within this model the primary responsibility for managing evaluation (conducted internally and externally) would be with the line department. On balance we consider them to be best placed to bring a strong understanding of the policy and program to the evaluation, as well as being an interface to the provision of program specific data and contact with internal and external program managers, for the evaluation task. In our view there is a strong case for a centralisation of the evaluation function within departments to both provide some separation of the evaluation and delivery roles, and as a means of pooling expertise and building a better corporate memory of evaluation findings.

Three other foci are also seen as being critical:

- Improved APS wide guidance on undertaking evaluations. A useful point of commencement would be the development of equivalents to the UK Green Book on appraisal evaluation in central government; Magenta Book which provides guidance on designing evaluations, and Aqua Book which provides guidance on producing quality analysis (Treasury (UK) 2011, 2015 and 2018).
- Improved Analytical capability. The UK Civil Service approach to the role of "professions" in policy making may provide a useful model.<sup>10</sup>
- Data access and availability.

---

<sup>10</sup> This approach has been designed to both boost skills and to ensure that there are common cross-government standards for the recruitment and promotion. The UK Civil Service is made up of 28 professions each of which has its own competency framework. Of particular relevance here is the recently introduced Analysis Function Strategy which aims to integrate analysis into decision making and delivery

## FORWARD LOOKING PRODDING QUESTIONS

### **Q1: How should the evaluation function and the enhancement of an evaluation culture be organised in the Australian Public Service - is there a role for greater centralisation?**

Australian and international practice has changed over time, with periods of both centralisation and decentralisation, although most countries have a stronger centralised element than current Australian practice.

Is it simply a question of letting managers manage, or are there flaws in this approach, and a need for a stronger centralised function to boost accountability and promote best practice? If such a function is seen as valuable what role should it have, simply monitoring that appropriate evaluations are being conducted, providing comprehensive evaluation guidance, or should it be involved in the conduct of evaluations? If undertaking an oversight role what incentives and structures might reduce the risk of this simply becoming a bureaucratic box ticking exercise?

Are stronger mechanisms needed to determine evaluation priorities and to guide decisions on the proportionate level of evaluation for policies and programs?

### **Q2: Should evaluation be focused on accountability within government, or should there be a stronger focus on external accountability and transparency?**

As a management tool evaluation is a function which can operate within government with departmental management and executive government as clients.

Is there a need to go beyond this to accountability to parliament, and more widely to stakeholders? Should this accountability extend to questions such as the appropriateness of the policy settings for the initiative, and if so how is this to be reconciled with concepts of Ministerial accountability and responsiveness to government?

Alternatively is there a case that external accountability of evaluations should only focus on the benefit/cost of policies/programs? Would this approach permit a stronger focus on learning from the internally focused elements of the evaluation?

To what extent should the default position be that evaluations are published? Are there legitimate constraints on the timing of this? What measures might be required to ensure ease of access (and knowledge about) available evaluations?

### **Q3: To what extent are improvements in the efficiency and effectiveness of monitoring and evaluation contingent upon reforms in investment in and access to data, including big data?**

Current evaluations frequently involve convoluted processes to access limited administrative data and the commissioning of bespoke, and at times costly, data collections. Program monitoring measures are often inhibited by a lack of data on specific outcomes. Is there a need for a much more strategic approach to data across the public sector, including ensuring that appropriate data collections, including evaluation specific requirements, are in

place prior to program implementation, to address these limitations, and should this be the priority area for evaluation development?

How can the public service make access to major administrative data sets across and within departments easier and at a lower cost (in terms of time), and should this also extend to providing access by researchers external to the APS? Is there a need for a body to take responsibility for actually implementing open data access? What is the level of risk of a privacy breach that should be accepted?

What resources are required and mechanisms be put in place to make major administrative data sets, particularly linked data sets available in a more timely way?

