



Australian Government
Productivity Commission

A Guide to Evaluation under the Indigenous Evaluation Strategy

October 2020

The Productivity Commission acknowledges the Traditional Owners of Country throughout Australia and their continuing connection to land, waters and community. We pay our respects to their Cultures, Country and Elders past and present.



The artwork used in this publication is adapted from *River of Knowledge* by Luke Penrith

© Commonwealth of Australia 2020

Except for the Commonwealth Coat of Arms and content supplied by third parties, this copyright work is licensed under a Creative Commons Attribution 3.0 Australia licence. To view a copy of this licence, visit creativecommons.org/licenses/by/3.0/au. In essence, you are free to copy, communicate and adapt the work, as long as you attribute the work to the Productivity Commission (but not in any way that suggests the Commission endorses you or your use) and abide by the other licence terms.

Use of the Commonwealth Coat of Arms

Terms of use for the Coat of Arms are available from the Department of the Prime Minister and Cabinet's website: www.pmc.gov.au/government/commonwealth-coat-arms

Third party copyright

Wherever a third party holds copyright in this material, the copyright remains with that party. Their permission may be required to use the material, please contact them directly.

Attribution

This work should be attributed as follows, Source: Productivity Commission, *A Guide to Evaluation under the Indigenous Evaluation Strategy*

If you have adapted, modified or transformed this work in anyway, please use the following, Source: based on Productivity Commission data, *A Guide to Evaluation under the Indigenous Evaluation Strategy*

Publications enquiries

Media, Publications and Web, phone: (03) 9653 2244 or email: mpw@pc.gov.au

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au).

Contents

A Guide to Evaluation under the Indigenous Evaluation Strategy	2
1 Building evaluation into policy and program design	5
2 Deciding what policies and programs to evaluate	12
3 Evaluation planning, design and conduct	16
4 Reporting and using evaluation findings	36
5 Building capability and a culture of evaluation	40
References	43
About the artist	44

A Guide to Evaluation under the Indigenous Evaluation Strategy

The Indigenous Evaluation Strategy (the Strategy) provides a whole-of-government framework for Australian Government agencies to use when selecting, planning, conducting and using evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people. The Strategy puts Aboriginal and Torres Strait Islander people at its centre, and recognises that governments need to draw on the perspectives, priorities and knowledges of Aboriginal and Torres Strait Islander people if outcomes are to improve.

2 A new approach to evaluation: an Indigenous Evaluation Strategy

For decades there have been calls to better understand how policies and programs are affecting Aboriginal and Torres Strait Islander people. However, there continues to be limited evidence about the effectiveness of many policies and programs designed to improve outcomes for Aboriginal and Torres Strait Islander people. Wide gaps between Aboriginal and Torres Strait Islander people and non-Indigenous people also persist across many indicators of wellbeing — life expectancy, health, educational attainment and employment — despite more than ten years of Closing the Gap initiatives.

Evaluation — the systematic assessment of a policy or program's design, implementation and outcomes — is about understanding what governments are doing right, what they are getting wrong and where they can do better. The Strategy is an opportunity to improve the quality and use of evaluations of Australian Government policies and programs affecting Aboriginal and Torres Strait Islander people, and Australians more generally.

