

ANZSOG



RESEARCH
INSIGHTS

Evaluation and learning from failure and success

Australia &
New Zealand
School Of
Government

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>>

Evaluation and learning from failure and success

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

March 2019

J. Rob Bray, ANU Centre for Social Research and Methods,
Australian National University

Matthew Gray, ANU Centre for Social Research and Methods,
Australian National University

Paul 't Hart, Utrecht University, Netherlands School
of Public Administration and ANZSOG

ANZSOG Reference Panel

This paper was prepared with the advice of a reference panel comprising:

- Chancellor Professor Tom Calma AO, University of Canberra
- Dr Nicholas Gruen, CEO, Lateral Economics
- Ms 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
- Mr 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

ANZSOG Reference Panel	3
Part 1: Background and issues	5
Background.....	5
Framework for considerations	6
Tensions and impediments.....	8
Initial background papers	9
Part 2: Answering the framing questions	13
Question 1. 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?	13
Question 2. How best to use evaluation capacity and capability, and is there a need for the public service to invest?	16
Question 3. Are there other changes to policy and program development processes which would enhance the capacity for their effective evaluation?	18
Question 4. How can the APS be more mindful of and learn from its own successes? What systems or processes of 'success detection and dissection' are needed above and beyond existing performance measurement routines that do not yield the rich and reflective feedback required?	21
Question 5. How can a sustained approach to learning from successes and failures be embedded within the APS?	23
Question 6. What leadership and stewardship are needed to foster a productive learning culture in the APS?.....	23
References	25

PART 1: BACKGROUND AND ISSUES

Background

In July 2018, the Australian Public Service (APS) Review Secretariat commissioned the Australia and New Zealand School of Government (ANZSOG) to undertake three streams of research to provide input for the Panel undertaking the Independent Review of the APS ('the Review'). One of these streams concerned 'evaluation and learning from failure and success'. The commissioning process involved both a context statement and a set of framing questions.

In response to this commission, ANZSOG presented two substantive papers and a linking paper in September 2018. The substantive papers were: *Governing by looking back: learning from successes and failures* by Jo Luetjens and Paul 't Hart ('the learning from success and failure paper'), and *Evaluation in the Australian Public Service: current state of play, some issues and future directions* by Matthew Gray and Rob Bray ('the evaluation paper'). These papers both involved extensive literature reviews and were informed by consultations with experts in the field, including current and past senior public servants. After receiving these papers, the Secretariat commissioned a second phase of the project which drew together the main issues arising from the earlier papers (attached as Appendices A to C) and responded to a set of six co-designed refined framing questions, listed below.

What should the APS' evaluation practices look like for the APS to deliver world-class performance for 2030 and beyond? What does the APS need to change to get there? In particular:

1. 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?
2. How best to use evaluation capacity and capability, and is there a need for the public service to invest?
3. Are there other changes to policy and program development processes which would enhance the capacity for their effective evaluation?

How should the APS be learning from its successes and failures to deliver world-class performance for 2030 and beyond? What does the APS need to change to get there? In particular:

4. How can the APS be more mindful of and learn from its own successes? What systems or processes of 'success detection and dissection' are needed above and beyond existing performance measurement routines that do not yield the rich and reflective feedback required?
5. How can a sustained approach to learning from successes and failures be embedded within the APS?
6. What leadership and stewardship are needed to foster a productive learning culture in the APS?

This paper addresses these questions. It comprises two parts. The first reports on the current structure within which evaluation operates in the APS and presents a vision of what an end goal may be for reform. This is followed by a discussion of what we have identified as the main tensions and impediments to the achievement of this. The section concludes with a brief summary of the original background papers which are attached at the end of this paper. The second part of the paper directly responds to the six framing questions.

Framework for considerations

Context – the current situation

It is almost universally acknowledged that a systematic approach to high-quality evaluation of policies and programs embedded in APS agencies, as part of a broader approach that reflects on and learns from their own and others' successes and failures, is pivotal to producing evidence to guide appropriate policy settings and effective and efficient public sector management.

However, it is equally widely acknowledged that such an approach is not reflected in current APS practice. In submissions to the Independent Review of the APS, submissions to the review of the *Public Governance, Performance and Accountability (PGPA) Act 2013* and the review report (Alexander and Thodey 2018), and in discussions with current and past public servants, it has been observed repeatedly that the APS currently does not learn well from experience, that its approach to evaluation is piecemeal both in scope and quality, and that this diminishes accountability and is a significant barrier to evidence-based policy-making. In their day-to-day activities, departments and agencies tend to focus on achieving the 'here and now' priorities of program and policy implementation and of ministers. In examining their past performance, departments and agencies are often more concerned with reputational risk, seeking to pre-empt or divert criticism rather than learning from experience and feedback. In this context, evaluation is often also seen as yesterday's news or a second-order and lower-priority issue. Evaluation processes and findings then become just one more thing that needs to be defensively managed, or an opportunity for quick grabs to justify current and future decisions and activities.

The conclusion from our analysis of the literature, submissions to the Review and our consultations is that the heart of this problem is not a lack of skills and capacity of public servants. Rather it is a product of cultural practices that have evolved within the APS, and of the environment in which the APS operates. Within the APS, this culture is further enabled by the lack of an institutional framework that firmly embeds the strategic importance, and processes, of institutional learning.

This paper considers the reasons for this disjuncture, drawing upon submissions to the Review, our consultations with current and past senior public servants, input from the ANZSOG Expert Panel, and Australian and international practitioner and expert literature. Following this review, the paper then turns to answering the six questions identified above to aid the Review Panel in its considerations.

In considering how to improve evaluation in the APS, it is important to be aware that there are differences of view about what the goals of evaluation should be. It is common to categorise the potential goals as: efficiency; effectiveness; and the appropriateness of policies and programs. While there is general agreement that evaluation should consider both questions of efficiency and effectiveness and these are largely the focus of the existing evaluation effort within government, the third, that of appropriateness, raises more significant questions. How should appropriateness be judged? Is it just in terms of government policies and priorities? Or does it extend to the aspirations of participants and their perceptions of their needs? Or from the perspective of a wider range of stakeholders, or from broader engagement with civil society?

An end goal

What would the end point look like if reforms were instituted?

In the first instance, the APS would take a rigorous evidence-informed approach to designing and implementing policies and programs, and providing advice to government. While the guiding objectives of policies and programs need to remain firmly founded in the objectives and understandings of elected governments, the APS should seek to ensure that development, implementation and management are guided by systematic efforts to determine 'what works' and 'what is required to make it work'. These efforts should be grounded in a cumulative knowledge base about what has worked in different settings in the past, and what key drivers and preconditions are essential to

making it work. It is an approach which learns from past experiences and builds into policies and programs the mechanisms to learn for the future, and to actively reshape activity as evidence is collected and understanding emerges.

The effectiveness of programs and policies, including their implementation, would be tested against clearly articulated program objectives, and wider consequences, within a systematic evaluation framework. Much of the evaluation activity would continue to be undertaken within government departments (or using external experts commissioned by them, whose performance is tracked). This would be part of a broader cohesive approach overseen, coordinated and supported by a central unit within the APS. The framework would not just address the specific evaluation of the design, delivery, outcomes and impacts of programs, but also provide scope for periodic thematic, cross-cutting, comparative and bird's eye reviews to identify patterns, trends and critical success/failure factors. Within departments, this approach, including systematic attention to post-implementation reviews, would provide early feedback on implementation and related issues and early warning of potential problems. A clear focus on defined objectives and measuring their achievement would enhance the orientation of implementation towards effectiveness and efficiency. These activities would be understood and championed at senior levels within the department, and be seen as an integral element of senior public servants' performance.

