ACADEMY OF THE SOCIAL SCIENCES IN AUSTRALIA

POLICY ROUNDTABLE

ADVANCING POLICY AND PROGRAM EVALUATION IN AUSTRALIAN GOVERNMENT

FRIDAY 21 MARCH 2025

In person: Treasury Building
Langton Crescent, Parkes, Canberra

ADVANCING POLICY AND PROGRAM EVALUATION IN AUSTRALIAN GOVERNMENT

Evaluation offers a range of tools to understand and make judgements about the performance of government policy and programs, which can help support accountability, manage risks, and drive improvement at all stages of development and implementation.

In March 2025, the Academy of Social Sciences in Australia (the Academy) and the Australian Centre for Evaluation (ACE) jointly convened a policy roundtable to explore how evaluation can support better policy and program design and implementation in a way that is ethical, robust and fit-for-purpose, as well as the capabilities and relationships needed to successfully do so.

The half-day event, held in Canberra, brought Australian and international research experts together with Australian Government officials and evaluation practitioners to consider how to improve evaluation research and practice to support better outcomes for the public. Presentations were followed by a facilitated discussion that tested the immediate actions government, academia and practice can take to support Australia's evaluation culture.

This summary provides an overview of the discussion, which took place under the Chatham House Rule.

Opening remarks

Opening remarks were provided by Professor Kate Darian-Smith FASSA, Academy President, Eleanor Williams, ACE Managing Director, and The Hon. Dr Andrew Leigh MP, FASSA, Assistant Minister for Competition, Charities, Treasury and Employment.

ACE was established in 2023 as the Australian Government's centralised evaluation function to drive cross-agency and cross-portfolio evaluation

activity and to build capacity in policy and program evaluation across the Australian Public Service (APS). ACE encourages both the creation and distillation of high-quality evaluation evidence, and the development of systems that support the use of and access to this evidence. In practice, this involves:

- Providing leadership and policy guidance for the APS.
- Working with partners to deliver high quality impact evaluations.
- Supporting agencies to plan and use evaluations.
- Building evaluation capability across the APS.

There is a long history of evaluation practice in Australia and a rich body of academic and practice knowledge about the challenges of doing evaluations. Government-led evaluations are just one part of a much broader ecosystem that shapes the use, culture and messaging about evaluations. The opening speakers agreed that drawing on this body of knowledge is imperative if ACE is to help realise the Australian Government's goal of measuring what matters and understanding what works for more effective, efficient policies and programs that improve the lives of Australians.

Session 1: The evaluation landscape in Australian Government

Session one focused on the existing evaluation landscape in Australia: how evaluations are understood by those who commission, conduct and use them, the recent trends in evaluation approaches, and the challenges that are distinct to Australian Government policy processes and decision-making.

Recent trends in evaluation include a focus on data sovereignty, linked data and the growth of co-designed evaluations, which recognise that government is only one of several stakeholders and evidence end-users. These trends were explored across a range of policy contexts.

The session included presentations from Professor Rosalie Viney FASSA, Mr Selwyn Button and Professor Ilan Katz.

In healthcare, the purpose of evaluation ranges from establishing causality and identifying unanticipated consequences to determining value for money. In pharmaceuticals, for example, the precise clinical efficacy of an intervention is generally established in a controlled setting. In the broader context of government policy, the purpose of the evaluation is to determine if public benefit can be achieved in the real world – with different populations, health contexts and complex regulatory and policy settings – and at a reasonable cost.

Evaluation practices in education were discussed. Much of the evidence in education has been developed without a practitioner lens and there are few researcher-practitioners. As a result, the practical implications of randomised control trials (RCTs) or meta-analyses of academic research may fail to reach or resonate with teachers and their day-to-day practice. Instead, teachers and school leaders are guided through informal networks and a 'if it works for them, I'll try it too' approach. It was agreed that researcher-practitioners must become an established feature of the educational environment, as is the culture in healthcare.

The evaluation challenges related to Closing the Gap were also discussed. Over the life of the commitments, there has not been an established evaluation strategy for determining what has and has not worked to deliver outcomes. Instead, there has been a focus on contractual compliance and counting activities and outputs, rather than the quality and impact of services and, importantly, if communities judge them to be fit-for-purpose. The burden of counting and reporting often falls to service providers and does little to ensure quality service delivery and continual improvement.