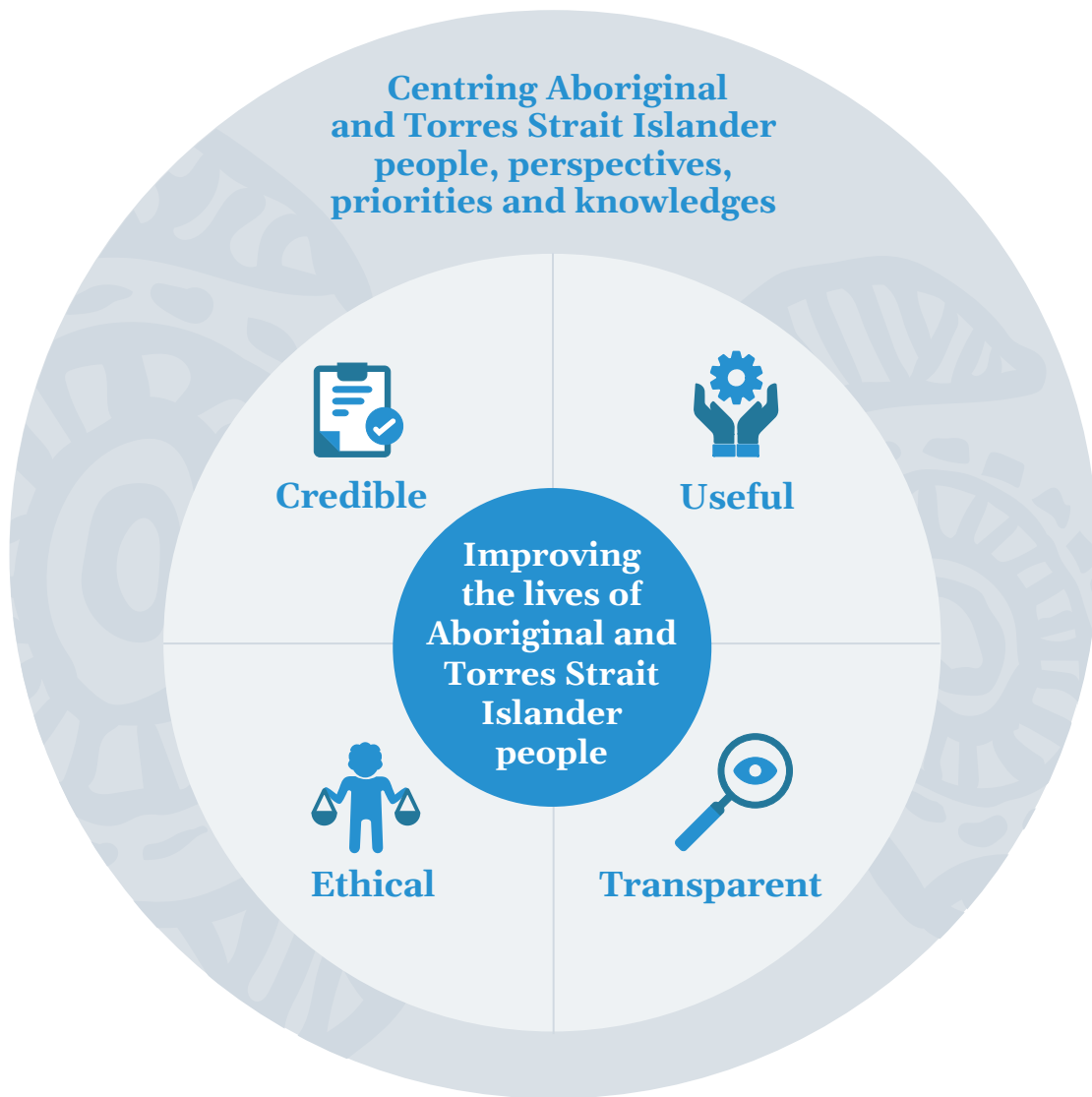
To achieve better policy outcomes, what Aboriginal and Torres Strait Islander people value, their expertise and lived experiences need to be reflected in what is evaluated, how evaluation is undertaken and the outcomes policies seek to achieve. If the outcomes of policies are not what is valued by Aboriginal and Torres Strait Islander people, then the policies have limited value and little prospect of improving the lives of Aboriginal and Torres Strait Islander people.

As well as providing guidance to agencies on planning, conducting and using evaluation, the Strategy supports more effective ways of working with Aboriginal and Torres Strait Islander people, and it includes governance arrangements that embed incentives to learn from and use evaluation findings in policy decision making. It is better policies, not evaluation per se, that will improve the lives of Aboriginal and Torres Strait Islander people.

Guiding principles for the Strategy

The overarching principle of the Strategy is centring Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges (figure 1). This principle is about recognising the strengths and diversity of Aboriginal and Torres Strait Islander people, communities, knowledges, histories and cultures. It is also about building partnerships with Aboriginal and Torres Strait Islander people to define policy and program outcomes, and decide on evaluation questions, how evaluations will be conducted and how evaluation findings will be interpreted.

Figure 1 Guiding principles for the Indigenous Evaluation Strategy



The overarching principle is also the lens through which the Strategy's other principles — credible, useful, ethical and transparent — should be interpreted. These principles frame how agencies should plan and conduct evaluations and how evaluations will be assessed.

Practical advice on implementing the Strategy

This guide has been developed as a companion to the Strategy. It provides practical advice for Australian Government agencies on how to conduct evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people.

The guide does not require users to have in-depth evaluation and methodological expertise. Rather, it is written for program managers and those commissioning and using evaluations. Links to further evaluation references, and more specialist sources, are provided throughout the guide.

The guide steps through the different stages of evaluating policies and programs and makes suggestions about how agencies can build a culture to support evaluation under the Strategy. It covers:

- building evaluation into policy and program design
- deciding what policies and programs to evaluate
- evaluation planning, design and conduct
- reporting and using evaluation findings
- building capability and a culture of evaluation.