For government, a more rigorous evaluation regime may carry some short-term cost in the form of critiques of existing activities. Indeed in public debate and some submissions to the Review, it has been suggested that one of the key impediments to effective evaluation within current structures and approaches is a reluctance of ministers to allow for the scrutiny of policies and programs due to such critiques. It however carries with it significant benefits for public policy. It provides a strong basis for decision-making on whether to maintain or modify programs, and, by being embedded in the learning framework of departments, should enhance policy advice, and program development and implementation. It would also enable stronger engagement with stakeholders, the wider community and the electorate. It would enable them to more concretely demonstrate the public value, and value for money, of government initiatives, and to explain why some activities have been terminated, or particular strategies not advanced.

Within the wider framework of departmental learning, however, evaluations are just one of a range of inputs. Organisational learning can also derive from past experiences, the experiences of other departments and of other policies, and from other countries. Drawing these together is an often neglected function, and would be supported by departments developing custodians and promulgators of these learnings. Underpinning these practices would be effective information systems, providing quality and timely data to departments and evaluators, and potentially more widely to researchers and the public. Such data should be coordinated across government to provide, where appropriate and possible, linked data which would allow for more comprehensive insights into the impact of government policies and programs, and the needs of the community.

In such a framework, policy and program failures should not be approached from a perspective of allocating blame, or successes from the perspective of reputation-enhancement and credit-claiming. Rather, both successes and failures should be seen as strengthening learning capacity, and the lessons internalised for ongoing work. Central to this culture is an acceptance that many answers are not known, especially what might work in addressing a problem. Recognising that policies or programs may not work, despite their apparent merit, or may have significant unanticipated consequences, will permit organisations to use trials and other mechanisms to test what might best work where, when and why, and thus learn what might work better in the future.

These activities would be an integral element of the APS accountability framework, including: accountability to ministers in providing them with clear and frank information on departmental activities including program outcomes, and policy advice and options formulated on the basis of the lessons learnt in past and present activities; and accountability to the Parliament through the minister, and, where appropriate, to the wider community. While the APS will continue to work in the era of the 24-hour news cycle, demands for instant grabs and a focus of 'gotcha'

reporting, this systematic approach to policy and accountability will reinforce the capacity and confidence of the APS to 'stand its ground' with authority.

Tensions and impediments

As alluded to in the context discussion above a range of tensions and impediments to such a vision exist. While some of these are potentially resolvable by the public service and government, others are not. A balance may need to be found which is not at either end of the spectrum of choices, but rather a workable middle path that seeks to balance the advantages and disadvantages of different approaches. The balance is likely to require adjustment over time – that is, a dynamic equilibrium. Four areas of tension are identified here. Two are related to public service accountability; the others involve the balance between the short and long term in organisational structures, and the tension between centralisation and decentralisation of functions.

Accountability¹

Within the tradition of executive government in Australia, government departments are responsible to their minister and through their minister to Parliament. An alternative paradigm is seen in offices such as that of the Auditor-General, which is accountable to Parliament. Yet other approaches, including some argued for in submissions to the Review and in the broader public debate, propose lines of accountability to a broader 'authorising environment' that includes societal stakeholders and/or the community as a whole. In consultations one view was that under existing accountability structures, as noted above, Ministerial commitment to particular programs and concerns about the potential consequence of evaluation findings sometimes can result in Ministers not being supportive of rigorous evaluation and the publicly release of findings.

The question for the APS, particularly with respect to evaluation, is which of these different pathways of accountability should be applied?² And, if more than one is to be adopted, can their respective responsibilities be balanced and aligned, and if so, how? We believe this to be one of the central challenges in addressing how the evaluation function can be most effectively approached.

Accountability structures and learning

A further dimension to the question of evaluation and accountability concerns how the APS can best learn from success and failure. Research has demonstrated that organisational learning is best achieved in an environment of openness and trust. In such an environment, forensic, thorough, intercollegial scrutiny of past actions and their consequences can take place without the shadow of blame and sanctions restraining the process of self-examination and lesson-drawing. It also allows and sustains 'double-loop' (reflective and adaptive) learning processes.

A tension can be seen between this kind of evaluation and the wider accountability framework that is primarily within the ambit of governments and the parliament, and potentially constrained by the role of other institutions such as the media and community pressures. There is no question that responsible government is a pivotal norm in a parliamentary system of government, and consequently public officials and organisations should be held

-
- ¹ In this paper we have used the language of accountability to describe the nature of the APS. We note that in much public policy debate the concept of 'independence' is frequently cited. In effect we consider this to be largely a question of language, in that independence is generally conceived of in terms of accountability to a concept such as 'professional standards', 'scientific neutrality' or 'the public good', or is used within the APS structure to signal being accountable to either some legislated criteria or a body such as the Parliament, rather than that to a minister and the government of a day.
- ² While the main pathway of accountability is to the Minister it is noted that there are some obligations beyond this including being subject to administrative law including FOI and ANAO scrutiny, and to the requirements of the PGPA Act (including the Performance Reports in agencies' Annual Reports). Although these can be seen as providing a 'degree of independence' to the APS at the same time the management of these in a way not to create any potential for the minister to be exposed is frequently seen as a mark of good managers.

accountable for poor decisions, systems, programs, management practices and lapses of judgement. However, the relentless emphasis on 'Gotcha!', 'No you don't' rituals of political accountability is inimical to learning from mistakes. It provokes defensive rather than open approaches from agencies and individuals. It also has a tendency to lead to scapegoating and dysfunctional forms of risk aversion.

Achieving effective organisations in the long term

The short term, the immediate and the urgent, often driven by the current media cycle and the pressures on and from ministers and their offices, frequently dominate the work of APS managers in departments and delivery agencies. This limits the capacity of APS managers to address long-term issues. It tends to discourage undertaking the reflective and developmental work that is needed to build and sustain strong cultures of institutional learning. APS managers are rewarded for being expert problem-solvers and real-time 'fixers'. In such an environment, building organisational capacity for learning and undertaking evaluations become second-order concerns, especially when evaluations often involve extensive time scales and findings are only obtained well after the immediate crisis has been resolved. As such, these functions are often seen as superfluous to immediate needs and inconsistent with lean and efficient organisational structures. The challenge is how to address these short-term pressures and reward structures to create balanced and sustainable organisations, more valuable (and widely valued) public policy, and robust accountability in the long run.

Centralisation and decentralisation

The tension between centralisation and decentralisation exists in most organisations and is a recurrent theme in public sector structures and reform proposals. Within the public service, the tension is also critically driven by the extent to which these organisations are hierarchical, reflecting the nature of existing lines of accountability. In the context of evaluation and learning from success and failure, the merits of centralisation or decentralisation should be considered at two levels. The first is whether, within departments and agencies, the evaluation function and the catalyst role for promoting learning should be centralised into a particular unit. If so, which other functions should it be associated with, and what should be the lines of reporting? The merits of such centralisation would include: highlighting the centrality of these functions across the agency; being able to identify systemic issues across the organisation and develop agency-wide understandings; and creating, maintaining and developing a critical mass of skilled staff who can focus on their evaluation role. In contrast, the decentralisation of the evaluation function allows for close alignment of evaluation and learning with the actual doing of the activity and drawing upon this expertise.