Should there be a requirement that evaluation related surveys be deposited into a data archive in order to ensure that they are available for future evaluation and research purposes, including permitting the conduct of replication studies, and use in meta-analyses?

#### **Q4: Is there too much emphasis on large-scale outcome evaluation to the detriment of more timely and operationally useful evaluation activity?**

Large-scale in-depth evaluations are expensive and take a substantial time for evaluation findings to be reached. By the time findings are available the world has changed and they are less relevant to decision making requirements.

Should there be more emphasis given to smaller-scale evaluative activity that is undertaken continuously? If so, how should decisions be made on which policies/programs should be subject to larger-scale outcome evaluations, and when this should happen? Is there a risk that a shift to smaller-scale evaluative activity will mean that there is a loss of accountability for outcomes?

#### **Q5: What is the realistic scope of policy and program experimentation and how can experimental design feed into better evaluation?**

It is recognised that randomised control trials and similar forms of experimentation can be used as a means of testing 'what works'. Equally these can be expensive and seen as forms of social experimentation with the community, as well as not necessarily informing on questions such as scalability. Should investments in this type of approach be made and how should this be managed?

To the extent such activity is currently being undertaken in the public sector, are there effective mechanisms for drawing together information on these, and learnings from them?

Should there be a deliberate strategy of seeking to use these methods more widely? Should this rely upon ground up initiatives, or is there value in top down approaches to actually identify and provide impetus to trials?

#### **Q6: Are there other simple changes to policy and program development processes which would enhance the capacity for their effective evaluation?**

A number of recurring problems are frequently observed in evaluations, including delayed commencement, problems of data access and problems of establishing counterfactuals. Are there some simple reforms to program development and approval processes which could address these?

Some issues which have been raised Include:

- a) The scope for outcome evaluations to be commenced (including commissioned) prior to program implementation (potentially as a requirement of policy/program approval), along with an investment in the collection of baseline data and direct and early participation in data specification;
- b) This process to also consider options such as phased introduction to assist the creation of counterfactual;
- c) Ensuring all contracts between the Commonwealth and organisations that deliver services and programs on behalf of the government have adequate obligations for consistent and appropriate data collection and the provision of this for evaluation purposes; and
- d) Ensuring appropriate legal and operational frameworks which will enable data-matching for the purposes of evaluation.

### **Q7: How best to use evaluation capacity and capability, and is there a need for the public service to invest?**

Does the APS have the technical skills to undertake or outsource evaluation effectively and to learn from evaluation findings? If not how can the APS recruit, develop and maintain these skills?

- a) Within departments should the evaluation capacity be centralised, both with respect to the design conduct and management of consultations? Are the gains of having a consolidated pool of evaluation skills outweighed by the cost of separation of these staff from the policy/program coalface, and does the function, once separated from 'core' operations risk being sidelined, or be seen as easy picking for 'efficiency gains'? ;
- b) To what extent should evaluations be conducted in-house or externally contracted? Is there a greater need for the public service to evaluate the quality of external evaluations, and evaluating organisations?; and
- c) What role do the data strategies considered under Q3 provide the opportunity to build a stronger evidence base for the public service to draw upon, including in pre-emptive evaluation of policy and program options.