At each of these stages, the guide outlines a set of questions to consider. Considering these questions will help ensure that those commissioning and using evaluations centre the perspectives, priorities and knowledges of Aboriginal and Torres Strait Islander people in the evaluation process. This in turn will improve the quality and usefulness of evaluations. Some hypothetical examples of evaluation in practice (in pink boxes) are also provided throughout the guide.

The Strategy does not replace, but rather complements and builds on Australian Government agencies' existing evaluation processes and frameworks. It provides an overarching whole-of-government framework and consistent accountabilities for all Australian Government agencies developing and implementing policies and programs that affect Aboriginal and Torres Strait Islander people.

While the guide is presented as a series of sequential stages, in practice, the process needs to be approached in a holistic way — how the later stages of an evaluation are implemented needs to be considered and integrated into policy and program development and administration from the outset. For example, while engagement is discussed in section 3, centring Aboriginal and Torres Strait Islander perspectives, priorities and knowledges should happen at each and every stage of the evaluation process.

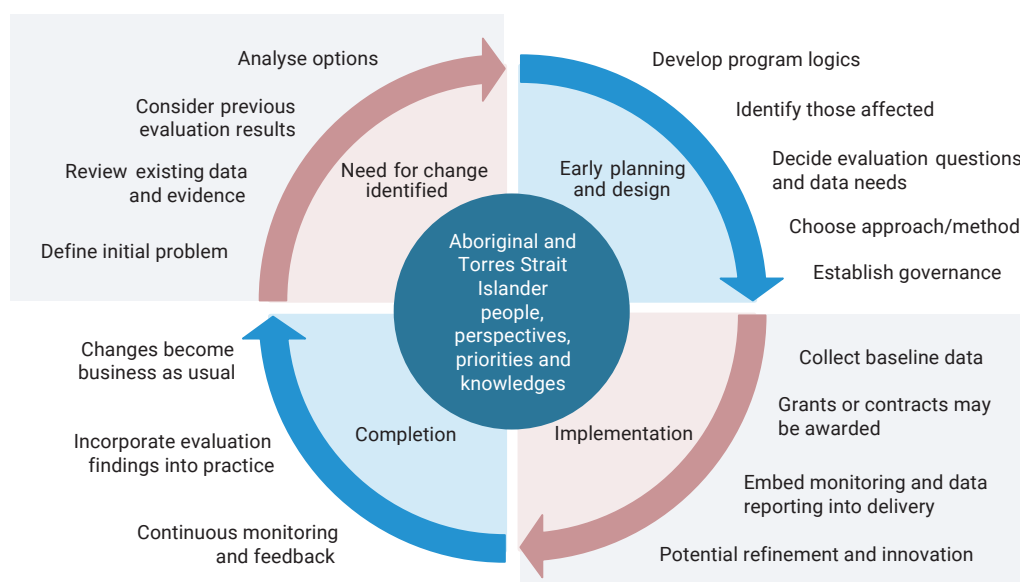
Section 1

Building evaluation into policy and program design

Evaluation is most effective when it is planned for early and when it is integrated into each stage of policy and program design – from setting policy objectives through to using evaluation findings to inform future policies and programs. And to ensure that Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges are centred, engagement also needs to be considered at each stage (figure 2).

Figure 2 Evaluation and engagement should be considered at every stage of policy and program design and implementation

5



Before a policy or program is implemented, evaluation should inform policy design. While a policy or program is being implemented, evaluation should inform thinking on ways to improve the policy or program. And after the policy or program has been implemented, evaluation can provide evidence on outcomes and impacts that can be drawn on to inform future policies and programs.

Drawing on evaluation evidence when designing a policy or program

Good practice policy and program design draws on existing evidence about what works and why, including evidence from previous evaluations. Policy makers should examine the existing evidence base at the earliest stages of policy and program design to see what lessons there are from similar policies or programs. This includes assessing the strengths (and weaknesses) of the evidence supporting the assumptions upon which a policy or program is based. It also involves looking at the evidence from trials or pilots of similar policies or programs and relevant behavioural insights on the best way to implement the policy or program and what the likely outcomes could be.

When assessing existing evidence, it is important to critically examine the strengths and limitations of the evidence and check that it is relevant for the proposed policy or program situation, setting and local context. An assessment of the existing evidence can also highlight gaps or unanswered questions that evaluation could seek to address (box 1).

Box 1 Review and synthesis of existing evidence

Working out what is already known about a particular form of intervention is not always easy, particularly when there are volumes of research evidence. And because not all evidence is of equal value, some way of differentiating between high- and low-quality studies is needed.

Evidence synthesis methodologies aggregate evaluation findings and review them in a systematic way (and generally also assess and rate the strength of the evidence).