The second level is whether evaluation should just be undertaken by the line departments, or whether there is a need for a centralised function across the APS. Similar issues arise in considering the second level as for the first level. The merit of centralising evaluation functions, in addition to giving leadership and guidance on evaluation across the APS, is frequently argued in terms of being able to take a whole-of-government approach, including setting priorities and holding departments accountable for their activities. This is a major issue in later considerations.

Initial background papers

The two substantive papers prepared for the first phase of this project were both concerned with how the APS can improve its performance in policy design, implementation and service delivery by reconstituting its engagement with evaluation and accountability. Both papers identified that issues of both organisation and culture are important.

In the covering note (Attachment A), this was presented as the two papers having overlapping domains. That is, while evaluations – largely viewed in the evaluation paper (Appendix B) as significant evaluations focusing on the impact and outcomes of policies and programs, including their cost effectiveness – can be one source of learnings, they are only one of a range of mechanisms. The learning from success and failure paper (Appendix C) addresses

a wider range of review mechanisms, including Royal Commissions and Senate Committees, as well as communities of practice and learning from abroad.

The greatest contrast between the two papers relates to the role of external ‘accountability’. The evaluation paper considers the degree to which the evaluation function of the APS should be maintained within the existing government structure of departments being accountable to their minister³ or if other pathways should be considered. It identified alternative approaches to accountability, for example, through an agency directly responsible to the Parliament (the model used by the Auditor-General’s office, and one, with variation, which has been adopted in some other countries), or an even wider accountability to the community as a whole. The paper further identified more open approaches to evaluation, which could coexist with existing structures. Possible approaches included releasing all evaluations, after at most a short delay for ministerial briefing, into the public domain. More open approaches to the release of program and related data would enable others to both assess program performance and test the evaluation findings. Alternatively, the learning from success and failure paper notes the growing dominance of “practices that are driven with an accountability focus as opposed to an institutional learning focus” and argued for practices which would “institutionalise and safeguard learning-enhancing values” (p.2).

Evaluation in the Australian Public Service: Current State of Play, Some Issues and Future Directions

The background evaluation paper reviews the role of evaluation in the Commonwealth. It notes that despite strong and persistent recognition – by Parliamentary committees, the Productivity Commission, the Australian National Audit Office and various public service inquiries – of inadequacies in Australian public sector evaluations and approaches to evaluation, little appears to have been done to address this. In contrast, some other countries have taken more strategic approaches, often accompanied by a central agency responsible for either undertaking or overseeing evaluations, and have invested more heavily.

The paper notes that evaluation is typically seen as encompassing three main forms: pre-emptive or ex ante evaluation of policy and program options and the need for these; formative or process evaluation, focused on implementation; and outcome evaluation. Evaluation is identified, in practice and across the literature, as having a range of objectives. While the emphasis placed on different elements differs, common elements include: accountability; identifying what works; enabling allocative efficiency through quantification of costs and benefits (often stated as cost-effectiveness); institutional learning; identifying unintended consequences; and the effectiveness, efficiency and appropriateness of the intervention.

The paper discusses a number of challenges for evaluations, in particular emphasising the need for:

- Policies and programs to have clearly articulated goals underpinned by a coherent program logic
- Adequate and timely data, both directly on the policy/program intervention, and on the population it affects; and clear counterfactuals (including potentially those derived from randomised control and other trials) against which impacts can be assessed.

In considering these requirements, it is recognised that not all circumstances call for ‘gold standard’ evaluations, but that mechanisms are required to determine priorities for the direction of effort. In addition, as noted earlier, it needs to be recognised that a substantial amount of the evaluation effort of the APS should not just be directed at the type of outcome/impact evaluations, the context in which this issue most frequently arises, but is associated with implementation and process evaluation including program monitoring and the refinement of policies and programs as these mechanisms identify potential issues, including what is working best.

³ Within this framework the government is then accountable through the electoral cycle to the population. The APS accountability to a minister encompasses several elements, including: the department’s role in designing and implementing a policy or program; the relevant advice the department provided to the minister; and the department’s performance in identifying issues and reporting on these, and on the performance of the activity, to the minister.

Turning to the conduct of evaluation in the APS, the paper reiterates Shand's (1998, p.15) still pertinent conclusion that "the major issues of evaluation are managerial rather than methodological".

It explores this to identify questions of organisation and culture that affect evaluations. With respect to organisation it considers questions including (de)centralisation, capacity-building and resourcing, data collection, management and access. In examining the question of culture, the paper emphasises the need for the public service to recognise that policies may not always be successful and programs may have weaknesses. The paper also discusses the question of accountability. It cites the strong citizen orientation for accountability proposed by Alexander and Thodey (2018, p.10), in contrast to the more traditional political accountability framework of executive government.

In formulating probing questions, the paper raises the option of seeking greater centralisation and coordination of evaluation within the APS; questions of accountability and transparency of government policies and programs; the critical role of information systems and data; and asked about the potential for policy experimentation.

It is important to note that similar strategies to some of those presented in the background paper, and suggested here, have been used in the past in Australia. An important question is why these strategies and arrangements have not endured and why there are mixed views about how effective they were and why they have not been sustained? Several reasons have been proffered to us including: lack of support at the most senior levels in the Public Service; poor quality of evaluations by some Departments; Ministerial and other opposition to rigorous evaluation of policies and programs (particularly when criticisms are made of conventional wisdom or "preferred" policies); delays in obtaining evaluation results when policy needs are urgent and operating to a different time frame; the fact that implementation of requirements such as centralised reporting, without cultural shifts, can lead to this becoming a 'tick-a-box' bureaucratic process; and contestation between departments (particularly between central and line agencies) that can result in central agencies being less than enthusiastic about others having a strong and well-resourced evidence base.

Governing by Looking Back: Learning from Successes and Failure

The second paper identifies six key activities which need to be both funded and committed to by the public service if it is to become a learning institution that effectively identifies its own and others' successes and failures as fertile 'data points' for self-improvement, rather than as bundles of threats and opportunities that need to be managed politically. These activities are:

- Continuously self-monitor and internally discuss past practices
- Generate and process multiple feedback streams from external sources
- Compare performances across time, contexts, sectors and jurisdictions
- Maintain a culture in which it is considered safe and valuable to expose errors and engage in critical self-examination
- Engage in reflective and adaptive ('double-loop') lesson-drawing that is prepared to question key tenets of the status quo and to transform lessons into practices
- Sustain these efforts over time, even after the focus has shifted away from the original agenda.

In addressing the need for these activities, the paper reports as a central issue the "structural imbalance in how the political system and the APS are tuned to detect and attend to government 'failures' ... as opposed to 'successes'" (p.4). As noted above, the focus on failure originates from the fundamentally political and often adversarial APS environment. This typically includes the media logic that 'good news is no news', the rituals of attack and defence between opposition and government, incident-driven Royal Commissions and other reviews, and investigations which seek to allocate blame for certain outcomes. This environment leads to a public service which is more oriented towards 'stand and defend', 'duck and run' or 'weaken and exploit' than to 'reflect and improve'.