## REFERENCES

- Alexander, Elizabeth and David Thodey. 2018. "[Public Governance, Performance and Accountability Act 2013 and Rule – Independent Review, Consultation Draft.](https://www.finance.gov.au/sites/all/themes/pgpa_independent_review/report/PGPA_Act_and_Rule_-_Draft_Report.pdf)" Mimeo, <[https://www.finance.gov.au/sites/all/themes/pgpa\\_independent\\_review/report/PGPA\\_Act\\_and\\_Rule\\_-\\_Draft\\_Report.pdf](https://www.finance.gov.au/sites/all/themes/pgpa_independent_review/report/PGPA_Act_and_Rule_-_Draft_Report.pdf)>.
- Althaus, Catherine, Peter Bridgman and Glyn Davis. 2007. *The Australian Policy Handbook* (4th edition). Crows Nest (NSW): Allen & Unwin.
- Ames, Phil and Michael Hiscox. 2016. *Guide to Developing Behavioural Interventions for Randomised Controlled Trials: Nine Guiding Questions*. Canberra: Department of the Prime Minister and Cabinet.
- Auditor General 2016a. "Delivery and Evaluation of Grant Programmes: Department of Industry, Innovation and Science, Department of Infrastructure and Regional Development, Department of Social Services." ANAO Report No. 25, 2015-16 Performance Audit. Canberra: Commonwealth of Australia.
- Auditor General 2016b. "Corporate Planning in the Australian Public Sector: Across Entities." ANAO Report No. 6, 2016-17 Performance Audit. Canberra: Commonwealth of Australia.
- Auditor General. 2018a. "The Implementation and Performance of the Cashless Debit Card Trial." ANAO Report No. 1 2018–19. Canberra: Commonwealth of Australia.
- Auditor General 2018b. "Defence's implementation of the first principles review." ANAO Report No. 34, 2017-18, Canberra: Commonwealth of Australia.
- Baird, Mark. 1998. "The Role of Evaluation." In *Public Sector Performance – The Critical Role of Evaluation*, edited by Keith Mackay, 7-12. Washington: World Bank.
- Baume, Peter. (Chair). 1979. *Through a glass, darkly. Evaluation in Australian Health and Welfare Services*. Canberra: Australian Government Publishing Service
- Coombs, H.C. (Chair). 1976. *Report – Royal Commission on Australian Government Administration*. Canberra: Australian Government Publishing Service.
- Department of Finance. 2017. "Annual performance statements for Commonwealth entities." Resource Management Guide No. 134, Canberra: Commonwealth of Australia.
- Department of Prime Minister and Cabinet (Office of Best Practice Regulation). 2016. "[Guidance Note: Post Implementation Reviews.](https://www.pmc.gov.au/sites/default/files/publications/017_Post-implementation_reviews_2.pdf)" <[https://www.pmc.gov.au/sites/default/files/publications/017\\_Post-implementation\\_reviews\\_2.pdf](https://www.pmc.gov.au/sites/default/files/publications/017_Post-implementation_reviews_2.pdf)>.
- Government Accountability Office (GAO) 2011. *Program Evaluation: Experienced Agencies Follow a Similar Model for Prioritizing Research*. Washington: Government Printing Office.
- Gruen, Nicholas. 2018. "Nicholas Gruen: why Australia needs an evaluator-general." *The Mandarin* 9 May 2018.
- Hall, Ralph. 2009. "Evaluation principles and practice." In *Evidence for Policy and Decision-making a Practical Guide*,. edited by George Argyrous, 27–44. Sydney: UNSW Press.
- Hellstern, Gerd-Michael. 1991. "Generating Knowledge and Refining Experience: The Task of Evaluation." In *The Public Sector – Challenge for Coordination and Learning*, edited by Franz-Xaver Kaufmann, 271–307. Berlin: Walter de Gruyter.
- Hood, Christopher. 1991. "A Public Management for All Seasons". *Public Administration* 69(1): 3–19.