Systematic reviews aggregate results that fit a pre-specified eligibility criteria in order to answer a specific research question. They aim to minimise bias by using explicit, systematic methods that are documented in advance with a protocol.

Systematic reviews search all the available research evidence for a particular question, which counters problems of selection bias that come from only identifying studies that are readily accessible or only published on major databases. They can help highlight gaps in the evidence (and identify evaluation questions that need to be answered) and methodological issues that may need to be considered in the design of an evaluation.

A meta-analysis pools statistical results from multiple studies as a way to increase explanatory power and improve estimates of the size of effects. Combining a diverse array of evidence can also provide better information about what works, why and for whom. A realist synthesis combines diverse evidence to provide information about what works for whom in what contexts (useful for situations where the same policies and programs do not work the same everywhere or for everyone).

A number of organisations collate and synthesise research and evaluation evidence. International examples include Cochrane and the Campbell Collaboration. There are also organisations that synthesise evidence relating to Aboriginal and Torres Strait Islander people. The Australian Indigenous HealthInfoNet, for example, provides an evidence base to inform practice and policy in Aboriginal and Torres Strait Islander health. The Lowitja Institute also publishes all its research on the Institute's website.

The Australian Institute of Health and Welfare is developing a Regional Insights for Indigenous Communities website (which will provide regional data on topics such as language and culture and health) and an Indigenous Mental Health and Suicide Prevention Clearinghouse. An Indigenous Research Exchange is also being developed by the Australian Institute of Aboriginal and Torres Strait Islander Studies.

Questions to consider when drawing on evidence when designing a policy or program

- Is there evidence from other policies and programs on the problem or issue you are seeking to address? Does it validate your hypothesis and drivers of change?
- How credible and relevant is the evidence to support or validate the success factors or outcomes you are seeking? Have you considered differences in context and the strengths and limitations of the existing evidence? Does the evidence shed light on any potential unintended effects or behavioural changes from similar policies or programs?
- Has the evidence been tested in Australia or in a similar setting and/or a similar population? Is there evidence demonstrating the effects on Aboriginal and Torres Strait Islander people?
- Have you looked at evidence from a range of sources, including insights from Aboriginal and Torres Strait Islander people and organisations?
- Does the evidence provide any guidance (resources, timing, capability) on how to implement the policy or program?

Planning for evaluation during policy and program design

Planning early and being clear about the objectives of a policy or program can result in both better policies and better evaluation. Developing an evaluation plan as part of the policy or program design has a number of benefits. It can help ensure:

- evaluations identify the objectives and outcomes of a policy or program and clarify what the evaluation questions should be
- the right data (such as baseline and monitoring data) are collected for the evaluations and data collection systems are in place from the start – this will improve the quality, rigour and usefulness of the evaluation
- agencies consider whether policies or programs can be designed or implemented in ways that support data collection and particular evaluation approaches
- Aboriginal and Torres Strait Islander people, and others affected by a policy or program, are engaged early on in the process for decisions about policy objectives and outcomes as well as evaluation questions, approaches, methods and how evaluation results will be interpreted, communicated and disseminated
- the different effects of mainstream policies and programs on Aboriginal and Torres Strait Islander people, and potentially other minority groups, are considered (and the data required to evaluate the effects are collected)
- adequate resources and realistic timeframes are allocated for the evaluation (noting that effective community engagement can take time).

Developing a *program logic* (also called a logic model or theory of change) is a critical first step in evaluation planning. A program logic describes the relationship between policy or program inputs, outputs and outcomes, and spells out the assumptions that underpin the relationships. It is a useful way to be explicit about the evidence you are using, how you expect to get results, what outcomes you will measure, and how you will measure the outcomes. While a critical early step, the program logic may need to be updated over time in response to analysis of new evidence.

Under the Strategy, the Commission proposes that New Policy Proposals developed by Australian Government agencies should have an 'Aboriginal and Torres Strait Islander Impact Assessment and Evaluation Plan (IAEP)' (box 2). The IAEP would establish a standard for the level of planning and engagement appropriate for the new proposal, based on the expected level of impact on Aboriginal and Torres Strait Islander people.

Box 2 Aboriginal and Torres Strait Islander Impact Assessment and Evaluation Plan

To help embed evaluation planning and engagement into the early stages of the policy design process, it is proposed that agencies prepare an Aboriginal and Torres Strait Islander Impact Assessment and Evaluation Plan (IAEP) for new policy or program proposals for Cabinet or Government (in the case of decisions by Ministers and/or statutory authorities) endorsement. The IAEP should detail:

- the Aboriginal and Torres Strait Islander people, organisations and communities that are expected to be affected by the proposed initiative
- a proposed engagement plan (including timeframes and cultural safety considerations)
- a proposed evaluation approach (including the scale of evaluation required)
- the data required to assess the policy's impact and how they would be collected
- an estimated evaluation budget.