On the positive side, the paper identifies some advisory Royal Commissions and expert-driven commissioned reviews, and even certain Senate Committees (citing in particular the approach adopted by the committee on superannuation) as striking a more constructive tone, and eliciting more reflective and productive agency and ministerial responses. It notes the largely untapped learning potential derived from domestic and international professional communities of interest and international organisations. The paper identifies some opportunities for learning lessons from overseas experience, but cautions against excess enthusiasm and opportunism in doing this (noting that these lessons are sometimes sold on the basis of ideological or other premises).

Reflecting these considerations, the paper draws attention to a two-fold challenge for the APS:

1. To institutionalise and safeguard learning-enhancing values of openness, curiosity, risk absorption, self-reflection and experimentation within the APS – even as the political Zeitgeist is one of blame avoidance, reputation management and consequently defensive responses to negative feedback (p.21)
2. To counterbalance the over-emphasis on failure-finding in its authorising environment with internal routines and methodologies of success-finding (and reflective practices about learning from it) at both the agency and the systemic level (p.21).

PART 2: ANSWERING THE FRAMING QUESTIONS

Question 1. 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?

In large part, evaluation is currently highly decentralised within the APS. While in some cases evaluation plans form part of Cabinet considerations of New Policy Proposals, this is not consistently the case. Responsibility for conducting or managing externally contracted evaluations is frequently distributed across departments. At the central level, even when evaluations are undertaken, there is no standard mechanism for the findings to be considered and addressed. Nor are there strong mechanisms at the central or departmental level for monitoring the quality of evaluations or for bringing the findings together in a strategic way. In addition, there are no strong mechanisms to build a history of what has worked and what has not, or to systematically review evaluations to learn cross-cutting lessons and provide guidance to future evaluation efforts. These arrangements mean that staff with skills in evaluation frequently only use them in evaluations within the program area in which they are working, and tend to work independently of others in the APS who are engaged in evaluation.

This decentralised approach allows priorities for evaluation to be set by ministers and departments, and often also involves a close link between the evaluation and the policy or program area. However, the weight of evidence derived from our consultations and in broader debate suggests the gains from this approach are disproportionately outweighed by the downsides. These include an absence of any strategic approach to evaluation; a failure to oversee the evaluation effort and the quality of evaluation; the dispersion and frequently under-utilisation of evaluation skills, along with limited capacity to build them; and a tendency for evaluation findings to be narrowly considered and soon lost from corporate memory. In contrast, systems of public administration in many other countries have a much more centralised focus on evaluation. In many of these cases, a more centralised focus on evaluation has been developed to redress the types of problems that have been identified in Australia.

A strong case therefore exists for developing a centralised evaluation role both at an APS-wide level and within departments. An APS-wide function would enable an integrated whole-of-government approach. It would be able to set priorities for, and identify gaps in, the evaluation effort; ensure an evaluation focus in Cabinet and related decision-making; identify APS-wide learnings; and provide leadership and guidance in evaluation across the APS. Although under such a structure some evaluations may be undertaken or overseen by the central function, it would be envisaged that the majority of evaluations would continue to be the responsibility of, and undertaken or managed by, individual departments. This would allow departments to use their strong subject knowledge, close contact and understanding of the program/policy area, and recognises that much evaluation work and gain is in policy and program delivery and modification.

In some formulations, such as that the conception of an Evaluator General as proposed by Nicholas Gruen, the introduction of a centralised evaluation role, but with a focus on an informational role and the provision of leadership and support, rather than any central coordination role, is in fact seen as an instrument which would lead to a much more robust and independent level of decentralised evaluation within departments. An allied concept, which emphasises decentralisation is that of “democratic experimentalism” (Dorf and Sabel 1998). While one element of this concerns devolution to levels of government closer to the public, and an emphasis on the role of courts in the context of the American constitution, the approach has some strong focus on local solution finding and testing of approaches.

Centralised APS-wide evaluation function

This would involve an agency taking an oversight role for evaluations, including developing a strategic approach, providing advice to government, and ensuring evaluation effort is more appropriately targeted and incorporated into policy proposals. Departments would be responsible for reporting to the agency on the evaluations they undertake and manage.

A central oversight role is critical to developing a strategic high-level approach to evaluation and strengthening an evaluation culture, including drawing on this to construct a more robust evidence-based policy approach in the APS.

There are several ways in which such a central evaluation function could be established. Our consultations undertaken as part of this project, along with submissions to the Review Panel, identified diverse views and it is clear there is no consensus. The lack of consensus relates to whether, as intimated in the preamble of this paper, the evaluation function is constructed within the relatively narrow interpretation of accountability in the Westminster system (with a primacy of ministerial accountability to Parliament), or alternatively has more direct lines of accountability to the Parliament or more widely to the community. These broader approaches, as noted, are likely to result in a greater evaluation focus on the appropriateness of policies and programs. Here we note two possible approaches reflecting these differing paths. We do not cover the broader questions of the appropriate accountability paths, which form part of wider questions around the nature of the APS and of the roles of the executive and Parliament.

Centralised function within current APS structures

This would see the centralised APS-wide evaluation function constituted as a unit, within an existing department. It would reflect the traditional structure of APS administration and the nature of Cabinet and ministerial accountability. The three major coordinating and economic departments (Treasury, Finance, and Prime Minister and Cabinet) were all identified as being possible locations, in our consultations, although with differing views on which would be best suited and why. Regardless of location, the leadership of the agency will need to embrace and develop the new function, and recruit and develop appropriate staff to undertake this wider and more strategic policy-related role.

The functions and approach of the agency, which underpin the leadership role that it is to provide, would include:

- Building strong links between evaluation and major policy/program decision-making at a government level (including overseeing that appropriate evaluation planning is included within policy and programs in the Cabinet process)
- Strategically managing commitment to provide evaluation results to stakeholders and the broader public
- Providing a whole-of-government focus on key evaluation priorities, including identifying gaps and other limitations within department activity
- Regular publication of a 'state of evaluation' report which identifies the strategies and priorities of evaluation and communicates key evaluation findings across the APS
- Ensuring that impact evaluations are publicly released by default, and that only in exceptional circumstances (e.g., those that have national security implications) are they not released
- Providing guidance and support for departmental central evaluation functions,⁴ including research into, and support for, implementing leading edge evaluation methodologies (including, where appropriate, policy experiments and trials), and for the professional development of the staff in these areas

⁴ There are several examples of this type of guidance. One example, and potentially a useful point to start, 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). Another interesting example is the recent changes in the United States which require

- Where determined by Cabinet, undertaking and managing specific evaluations.

To be effective and to achieve the functions outlined above, the central unit would require a degree of independence, potentially having direct responsibility to the relevant agency head.⁵

An independent evaluation function

An alternative approach is to establish an independent evaluation function as an independent statutory agency accountable to Parliament. Under this model, the functions would be similar to those for a central unit within an existing APS central agency, but with some differences. The independent function would:

- Have less engagement in the Cabinet process
- Have a stronger focus on the conduct of evaluations, at least of major policies and programs, including a mandate to undertake evaluations where deemed to be appropriate, but with more limited scope for interaction with departments' current evaluation activities
- Have a stronger focus on the dissemination of evaluation findings including its direct role in reporting to parliament and, dependent upon the model chosen, to external stakeholders.