- Joint Committee of Public Accounts and Audit (JCPAA) 2015. Report 453 Development of the Commonwealth Performance Framework, Canberra: Commonwealth of Australia.
- Keating, Michael. 2001. "Public Management Reform and Economic and Social Development." *OECD Journal on Budgeting* 1(2): 141–212.
- Lahey, Robert and Steffen Bohni Nielsen. 2013. "Rethinking the Relationship Among Monitoring, Evaluation, and Results-Based Management: Observations From Canada." In *Performance Management and evaluation. New Directions for Evaluation*. Edited by Steffen Bohni Nielsen and David E. K. Hunter 137:45-56.
- Mackay, Keith. 2011. "The performance framework of the Australian Government, 1987 to 2011", *OECD Journal on Budgeting*. 2011(3): 1-48.
- Mark, Katharine and John R. Pfeiffer. 2011. "Monitoring and Evaluation in the United States Government: An Overview". ECD Working Paper Series No. 26, Washington: World Bank.
- Moran, Terry. (Chair). 2010. *Ahead of the Game: Blueprint for the Reform of Australian Government Administration*. Canberra: Australian Government Department of the Prime Minister and Cabinet.
- National Audit Office (UK). 2013. *Evaluation in government*. London: National Audit Office.
- O’Faircheallaigh, Ciaran, John Wanna and Patrick Weller. 1999. *Public Sector Management in Australia (Second Edition)*. South Yarra, Vic.: Macmillan Education Australia.
- Osborne, David and Ted Gaebler. 1992. *Reinventing Government – How the Entrepreneurial Spirit is Transforming the Public Sector*. New York: Addison-Wesley.
- Podger, Andrew and Maarten de Jong. Forthcoming. "The evolution of performance budgeting amidst other public financial management reforms: The experience of Australia and the Netherlands."
- Productivity Commission 2016. *National education evidence base*. (Productivity Commission Inquiry Report No. 80.) Canberra: Commonwealth of Australia.
- Regan, Sue and David Stanton. 2018. "Social policy inquiries in Australia: The Henderson Poverty Inquiry in context." TPI Working Paper No. 8/2018, Canberra: Tax-Transfer Policy Institute, Australian National University.
- Rogers, Patricia J. and Dugan I. Fraser. 2014. "Development evaluation". In *International Development: Ideas, Experience and Prospects*, edited by Bruce Currie-Alder, Ravi Kanbur, David M. Malone, and Rohinton Medhora, Chapter 9: 1–45. Oxford: Oxford University Press.
- Rossi, Peter H. and Howard E. Freeman. 1989. *Evaluation A Systematic Approach (Fourth Edition)*. Newbury Park, US: Sage.
- Shand, David 1998. "Evaluation Capacity and the Public Sector." In *Public Sector Performance – The Critical Role of Evaluation*, edited by Keith Mackay, 15–18. Washington: World Bank.
- Sharp, Colin A. 2003. "Development of Program Evaluation in Australia and the Australasian Evaluation Society – the early decades." *Evaluation Journal of Australia* 3(2): 6–15.
- Shergold, Peter. 2015. *Learning from Failure*. Canberra: Public Service Commission.
- Talbot, Colin. 2010. "Performance in Government: The Evolving System of Performance and Evaluation Measurement, Monitoring, and Management in the United Kingdom.", ECD Working Paper Series No. 24, Washington: World Bank.
- Tanner, Lindsay. 2008. "Operation Sunlight. Enhancing Budget Transparency", December 2008. Canberra: Australian Government
- Treasury (UK). 2011. *The Magenta Book – Guidance for evaluation*. London: HM Treasury.

- Treasury (UK). 2015. *The Aqua Book – Guidance on producing quality analysis for government*. London: HM Treasury.
- Treasury (UK). 2018. *The Green Book – Central Government Guidance on Appraisal and Evaluation*. London: HM Treasury.
- Volker, Derek. 1996. "Using Performance Measures to Shape Policy." In *New Ideas, Better Government (Australian Fulbright Papers: 4)*, edited by Glyn Davis and Patrick Weller, 151–163. St Leonards: Allen & Unwin.
- Wollmann, Hellmut. 2003. "Evaluation in public-sector reform: Toward a 'third wave' of evaluation?" In *Evaluation in Public-Sector Reform*, edited by Hellmut Wollmann, 1–11. Cheltenham, UK: Edward Elgar Publishing.
- Wong, Penny Y-Y. 2013. "Public Governance, Performance and Accountability Bill 2013 – Revised Explanatory Memorandum." Canberra: Senate.





Australia &  
New Zealand  
School Of  
Government

**ANZSOG**