When preparing an IAEP, agencies will need to make an assessment of the impact of the policy or program on Aboriginal and Torres Strait Islander people. Policies expected to have a proportionally high level of impact should have higher engagement requirements, and more thorough evaluation planning, than those expected to have a lower impact.

Evaluation in practice: Budgeting for evaluation during policy and program development

A new Indigenous ranger program is being developed for roll out in remote areas. The Impact Assessment and Evaluation Plan prepared as part of the New Policy Proposal has identified that the program could have a significant effect on Aboriginal and Torres Strait Islander people's wellbeing. The plan is for the evaluation to include quantitative analysis of administrative data as well as qualitative case studies in several locations where the program is operating.

When estimating the cost of the future evaluation, the agency budgeted extra time and money (compared with an earlier evaluation of a similar program rolled out in regional areas) on the basis that travel costs were likely to be higher, and evaluators would need to visit each case study area several times to build relationships with community members. In some areas where the program is proposed to run, English is spoken as a second or third language by many people. The evaluation budget allowed for translation of surveys and responses.

The proposed timeline for the evaluation also factored in the need for a formal ethics review, as well as meeting the requirements of research protocols, which are in place for those communities expected to be involved in the case study.

Policy design and evaluation

At the early stages of policy and program design, agencies should think about whether the policy or program can be designed or implemented in a way that aids evaluation. Getting things right from the start is important for lifting the quality and usefulness of evaluation results.

Rigorous impact evaluation requires an estimate of the counterfactual: what would have occurred in the absence of the policy or program? If designed well, piloting, randomisation and/or phased introduction of a policy or program can help evaluators compare the outcomes of a policy or program for participants with those for non-participants.

Pilots and trials can also allow policy makers to experiment and/or be innovative with policies and programs that otherwise might be too risky or expensive to put in place on a larger scale. By building evaluation into a pilot, lessons on implementation and impact can be used to refine or end a policy or program before it is rolled out more widely.

However, pilots need to be well designed and implemented. This includes the pilot or trial having:

- a clear purpose from the start so appropriate methods and timeframes can be put in place. The methodology, timeframe and scale of any pilot or trial needs to be such that the evaluation results produced are not misleading (for example, the outcomes of some policies will only be achieved after a policy has been in place for a certain period of time, which means too short a timeframe to test for outcomes could produce misleading or false results)
- mechanisms in place that allow the policy or program to be adapted or stopped (if needed) because of evaluation findings
- adequate resources and assessments of ethical and cultural safety risks.

Evaluation in practice: Lessons from a pilot program

An agency is developing a new legal assistance program targeted at young people. The program uses an innovative approach to provide integrated legal and other support services for young people at risk of incarceration. Because this approach has not been tried before in Australia, the agency is introducing it as a pilot program before rolling it out across Australia. The aim is to see if the approach improves outcomes, as well as to iron out any implementation issues.

The sites for the pilot are chosen to ensure that a well-matched control group of non-participating sites can be used for comparison. Both a process evaluation — focusing on whether the program is being implemented as planned — and an impact evaluation — to be conducted at the conclusion of the pilot — are planned. The pilot sites are chosen to reflect the diversity of settings in which the program might eventually operate, as well as to allow inferences to be made about the impacts on key participant groups, including Aboriginal and Torres Strait Islander people.

The agency uses the results to decide whether the program should be expanded, and if so, how it should be implemented to maximise positive outcomes. As a key objective of the pilot is to learn whether this innovative approach is suitable for use in Australia, the agency shares the lessons from the evaluation.

Questions to consider when planning for evaluation during policy and program design