While an independent agency should also seek to take a leadership role in promoting evaluation within the public service there are mixed views as to the extent this would occur. One perspective is that it would lead to a stronger evaluation culture and enhance the capacity and autonomy of departments as they will be able to proceed knowing that they have a strong evidence base for their policies and programs. Alternatively, it has been suggested that it is likely to be perceived as being outside of government, and relationships are likely to be more arm's length.

Commentary

We would consider both of these approaches to be feasible, with the Auditor-General being a model that could be considered for the independent function. Each approach has strengths and weaknesses. It is our view that the internal approach would be a stronger mechanism for building an evaluation culture within the APS, although potentially with a primary focus on efficiency and effectiveness within the program or policy framework. In contrast, the second approach may be more effective in increasing the external focus on the evaluation of some government programs and extend the ambit of evaluation to encompass more aspects such as appropriateness.

A complementary mechanism to the internal approach would be for the Auditor General to also take a stronger role with respect to the reviewing of agencies' performing reporting arrangements under the PGPA Act to have a specific focus on evaluation, in terms of the extent and priorities of evaluation being undertaken, and how this is reported on, and incorporated, in terms of performance reporting.

Within department centralisation

Regardless of the model chosen, as has been noted, it is probable that a significant level of evaluation activity will continue to be undertaken within departments. We consider there to be a strong case (as detailed in the background paper and summarised earlier) for this type of centralisation. It will assist departments to develop their evaluation capacity, including through a critical mass of skilled staff, and to provide a stronger focus on, and

each Federal department to have a Chief Evaluation Officer to coordinate evidence-building activities and an official with statistical expertise to advise on statistical policy, techniques and procedures ([H.R.4174 – Foundations for Evidence-Based Policymaking Act of 2007](https://www.congress.gov/bill/115th-congress/house-bill/4174/text) <https://www.congress.gov/bill/115th-congress/house-bill/4174/text>). An example from the Canadian system is the [Centre of Excellence for Evaluation](https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation.html) <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation.html> which is designed to be the Canadian government's evaluation policy centre).

⁵ Reflecting the ideas advanced in some submissions, such a role could be cast as the Australian Government Evaluator, or Evaluator General. Regardless of the nomenclature, it is envisaged that the position would need to be senior with a high profile both within and outside of the SES, and, as noted, a degree of independence, even if constituted within existing APS structures.

leadership in, departmental learning across the breadth of the portfolio activities.⁶ This includes the scope to identify systemic issues across the department.

Question 2. How best to use evaluation capacity and capability, and is there a need for the public service to invest?

The question of resource utilisation and investment has several dimensions. These are considered below with respect to: APS skills; the adequacy of resourcing for evaluation, including the balance of evaluation effort; investment in data; and the best use of external evaluation skills.

APS skills

As already noted, our research and consultations led us to conclude that the major impediments to evaluation capacity and capability in the APS are cultural and organisational rather than a lack of skills or analytical capability.⁷ For this reason, we see the major response to improve use of these skills is to implement the evaluation approach outlined in this paper.

In the first instance, creating a central APS function will provide the capacity for recruiting staff with strong evaluation skills from across the service, and provide them with an environment in which they can further develop these skills. Similarly, the centralisation of the evaluation function within departments provides the opportunity to develop a critical mass of evaluators, and enhance the capacity for skill development and training (e.g., in cutting-edge techniques) and the training of the next generation of evaluators.⁸ However, issues of workforce capability are secondary to organisational and cultural issues. Developing and valuing the evaluation function within the APS, including through the creation of career paths, will enhance the incentives for departments and individuals to invest in skills development.⁹ More generally, beyond those working on evaluation, attention also need to be given to the wider development of the evaluation literacy of public servants across the APS, so they can more effectively use the products of evaluations, and indeed the development of data and related interpretation skills to best use these sources in policy and program development.

Resource adequacy

Within the time constraints of this project we have been unable to rigorously address the question of allocating resources to the evaluation function. This question has two key elements: the total quantum of resources allocated to evaluation; and the balance of current resource application.

With respect to the level of resources currently allocated for evaluation, the limited level of evaluation currently being performed in the APS can be interpreted as evidence of this being insufficient. Such a perspective was

⁶ Some submissions proposed this type of unit to be headed at a senior SES level by a Chief Evaluation Officer, in a similar role and with similar status to Chief Information Officers and similar posts. As discussed later with respect to the second set of questions, the learning role of such a unit creates the potential for combining this role with the responsibility for promoting a learning culture.

⁷ We recognise that different assessments of this capability exist; for example, some submissions to the Review highlighted issues of 'deskilling'. Our view is that while the question of further skills development is important, as is the question of ensuring appropriate levels of resourcing, discussed further below, that neither of these acts in their own right, in contrast to the issues of organisational structure and culture, as binding constraints on the capacity of the APS to undertake effective evaluation.

⁸ While the university sector already provides opportunities for formal learning of evaluation and related skills, these mechanisms may also provide enhanced opportunities for the APS to better articulate its training and education needs. This could include scope for co-design of more tailored approaches to meet APS needs.

⁹ One strategy is that embodied in the UK Civil Service approach to the role of 'professions' in policy-making as a means of improving analytical capability, including specifically those capabilities related to evaluation. This approach, in addition to having a role in boosting skills, also seeks to ensure that there are common cross-government standards for recruitment and promotion. The UK Civil Service is made up of 28 professions, each of which has its own competency framework.

strongly argued in a number of submissions to the Review, and was also quite widely reflected in our consultations with senior public servants.¹⁰ Taking these perspectives together we consider that there is a strong prima facie case that inadequate resources are currently being committed to evaluation across the APS.

One approach to addressing this (regardless of how the structural reforms discussed in Question 1 are implemented) is through a requirement that all policy and program proposals include an evaluation plan and, where appropriate, specify the resources for this. However, this is only a partial response, recognising that many policies and programs are effectively ongoing and changes in these are often undertaken within existing funding.

The second key element relates to the balance of current evaluation activity. Several questions can be asked about this. These include the balance between smaller-scale evaluative activity on a continuous basis, and large-scale in-depth implementation and impact evaluations that are expensive and take a substantial time before evaluation findings are reached. Another is whether the current program of evaluation reflects the relative priorities for evaluation, including the extent to which new, as opposed to continuing, programs are the primary focus.

Both of these elements were raised as an issue in our consultations, particularly by current APS leadership. The proposal for a centralised Commonwealth function with a focus on developing a systematic and priority approach to evaluation will help to resolve this.

Investment in data

The importance of ongoing investment in data cannot be overemphasised. Existing investment includes that through the Australian Bureau of Statistics, the funding of surveys such as the Household, Income and Labour Dynamics in Australia (HILDA) Survey, and the development of administrative datasets, including the Multi-Agency Data Integration Project (MADIP). Data from these sources provides an essential tool for evaluation. It is used in both evaluation of policy options and in outcome evaluations. It is often extensively drawn upon to identify counterfactuals, baseline data, and the nature and characteristics of the target population, as well as information about the broader environment. Investment in these activities needs to be maintained, if not enhanced. As discussed in Question 3, a related question concerns the need for program and other management information systems to be developed with a cognisance of their importance in evaluations. An important pre-condition for the more effective use of administrative datasets is having appropriate legal and operational frameworks in place. This not only allows easy provision of data for evaluators,¹¹ but will also enable data-matching for the purposes of evaluation. Using data well in evaluations will also help justify the quality of data that is available for future evaluations.