There is also a privileging of government data over community-controlled data, which generates data gaps. The subsequent partial picture of progress can have unintended consequences. For example, participants heard that in health settings Aboriginal and Torres Strait Islander mothers and young babies presenting in hospitals have ended up in the child protection system because a record of previous care in community-controlled clinics was not available. The speakers agreed that addressing this challenge is not a matter of taking data and evaluation capabilities out of the community-controlled sector and centralising it within government. As the Productivity Commission Review of the National Agreement on Closing the Gap makes clear, it is a matter of investing in data linkage, building community-controlled capability alongside that of government, and ensuring evaluation frameworks are inclusive of different kinds of evidence.

The speakers discussed the challenges often faced in policy and program evaluation, including rigour, data availability and insufficient sample sizes, and unclear objectives. They agreed that it is difficult to withdraw policies and programs once implemented. For example, a lifesaving medicine that requires much greater investment than originally planned may not be easily withdrawn. To protect against this, participants stressed the importance of a clear program logic, maintaining the rigour of processes and committing to the questions asked, even if they reveal things that are not welcomed or expected.

"The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."
– Donald T. Campbell

Particular attention was given to timing and design – with evaluations often being sought too late in the decision-making process – and the 'evaluability' of policies that are rolled out without consideration of evaluation, necessitating weaker retrospective evaluation design.

The session concluded with participants agreeing that evaluation cannot be a 'nice to have'; it must be built into policy and program design and implementation processes and evaluation work must be published to ensure a continuous stream of evidence is available when called on. One option explored was legislating evaluation and publication, as was the case with the US Welfare-to-Work Grants Program. However, some cautioned against only legislating experimental and quasi-experimental evaluation approaches.

Session 2: Fit-for-purpose evaluation: Established and emerging tools

Session two explored what fit-for-purpose evaluation means for the different contexts in which evaluation takes place, the different stages of policy and program design and implementation, and for different end-users. A range of approaches and feasibility and ethical considerations were discussed against the backdrop of the Commonwealth Evaluation Policy and its key principles. Presentations were provided by Associate Professor Jenny Povey, Emeritus Professor Elliot Stern and Mr Andrew Hawkins.

There are many things policy makers wish to know: what is the nature of a problem, how might it best be addressed, what are the contextual factors, what are the potential implications, what are the likely risks, costs and benefits, and how might these vary in different places or with different communities or populations. Evaluation can help answer many of these questions, but not always prospectively, quickly, or with high levels of confidence.

The impacts and outcomes of an intervention often change over time, sometimes in unanticipated ways. For example, an intervention that is not immediately successful can produce downstream benefits. In this context, quasi-experimental methods are vulnerable to the quality and availability of data, while datasets that do exist have often been constructed for a specific purpose or to favour a particular population group.

A common theme across the three presentations was that which evaluation approach should be considered 'fit-for-purpose' is not static, it emerges and evolves through conversations between evaluators and commissioners (and in certain contexts, those being evaluated) as they seek to answer:

1. What can we know?
2. What do we already know?
3. What do we want to know?
4. What can we afford to know (the value of finding out should not be worth more than the information to be gained)?

These questions help match methods to purpose and explain the nuances of evaluation depending on a policy or program's maturity.

ACE has a particular focus on RCTs and quasi-experimental approaches to testing policy and program impact and effectiveness. When ethical, feasible and well-designed, RCTs are an appropriate method for evaluating performance against a predetermined set of outcomes to support policy and investment decisions. The discussion explored the particular points for consideration to ensure RCTs are fit-for-purpose. In complex circumstances, for instance, it is equally important to measure why a program or policy works, for whom and in what context, as well as the complementary process evaluation required to understand the generalisability of results. The results of an evaluation that fails to account for these points may produce results with implications for program funding decisions and unintended consequences for

vulnerable groups. In complex circumstances, it is important to first establish if an RCT is feasible and what other methods and existing data can support evaluation. Here the speakers reiterated the importance of publishing evaluation work. In recognition of the value of qualitive components, ACE conducts all RCTs as mixed-method evaluations. The strength of this approach can be enhanced by the addition of multi-arm trials and stratified data around cohorts and locations to explore the range of factors shaping outcomes.

Recent improvements to Australia's linked data assets, such as the Person-Level Integrated Data Asset (PLIDA) and Business Longitudinal Analysis Data Environment (BLADE), have helped address some of these challenges. Linked data provides rich insights into populations and their interactions with services and can help inform policy and program design decisions. Improved data assets can reduce evaluation costs, support mixed and multi-method evaluations, and enable new questions to be answered about the long-term impacts of well-established or discontinued policies and programs.