- Will the policy or program have a significant and/or different effect on Aboriginal and Torres Strait Islander people? If so, is this reflected in the Aboriginal and Torres Strait Islander Impact Assessment and Evaluation Plan?
- Are the program objectives clearly identified? How are Aboriginal and Torres Strait Islander people involved in developing the program logic or theory of change? Does the program logic or theory of change reflect the outcomes that are valued by Aboriginal and Torres Strait Islander people? What are the priorities of the Aboriginal and Torres Strait Islander people who will be affected by this policy or program? What are the gaps in the existing evidence base that evaluation of the proposed policy or program could fill?
- Could there be unintended negative consequences of the policy or program for Aboriginal and Torres Strait Islander people's wellbeing that are not captured in the program logic?
- Does the evaluation plan identify opportunities to engage with Aboriginal and Torres Strait Islander people, and other stakeholders, during the evaluation?
- Does the proposed policy or program budget include sufficient resources for future evaluation, including to collect data and engage with Aboriginal and Torres Strait Islander people and others affected by the policy or program when developing an evaluation plan and undertaking evaluation? Are the time and resources allocated to evaluation proportionate to the size and importance of the policy or program being evaluated? Does the evaluation budget include resources to provide feedback on evaluation findings and opportunities to strengthen evaluation capability of Aboriginal and Torres Strait Islander people, organisations and communities?
- Can the policy or program be designed or implemented in a way that aids evaluation (such as by piloting, randomly allocating participants and non-participants or a phased introduction)? Does the evaluation design allow for inferences to be made about how a pilot program might work in different contexts or for different groups of participants?
- How have Aboriginal and Torres Strait Islander people informed policy and program development and evaluation plans?

Further reading on building evaluation into policy and program design

Indigenous Evaluation Strategy, Background Paper, chapter 5.

Better Evaluation nd, 'Define', https://www.betterevaluation.org/rainbow_framework/define.

HM Treasury 2020, *The Magenta Book: Central Government Guidance on Evaluation*.

Jowell, R. 2003, *Trying It Out: The Role of 'Pilots' in Policy-Making, Report of a Review of Government Pilots*, UK Government Chief Social Researcher's Office, December.

Section 2

Deciding what policies and programs to evaluate

It is not practical or feasible to evaluate every policy or program affecting Aboriginal and Torres Strait Islander people, so agencies need to decide what to evaluate. Setting evaluation priorities can help ensure limited evaluation resources are used in the most effective way to improve the lives of Aboriginal and Torres Strait Islander people.

12

When deciding what to evaluate, agencies need to identify which of their new and existing policies and programs should be considered for evaluation under the Strategy. Indigenous-specific and mainstream policies and programs to be evaluated under the Strategy include those that either align with the government-wide evaluation priorities (figure 3), or have a disproportionate and/or significant impact on Aboriginal and Torres Strait Islander people.

Figure 3 Proposed interim government-wide evaluation priorities
Based on the National Agreement on Closing the Gap

The Strategy should prioritise evaluation of government efforts ...			
... delivered as part of a formal partnership and/or shared decision making arrangement (Priority 1) that build up the Aboriginal and Torres Strait Islander community controlled sector (Priority 2) (particularly mainstream ones) that eliminate racism, embed and practice meaningful cultural safety, deliver services in partnership with Aboriginal and Torres Strait Islander people (Priority 3) involved in data sharing arrangements (Priority 4) ...
... and focused across (and within) the following policy domains :			
<ul style="list-style-type: none">• Education• Housing• Health and wellbeing	<ul style="list-style-type: none">• Land and water• Justice• Culture and languages	<ul style="list-style-type: none">• Safety (Families, children and youth)• Employment (Economic development)	

Source: adapted from Joint Council on Closing the Gap (2020).

Agencies should then use a criteria-based priority setting process (box 3) to determine which of the policies and programs should be prioritised for evaluation (and the extent to which they should be evaluated). Priorities should be determined based on policy and program impact, risk profile, strategic significance and expenditure, as well as on Aboriginal and Torres Strait Islander people's priorities, and should be informed by the National Agreement on Closing the Gap.

Box 3 Using multi-criteria impact assessment to determine agency evaluation priorities

Agencies should develop a fit-for-purpose multi-criteria impact assessment tool to help categorise the relative significance of policies and programs. To ensure consistency across agencies around prioritisation decisions, the criteria used for determining evaluation priorities should include:

- **Impact of the policy or program** – how likely the policy or program will impact the lives of Aboriginal and Torres Strait Islander people (taking into account the total Aboriginal and Torres Strait Islander population affected or whether a policy or program disproportionately affects Aboriginal and Torres Strait Islander people).
- **Strategic significance** – how important and valuable the policy or program is in terms of the competing priority areas within government (particularly government-wide evaluation priorities under the Strategy).
- **Risk profile** – the policy or program's overall risk. This may include how difficult it is to estimate the impact of the policy or program (the outcomes are uncertain or difficult to measure); or the risk that a policy or program could disproportionately affect the rights and lives of Aboriginal and Torres Strait Islander people.
- **Program expenditure** – the overall funding for the policy or program. This could be measured in absolute terms, or relative to the cost of other programs delivered or managed by the agency.

Each of the criteria should be underpinned by the perspectives, priorities and knowledges of Aboriginal and Torres Strait Islander people. Agencies should have appropriate governance arrangements in place to ensure Aboriginal and Torres Strait Islander people are engaged in these decisions.