With many government services and programs delivered by for-profit and community organisations on behalf of the Commonwealth, all contracts between the Commonwealth and these organisations should have adequate obligations for consistent and appropriate data collection and providing this data for evaluation purposes.

The effective use of external evaluators

A large amount of APS evaluation activity is undertaken by external evaluators. From consultations, the literature and our own experience as evaluators, our assessment is that there are advantages and disadvantages of both using external evaluators and in-house evaluation. However, these largely relate to the nature of specific evaluations, the need to mobilise specialist skills, or to coordinate data collections or diverse teams, and are less to do with questions of quality and independence.

¹⁰ Within the scope of this project it was not possible to undertake any systematic inquiry into the actual level of evaluation activity, the adequacy of funding, or the gaps and the cost of addressing these. An alternative approach would be to benchmark Australian activity against other countries, especially those seen as being best performers. Again though, this is a far from trivial activity, even for the first step of quantifying existing spending.

¹¹ One case in point is the extent to which departments claim ownership of particular datasets.

It is clear that there is scope for more effective management of external evaluators. This includes addressing the tendency of some departments to micromanage the activities, and a lack of appreciation of how a department's internal processes, including issues such as data access, can significantly impede the efficient conduct of the evaluation. More generally, the APS does not appear to have good processes in place to consistently evaluate the quality of externally conducted work or take performance into account when awarding future contracts.

Question 3. Are there other changes to policy and program development processes which would enhance the capacity for their effective evaluation?

From our consultations, submissions to the Review and our experience as evaluators, we identified a number of changes which could enhance the capacity for successful evaluation of policies and programs. Here we discuss five aspects that we consider to be priorities:

- Integrating the development of evaluation strategies into the policy/program development process including funding approval
- The role of mechanisms that will allow for strong counterfactuals, including the use of randomised control trials and other forms of experimentation
- The need to consider evaluation in the long term (including after the cessation of programs)
- Paying attention to the evaluation of existing and ongoing activities
- The integration of evaluation findings into both internal and external program reporting and performance measurement.

New policy and program development, approval and implementation

The capacity for effective evaluation of policies and programs is greatly enhanced if evaluations are built into the policy development process. This integration has two elements. The first is that the mechanism for implementation and process evaluation, and the refinement of programs and policies based on the findings from these are built into the program. The second is that the necessary steps to enable outcome and impact evaluations are embedded early in the process. While such integration needs to commence earlier in the policy development process, the logical place for this to be formalised is in New Policy Proposals. All New Policy Proposals should be required to include:

- A clear program logic and the full range of specific outcomes and impacts sought from the policy/program (including where appropriate numeric targets). These outcomes and impacts should also include the range of aspirational objectives articulated in the rationale for the policy or program. New Policy Proposals should also include a description of the evidence base upon which the program logic is based. This not only provides a rigour to the program logic model, but aids evaluation by providing the theoretical and empirical rationale for how it is expected the policy or program will achieve its objectives.
- A clear commitment to conducting Post Implementation Reviews (usually but not always done within departments), including the timing and accountability framework for these.
- The proposed evaluation strategy and associated funding for evaluation. If no formal evaluation is proposed, a rationale should be provided. (This should also include the timetable for the implementation, in particular taking account of the discussion below on baseline data.)

In addition:

- The development of any information systems for a new policy or program needs to include specifications for providing data for evaluation purposes. These specifications need to be included in the evaluation strategy.
- To the extent possible, the program or policy implementation should be phased to allow for the evaluation to start (including the commissioning process if external evaluators are being used) before program

implementation. Program and policy implementation should also allow for investment in the collection of baseline data, where appropriate, and the direct and early participation of the evaluators in data/systems specification.

Constructing counterfactuals and innovative approaches to identifying what works

A key challenge for many evaluations is constructing a robust counterfactual against which the policy and program outcomes can be measured. It is common for policies or programs to affect the focal group as a whole. Hence, it is not easy to identify the specific effect of the program on this population, particularly if the group may be affected by multiple interventions, or there may be other changes in the environment. In classical evaluation terms, there is a need for a control or comparison group.

In addition, as identified in the background and issues section of this paper it needs to be recognised that in public policy the potential efficacy of interventions is often uncertain.

Taking these two issues together generates a strong case for considering mechanisms which allow for the construction of robust counterfactuals, or allow for policy or programs to be tested in various ways before being fully implemented, or both. Several mechanisms exist for such approaches, including:

- *Pilot programs.* This strategy is quite commonly applied in Australia, and involves the introduction of the policy or program in a limited way, typically to a specific population subgroup or location. It provides the capacity for those populations/locations not selected to be used as a control. Unfortunately, selection of the population or location is frequently based on criteria such as high needs. This reduces the ability to use populations or locations that do not meet these criteria as a control. Another limitation is that pilot programs may be a 'foot in-the-door' way to get a program introduced and not as a genuine pilot.
- *Staged introduction.* This is less frequently used, but is a useful technique when it is anticipated that much of the effect will be in the short term. Areas where the policy was initially introduced can be compared with those yet to be phased in. This type of approach is less applicable when the program impact is likely to be achieved over longer periods.
- *Randomised control trials.* This approach, often proposed as the gold standard for measuring the effectiveness of an intervention, involves the random allocation of individuals (or potentially other units such as communities) into either a treatment group, which is exposed to the policy or program, and a control group which is not. The difference in outcomes between the two groups is taken as a measure of the program impact.
- *Experimental approaches such as rapid prototyping.* As identified earlier approaches such as "democratic experimentalism" may also provide scope for experimental approaches, driven by local assessment of circumstances and needs. These types of strategies can involve risk-embracing approaches that allow potential policy and program options to be developed and tested in a time- and cost-effective way. Such approaches have been characterised as 'failing fast, failing cheap, failing forward'. In contrast to the above approaches, which have a strong emphasis on establishing counterfactuals, experimental approaches are more concerned with identifying whether a particular approach may have an impact, and what issues may arise in implementation. Evaluation findings are frequently derived from qualitative rather than quantitative methods. A focus of the methodology is using the findings from the prototyping to test different approaches, including potentially 'high-risk' strategies, within the policy development process, and providing an evidence base to guide the optimal approach.

Each of these approaches has positive and negative aspects, and limitations as to when they are appropriate or practical. From an evaluation/program development perspective, a key issue is that they generally do not address the question of scalability. In addition, the community may see them as forms of social experimentation, or deem them to be inequitable because some people are subject to constraints, or alternatively offered some benefit, which others are not. While it can be argued that this form of experimentation is appropriate and non-discriminatory, especially if there is some uncertainty about the effectiveness of the measure, this argument can be hard to

sustain. When there is uncertainty about the effectiveness of some policies or programs, one inhibition within the APS may relate to the higher risk of such trials proving to be unsuccessful, and the associated potential for negative blame-placing.