International perspectives on evaluation were also discussed. In Australia, there is often a narrow focus on the policy settings where evaluation first became impactful, like healthcare and broad social services. Elsewhere, governments and other actors are exploring how interventions can take a place-based approach to target complex and interconnected factors, like entrenched disadvantage, energy transitions, labour market participation, and public health and behaviour change. Participants heard that this difference was in part because of certain cultural norms and the level of acceptance of taking on more complex evaluations. Technical capability and data availability – with a focus in Australia on individuals rather than systems – and siloed, project-based budgets were also identified as contributing factors.

The challenge ahead for Australia is that evaluations will need to deal with the increasingly complex issues facing government – which have high levels of data uncertainty – and this must be done in the face of resource constraints.

There has been much work over many years on place-based evaluations and participants discussed the governance structures, relationships, and information aggregation methods needed. Despite work to date, there is still much to be done to reach consensus on how best to capture the outcomes and impacts of place-based interventions within the current policy settings and, moreover, the system-level changes needed to implement and evaluate complex, cross-cutting policies and programs. Place-based approaches are an explicit priority and focus of the current government, and participants were generally confident that this authorising environment will help build the culture and capabilities for place-based evaluation.

Returning to a common theme throughout the roundtable, speakers and participants agreed that relational, collaborative approaches would help the Australian Government respond to the complex challenges it faces. They agreed that evaluation can help 'steer the ship of policy', yet to do so, evaluators must first be invited into the design process.

The strengths of United Kingdom (UK) and European evaluation practice was discussed, such as multi-method, multi-disciplinary teams that bring together researchers, evaluators, decision-makers and other content specialists. These teams work together to establish shared program logics and bring to the evaluation a range of skills, knowledge and resources otherwise not available. A longstanding feature of the UK evidence ecosystem is the What Works Network, which participants agreed is a useful model for brokering relationships for fit-for-purpose evaluations from which Australia can learn. What Works is made up of nine full and three affiliate member centres across a range of policy areas. Each centre works closely with government to tailor their research and outputs to the needs of decision-makers and to forward a broader agenda that situates evidence and evaluation alongside the other conventions of working in government.

Session 3: Realising the potential of data and research for evaluation

The final session examined the necessary investments in capabilities, workforce, infrastructure, and relationships to strengthen and sustain an evaluation culture in Australian Government. Presentations were made by Scientia Professor Carla Treloar FASSA, Dr Phillip Gould, and Professor Deborah Cobb-Clark FASSA.

The session opened with a discussion on co-design and the powered dynamics of evidence collection and analysis. Evaluations, particularly large evaluations for government, have historically dealt with normative evidence that is not particularly concerned with the power dynamics implicit in the collection and use of this evidence. Participants heard that evaluations in the healthcare context, for instance, have not tended to consult with or record what is important to the patient about their outcomes. It was posited that co-design can help evaluators make decisions during the program logic phase about what data to collect and what questions to ask from the perspective of those being evaluated. This can provide for more comprehensive and ethical evaluations, as well as pathways to draw on community-controlled data and other expertise. The suggestion was made to include co-design principles and requirements to assess power dynamics in the Commonwealth Evaluation Toolkit. The Commonwealth Charter of Partnerships and Engagement could support and guide good practice in this regard.

The ABS and other Australian Government agencies are working to build high-quality, large-scale, integrated data assets and broker access to enable deeper insights. These assets are actively supporting better policy and program evaluations and innovative research. However, there are challenges that remain. These assets are not immune to gaps, measurement errors or selection bias, often because they have not been constructed with evaluation and research in mind. Additionally, the technical capacity to meet

data access requests is lacking across the APS and a cost recovery model has been adopted for some data services. Protracted approval processes and cost limit what research and evaluation can be done, and is particularly prohibitive for early and mid-career researchers who may lack institutional support. An enduring public value question for the Australian Government is how to translate the research and evaluation reports that use these assets into something understandable and useful for other end-users.

It was agreed that sustainable, long-term funding is needed to meet growing demand and ensure access to researchers at all levels and projects of all sizes. As well as funding models, the various relational aspects and pathways needed to better harness Australia's data assets were discussed. The speakers and participants encouraged deeper research-APS partnerships to draw together diverse expertise, help identify and fill data gaps, and develop the capabilities to use assets. Such collaboration can help foster shared buy-in into projects and make researchers more invaluable to the APS and policy formation.