To determine a ranking of policies and programs for evaluation, agencies should:

- assign scores for each policy or program against each criterion above, based on defined thresholds
- calculate a weighted average score for each policy or program against the criteria to determine where it sits along a continuum of overall 'significance'. ('Weights' could be applied to all criteria to ensure outcomes are not biased against criteria that may be less important across policies and programs within an agency, or where the criteria may be less reliably measured.)

Based on resources and the level of activity, agencies would then establish a threshold so that all 'significant' policies and programs would be resourced to be rigorously evaluated. Lower priority policies or programs could focus their evaluative activity on improving data collection towards more meaningful outcome measures, monitoring and performance audits, or other evaluative activities that facilitate learning by doing, such as user surveys or reviews.

The impact assessment methodology should be publicly available, and should be reviewed and updated over time to ensure it remains relevant.

Evaluation in practice: Balancing impact, risk, expenditure and strategic significance when determining evaluation priorities

An agency is applying its multi-criteria impact assessment tool to prioritise three programs for evaluation: a pilot intensive family support program being trialled in a remote community; a large mainstream education program; and an established Indigenous-specific education program.

The agency uses a five-point scale to score each program against the criterion – based on impact, strategic significance, risk profile and program expenditure. For example, the impact of the program for Aboriginal and Torres Strait Islander people could be given a score of five if it is a program aimed entirely at Aboriginal and Torres Strait Islander people, or one if it is expected to have less than 5 per cent Aboriginal and Torres Strait Islander clients.

Once a score has been allocated to each criterion, a weighted score is calculated for each program. The agency has decided to weight all criteria equally.

<i>Program</i>	<i>Impact of program</i>	<i>Strategic significance</i>	<i>Risk profile</i>	<i>Program expenditure</i>	<i>Weighted score</i>
Pilot study in a community	4	5	5	2	4
Mainstream education program	2	4	5	5	4
Established Aboriginal and Torres Strait Islander education program	5	5	2	2	3.5

The results of the prioritisation process show that the pilot study and the mainstream education program are considered the highest priorities for evaluation.

- The pilot study has only a relatively 'small' aggregate impact in terms of expenditure and population, but it is assessed as being high priority for evaluation because the potential risk of the program to a child's wellbeing is significant and it is a particularly innovative program in a strategically-important policy area where the existing evidence base is poor.
- The mainstream education program has a relatively small proportion of Aboriginal and Torres Strait Islander participants, but has a large budget and is designed to affect school attainment, an outcome that has been identified as being a very high priority to Aboriginal and Torres Strait Islander people and the government.
- The established Aboriginal and Torres Strait Islander education program is ranked as a lower priority for evaluation because it has a relatively small budget and has already been subject to recent evaluation.

To provide transparency around how evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people are prioritised, each Australian Government department should centrally coordinate, on an annual basis, a rolling Three Year Evaluation Forward Work Plan, which details:

- policies and programs within their portfolio that contribute to government-wide evaluation priorities aimed at improving the lives of Aboriginal and Torres Strait Islander people
- how the department identified high priority policies and programs (based on the above criteria)
- a plan for how/when over the next three years the department's identified policies and programs will be evaluated (or how they will become ready for evaluation)
- how Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges were centred as part of the prioritisation process.

Questions to consider when deciding what policies and programs to evaluate

- What policies and programs administered by your agency should be considered for evaluation under the Indigenous Evaluation Strategy?
- Have Aboriginal and Torres Strait Islander people been engaged in the process of deciding what policies and programs to evaluate?
- Do evaluation priorities align with the priorities of the National Agreement on Closing the Gap?
- Is an evaluation work plan in place? Is it publicly available?

Further reading on deciding what to evaluate

Indigenous Evaluation Strategy, Background Paper, chapter 6.

Department of Industry, Innovation and Science 2019, *Evaluation Strategy 2017 to 2021*, Canberra.

Joint Council on Closing the Gap 2020, *National Agreement on Closing the Gap*.

Section 3

Evaluation planning, design and conduct

Once an agency has decided to evaluate a policy or program, detailed evaluation planning should take place. This involves deciding on the evaluation questions, the type of evaluation, the approach and methods that will be used, the data needed to answer the evaluation questions, and who will conduct the evaluation. The evaluation plan should also identify key stakeholders, and include an engagement plan, a budget and an evaluation timeline.

16

This section outlines some of the key decisions that agencies need to make during an evaluation. Some of these decisions will have to be made during the planning phase, others while the evaluation is being conducted, or later when the findings are released and acted on (and some decisions will need to be revisited over the course of an evaluation). Knowing what key decisions need to be made before, during and after an evaluation can help agencies pre-empt problems and potentially address them during the planning phase.