Notwithstanding these limitations, there appears to be strong potential for a greater application of trials and other approaches within the APS. We would envisage that if a central APS evaluation function is established, one of its roles would be to work with departments to identify where these approaches could be adopted, and provide support in their design, development, implementation and evaluation.

Evaluating programs in the long term

Many government interventions, including in health, education, social services, and employment and training, may have effects that only manifest over extended time periods. Effects usually extend well beyond the usual time scales of evaluations, and indeed may often occur well after the program has been terminated.¹² It is important that a framework be developed which allows this to be considered.

Elements of such a framework include:

- Ensuring records on interventions, including details of the individuals affected, are collected and held, along with (where appropriate) permissions for the use of these records and for data-matching
- Regularly reviewing the opportunities for using administrative data-matching as a means of establishing longitudinal records that may provide information on outcomes over time
- Considering the need for establishing specific panel surveys to follow participants over time. (Several existing studies contain elements of this approach, for example, the Longitudinal Study of Australian Children, the Longitudinal Survey of Australian Youth, and the Beginning a New Life in Australia study of humanitarian migrants to Australia)
- Continuing to support data collection, such as the HILDA survey, which allows analysis of the long-term effects of a diverse range of policies and programs, in particular those which may affect a significant group of the population. These data collections, along with the development and release of administrative data on programs provides the opportunity for evaluation activity to also be undertaken by community groups, academics and others. It would also provide the opportunity for the development of partnerships between the APS and these groups to undertake evaluation and related activities.
- Regularly considering the potential to support specific evaluations that can take a long-term backward look at the impact of policies and programs, even after they are no longer in operation. (In some program areas, there may also be a wider public policy interest in this type of work. Possible options include making data available to academics and others with a research interest in the field.)

Again, we see the establishment of a centralised evaluation function within the APS as one way to provide the impetus for these kinds of strategies, including identifying priorities and opportunities for undertaking particular evaluations, and guiding the development of linked datasets that may permit this type of work to be undertaken in a cost-effective manner.

Evaluating existing policies and programs

While systematic planning of evaluation before introducing programs and policies is highly desirable, focusing only on new policy and programs is not sufficient. A clear strategy for reviewing and evaluating existing and ongoing policies and programs is required.

Evaluating existing policies and programs is necessary because:

¹² Frequently cited examples are the Perry Pre-School and Abecedarian programs in the US where long-term evaluation has identified significant positive returns, well into adulthood, from early childhood intervention.

- A very high proportion of government spending is on an ongoing basis
- The nature and outcomes of these programs may have evolved or drifted away from their initial rationale or purpose
- Such long-term programs have the potential to become embedded or institutionalised by the participants or community in ways which may have significantly affected their outcomes.

The evaluation approach to be adopted will depend on the nature of the program or policy. Potential approaches range from operational efficiency reviews, re-establishment and analysis of program logics and how these are supported by the current evidence base, through to policy reviews (underpinned by the systematic analysis of the program and its impacts).

An initial starting point would be to examine the potential of approaches such as the comprehensive re-establishment of portfolio evaluation plans, which identify the extent to which existing programs have been subject to evaluation and the proposed timing of future activity.

Evaluation and internal and external program reporting and performance measurement

Closely associated with the policy and program development process are processes for program reporting and performance measurement.

The production of key performance indicators (KPIs) has become an entrenched element of APS reporting over several decades. Significant effort has been applied to establish indicators that contain a strong focus on the outcomes achieved. Notwithstanding this, across most departments, these indicators give at best only a limited insight into the efficiency and effectiveness of a program, and are frequently reported on across an aggregation of different initiatives. This reflects their role as being a series of 'key' or headline indicators.

Given this, there is considerable scope for evaluation activity to complement the production of KPIs, and provide a potential evidence base for revising existing reporting frameworks. Information from recent evaluations could be incorporated into the performance indicators, for example, information relating to cost-effectiveness and the identification of outcomes (both outcomes relative to program objectives and 'unexpected consequences' of programs).

Within the broader reporting framework, a clear priority is for evaluation findings to be reported in the public domain. As discussed earlier, this could occur through the rapid publication of all evaluation reports (along with additional material such more accessible summary reports and access to data), as well as mechanisms such as reporting evaluation activity and outcomes in departmental annual reports.

There is also an important role for the internal consideration of evaluation findings. Findings can be used as a tool for building learning organisations, and building and reinforcing robust corporate memories that can draw on lessons of the past in current and future policy and program development. In this latter regard, attention needs to be paid not just to the findings of outcome evaluations, but also to process evaluations, including elements such as Program Implementation Reviews. These strategies will also tend to complement the relatively narrow focus of existing performance indicators.

Question 4. How can the APS be more mindful of and learn from its own successes? What systems or processes of 'success detection and dissection' are needed above and beyond existing performance measurement routines that do not yield the rich and reflective feedback required?

The learning from successes and failures paper argues that there is a structural imbalance in how the political system and the APS detect and attend to government failures as opposed to its successes. Potential alleged and

apparent failures are the focus of much attention, accountability processes and 'blame games'. In contrast, high-performing programs and practices are mostly taken for granted and therefore go unnoticed and underappreciated (the occasional granting of awards and commendations notwithstanding).

This imbalance manifests itself in APS accountability mechanisms and consequently in the APS's engagement with feedback. Neither is conducive to a productive learning culture within the APS. The often highly publicised and politicised framing and processing of failures breed a fundamentally defensive propensity to minimise political and reputational risk. This crowds out any inclination to treat failures as data points. The underreporting and non-acknowledgement of successes leads to them being overlooked as potential levers for continuous improvement and lesson-drawing.

To redress this imbalance and more effectively mine the learning potential of both successes and failures, the APS could adopt a series of practices focused on detecting and exploiting these lessons. These practices should be seen as complementary to a more robust generic evaluation regime as presented in our recommendations in Questions 1–3. Successes and failures represent the outer edges of the government's performance spectrum. They therefore offer particularly salient and urgent data points and cause for reflection above and beyond feedback from regular evaluation cycles.

Specifically, there is a suite of complementary options that the Review Panel could consider:

- The building up of *critical-incident/near-miss reporting systems*, particularly within delivery and regulatory agencies. These can be modelled on good practices currently extant within, among others, the process industries, the aviation sector and hospitals.
- Adopting *'whole system in the room' debriefs*. These are carefully prepared and facilitated Chatham House rule exercises where critical cases (near misses, explicit failures, ongoing or ad hoc instances of high performance) are reconstructed and reflected upon, drawing on the perspectives of designers, (co-)producers, deliverers, and targets/recipients of policies and programs. The focus lies on what may be learned from the experience, by whom and how this learning can be actioned.
- *'Learning from our stakeholders'* exercises. These can take the shape of focus group and/or fishbowl sessions. In these sessions, clients, stakeholders (including otherwise 'soft voices' in the sector), and independent experts of policies and programs are explicitly encouraged to articulate their experience of tensions, disappointments and frustrations, as well as of highly positive, constructive and impressive performances of the agency in question. Agency representatives observe but do not speak, let alone defend, during these sessions. The feedback obtained from these sessions can be compiled, analysed and used to craft (ad hoc) unit- or program-level and (annual) agency-wide 'Learning from our Stakeholders' reports. These reports can be used to feed into strategy, innovation, design and improvement processes within the agency (inspired by, for example, Bason, 2017).
- *Expanding and better using existing routines of recognising professional achievement*. Awards events and competitions could be organised in a tiered, multilevel fashion (from branch to agency to APS systematic level). They could be designed and leveraged not just to put a positive spotlight on certain high-performing and dedicated individuals and teams, but to generate a series of standardised case narratives describing the nature, operative mechanisms and boundary conditions of successful performance. Such leveraging can take several forms, such as:
 - staging annual agency-level Learning Festivals open to all staff, where success cases are presented, subjected to 'critical friend' scrutiny and form the basis of 'lesson-drawing' workshops;
 - presenting award-winning cases from across the APS at dedicated APSC-run Learning Conferences and/or IPAA/ANZSOG annual conferences open to the public;