Participants also reflected on the role of the ACE and Evaluation Profession to build better relationships, and there was interest in transactional versus relational contracting, including more flexible procurement that allows evaluators and commissioners to connect to discuss data and program logic before commissioning.

Concluding remarks and next steps

While the roundtable established that much of the necessary evaluation ecosystem exists, collective effort is needed to strengthen it, align different elements and, critically, build the institutional capacity in government for evaluation and evaluation partnerships. Australia's five Learned Academies bring together the nation's leading experts across the full breadth of research disciplines, able to provide trusted, independent

advice to inform policy and develop innovative solutions. It was agreed that there is a role for all Academies to support ACE to improve the quality and use of policy and program evaluation to deliver better outcomes for the Australian public.

The roundtable aimed to generate a shared agenda for government, academia and practice to improve the planning, conduct and use of evaluation. Four key areas for action were identified:

1. **Policy content** and how the subject area expertise of Australia's five Learned Academies can be brought into policy and program evaluation.
2. **Methodology** and how the Academies can assist ACE to build capabilities and deeply embed approaches that are ethical, robust, and fit-for-purpose.
3. **Data and the resources, relationships and approaches** needed for consistency across agencies in terms of access, utility, quality, and democratisation.
4. **Research and evaluation** are not the same thing, and differentiating between the two and embedding them at different points in the policy process ensures they are leveraged in different ways.

The Academy and ACE will continue to collaborate to engage with Australia's other Learned Academies and pursue opportunities to positively shape the environment for evaluation. This could take the form of working with the Evaluation Profession to better understand its needs and connect it to cross-Academies expertise, or a series of smaller, more focussed workshops jointly led by the Academies and ACE that target discrete themes emerging from this roundtable.

PARTICIPANTS

Ms Danielle Aeukens	Branch Manager, Performance and Evaluation | **Department of Social Services**
Mr Scott Bayley	Managing Director | **Scott Bayley Evaluation Services**
Mr Selwyn Button	Commissioner | **Productivity Commission**
Professor Deborah Cobb-Clark AO, FASSA	Deputy Head (Research) School of Economics and Deputy Director Life Course Centre | **University of Sydney**
Dr Honae Cuffe	Policy Director | **Academy of the Social Sciences in Australia**
Mr Bruce Cunningham	Assistant Secretary, Employment Evaluation | **Department of Employment and Workplace Relations**
Professor Kate Darian-Smith FASSA	President | **Academy of Social Sciences in Australia**
Professor Karen Fisher FASSA	Social Policy Research Centre | **University of New South Wales Sydney**
Dr Phillip Gould	Deputy Australian Statistician, Data & Statistical Practices Group | **Australian Bureau of Statistics**
Mr Harry Greenwell	Director, Impact Evaluation Unit, Australian Centre for Evaluation | **Treasury**
Professor Jane Hall AO, FASSA, FAHMS	Centre for Health Economics Research and Evaluation | **University of Technology Sydney**
Professor Brian Head FASSA	Faculty of Humanities, Arts and Social Sciences | **University of Queensland**
Dr Chris Hatherly	CEO | **Academy of the Social Sciences in Australia**
Mr Andrew Hawkins	Partner and Chief Evaluator | **ARTD Consultants**
Mr David Kalisch FASSA	Faculty of Business, Government and Law | **University of Canberra**
Professor Ilan Katz	Social Policy Research Centre | **University of New South Wales Sydney**
Dr Susie Kluth	Assistant Secretary, Analysis, Data and Measurement | **Department of Education**

PARTICIPANTS

The Hon Dr Andrew Leigh MP FASSA	Assistant Minister for Competition, Charities, and Treasury and Assistant Minister for Employment
Associate Professor Jenny Povey	Deputy Director (Training) and Principal Research Fellow, Institute for Social Science Research | **University of Queensland**
Emeritus Professor Elliot Stern	Emeritus Professor of Evaluation Research | **Lancaster University**
Scientia Professor Carla Treloar FASSA	Centre for Social Research in Health | **University of New South Wales Sydney**
Professor Rosalie Viney FASSA	Director, Centre for Health and Economics Research and Evaluation | **University of Technology Sydney**
Professor Mark Western FASSA	Research Director, The Queensland Commitment | **University of Queensland**
Ms Eleanor Williams	Managing Director, Australian Centre for Evaluation | **Treasury**