- widely disseminating conference proceedings featuring both individual case histories as well as comparative, thematic and lesson-drawing reflections by commissioned observers, across and beyond the APS.¹³
- *Training evaluators across the APS in the methodology and tools of positive policy evaluation*, as well as encouraging their use by other review and accountability bodies such as the Auditor-General, the Ombudsman, and (through inculcation in the Ministerial and Parliamentary Services division) the Senate and House. These tools include: Appreciative Inquiry, the Success Case method, the Most Significant Change method, tracking and analysing instances of Positive Deviance, and Developmental Evaluation strategies.¹⁴

Question 5. How can a sustained approach to learning from successes and failures be embedded within the APS?

Echoing the need for a balanced approach between centralised and agency-level organisation of the evaluation function, a similar approach to organising the closely related function of learning from successes and failures could be considered. These two functions can be organised within the general model for organising evaluation within the APS that was presented above (under Question 1). However, given the close link between the evaluation system and the requirements of ministerial accountability, there is considerable merit in organising the explicitly and exclusively learning-oriented practices, which are even more internally focused, in a separate structure.

This alternative scenario might take the form of building up a system of Chief Learning Officers (CLOs). It is important to note that the role of CLO has existed in the corporate sector since US CEO Jack Welch called it into being in the mid-1990s. CLOs can also be found in some public agencies around the world (though more sparingly). However, this role has traditionally been designed as the HR-focused responsibility of encouraging the organisation to take a strategic approach to identifying and meeting the learning and development needs of individual employees and teams. The CLO function envisaged in our paper is broader and has a primary focus on stewardship of organisational (i.e., institutional) learning.

At the level of individual departments and agencies, the role could be allocated to a Deputy Secretary, or in smaller agencies First Assistant Secretary. It may seem tempting to allocate the role to executives already tasked with corporate roles, but there are grounds for caution here. It is important that salience and ownership of learning activities is felt among those responsible for day-to-day policy design, advice to ministers and operations. Rotating the CLO role across the senior management team may also be considered. CLOs would be supported by a small permanent staff and operational budget. They would be responsible for undertaking and embedding the agency's chosen repertoire of practices for detecting and learning from both success and failure (a menu of options to this effect was presented above at Question 4).

Stewardship of this APS-wide system of CLOs, and encouraging the cross-agency, cross-sectoral flow of lesson-generating and lesson-drawing exercises at the central level, could be situated with the Deputy Public Service Commissioner.

Question 6. What leadership and stewardship are needed to foster a productive learning culture in the APS?

Public service systems operate within a political environment. The public service is subject to ministerial direction and integrally included in the systems of political accountability. It also faces multiple other forms and forums of accountability in serving its political masters while also serving its clients and the public interest at large. In the

¹³ This practice is derived from the European Public Sector Awards experience, run bi-annually in tiered fashion at national and European level. See: [EPSA – European Public Sector Awards](http://www.epsa-projects.eu/index.php/EPSA-Projects.eu) <www.epsa-projects.eu/index.php/EPSA-Projects.eu>.

¹⁴ As described and referenced in the learning from success and failure paper (at Appendix B).

Westminster system, the rationale and modus operandi of the public service are traditionally entirely contingent upon the principle of ministerial responsibility – although with partial exceptions in the shape of statutory authorities. Moreover, the public service is expected to be nearly completely and increasingly proactively transparent about its structures, processes, outputs and outcomes. Public service performance is monitored and scrutinised relentlessly by a wide variety of observers and stakeholders, many of which have political agendas in doing so.

In such an environment it is easier said than done to establish and maintain a strong culture of evaluation and learning. It is one thing to set up structures, assign roles, allocate some funding and undertake some training along the lines of what has been recommended in this paper. It is quite another to make sure these investments pay off by actively contributing towards more rigorous ('evidence-based') and more reflective practices of policy design, policy advice, implementation, delivery and oversight. This is particularly so when these practices are up against the array of mechanisms that favour risk avoidance and reactive problem-solving.

Beyond what the Review Panel chooses to recommend in this regard, truly strategic and concerted *leadership* by individual secretaries, as well as the Secretaries Board, is needed to embrace the ideal of the APS as a learning culture, and then effectively advocate for the necessary structures and roles, and appropriation of the requisite resources.

Perhaps most importantly, once a more robust system of evaluation and learning is up and running, it will require active *stewardship* to conserve and protect the integrity of such learning cultures whenever the forces of political expediency and short-termism threaten to undermine them. It needs to become both normal and safe within the APS to forensically – methodically, dispassionately – take stock and 'look back' at how and how well policies, programs and projects are performing; to actively seek out voices from clients, stakeholders and critics; to ask hard questions about what is valuable and what is not; and to re-examine beliefs and assumptions on which policy decisions were made and programs were designed in light of the subsequent experiences after they were put into practice.

To exercise the requisite leadership and stewardship is both necessary and delicate. At times, it will require careful balancing of responsiveness to ministers with guardianship of institutional capacity. Ideally, these two imperatives can be suitably aligned; the crucial tests come when there is the next efficiency round, the next machinery of government change, or the next imperious ministerial office threatening to cut or sideline evaluation capacity or infect learning processes with political danger. It would help greatly, however, if the will and courage required to exercise leadership and stewardship are uniformly and broadly distributed across the SES, particularly at Band 2/3 levels. At these critical moments, leaders will need to find it within themselves to hold firm against political pressures and protect the capacity of the public service to do its job in the best possible way. Having colleagues and others within their agencies seeing them doing that, and doing that well, provides probably the best impetus for the cause of evaluation and learning in the Australian public service.

REFERENCES

- Alexander, Elizabeth and David Thodey. 2018. *[Independent Review Into the operation of the Public Governance, Performance and Accountability Act 2013 and Rule](https://www.finance.gov.au/sites/all/themes/pgpa_independent_review/report/PGPA_Independent_Review_-_Final_Report.pdf)*. Canberra: Department of Finance, <https://www.finance.gov.au/sites/all/themes/pgpa_independent_review/report/PGPA_Independent_Review_-_Final_Report.pdf>.
- Bason, Christian. 2017. *Leading Public Design: Human Centered Governance*, Bristol: Policy Press.
- Dorf, Michael, and Charles Sabel. 1998. "A Constitution of Democratic Experimentalism". *Columbia Law Review* 98(2): 267-473.
- Gruen, Nicholas. 2018. "Nicholas Gruen: why Australia needs an evaluator-general." *The Mandarin* 9 May 2018.
- 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.
- 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.



Australia &
New Zealand
School Of
Government

ANZSOG