Identifying and engaging with those affected by the policy or program

Engagement should take place early and throughout the stages of an evaluation. Engaging on evaluation questions, approaches, methods and reporting is key to Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges being centred in evaluations.

Engagement will look different depending on the policy or program being evaluated, the groups and communities affected, and the scope and purpose of an evaluation. Some evaluations will be Aboriginal and Torres Strait Islander led or co-designed. Such engagement approaches can work well when evaluating complex problems that require working collaboratively with those receiving a program to find policy solutions. The National Agreement on Closing the Gap identifies formal partnership and shared decision making as a priority reform area and has articulated the elements of a strong partnership (including adequate funding for Aboriginal and Torres Strait Islander parties to be partners with governments in formal partnerships).

A policy or program that is assessed as being a priority for improving the lives of Aboriginal and Torres Strait Islander people should look to having more involved forms of engagement (where those affected by a policy or program can have influence over evaluation decisions). Policies and programs that are assigned a lower priority could require less involved and ongoing engagement. What is important is that Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges are taken into account. For mainstream policies and programs, this will assist agencies to identify policies or programs that could work differently for Aboriginal and Torres Strait Islander people (it will also ensure the data required to evaluate the effects are collected).

Engagement is not without costs. Government agencies need to consider what resources and time are needed for those affected by a policy or program to be able to engage effectively in an evaluation.

Getting engagement right is critical for high-quality evaluations — many of the concerns raised about evaluation quality relate to a lack of engagement with Aboriginal and Torres Strait Islander people, and a lack of regard for Aboriginal and Torres Strait Islander knowledges.

An engagement plan should be developed to help agencies identify opportunities for, and barriers to, engagement. Opportunities for engagement with Aboriginal and Torres Strait Islander people during an evaluation could include:

- engaging with Aboriginal and Torres Strait Islander peak bodies or community representatives during the evaluation planning phase
- Aboriginal and Torres Strait Islander leadership and/or representation on evaluation governance or steering committees
- engaging Aboriginal and Torres Strait Islander evaluators to lead evaluation planning and conduct
- governance arrangements for evaluations within an agency (for example, the agency's evaluation committee) including Aboriginal and Torres Strait Islander members.

17

Evaluation in practice: Engaging with Aboriginal and Torres Strait Islander people on evaluation design

To build engagement with Aboriginal and Torres Strait Islander people into the design of an evaluation of an Indigenous-specific early childhood program, an agency developed an evaluation plan as a separate project.

External evaluators with experience and expertise working with Aboriginal and Torres Strait Islander people were commissioned to develop a detailed evaluation plan for the evaluation. The contract stipulated that engagement with Aboriginal and Torres Strait Islander people was to be a key deliverable for the project and that Aboriginal and Torres Strait Islander people were to be engaged in the planning phase of the evaluation, and throughout the course of the evaluation.

As part of the evaluation plan project, a steering committee with majority Aboriginal and Torres Strait Islander members was set up. The steering committee worked with the evaluators to develop the evaluation questions, choose methods and suitable indicators, provide guidance on ethics and develop a stakeholder engagement plan. The steering committee continued to guide the evaluation during the conduct phase.

Questions to consider when identifying and engaging with those affected by the policy or program

- Who will be affected by the policy or program? Have Aboriginal and Torres Strait Islander people with an interest in the evaluation been identified?
- Have Aboriginal and Torres Strait Islander people had input into how engagement takes place? Do engagement methods take account of contextual factors? Have you considered whether Aboriginal and Torres Strait Islander languages should be used? Has sufficient time been allocated for engagement, including relationship building and establishing trust with those affected by the policy or program?
- Can Aboriginal and Torres Strait Islander people participate in the evaluation on an equal footing with others? Are there language barriers or other factors that might limit Aboriginal and Torres Strait Islander people participating?
- Have cultural norms and standards for engaging with Aboriginal and Torres Strait Islander communities been adopted? Have any potential barriers to engagement been identified and mitigated? Do resources need to be provided to enable Aboriginal and Torres Strait Islander people to participate in the evaluation?
- What governance arrangements need to be put in place? Are Aboriginal and Torres Strait Islander people represented in evaluation governance arrangements?
- What mechanisms are in place to make evaluation processes transparent?

Deciding who will conduct the evaluation

Deciding whether an evaluation will be conducted in-house by the agency or contracted out is usually made after some evaluation planning has taken place, but in most cases evaluators will be involved in developing a more detailed evaluation plan.

Independence of evaluators from policy and program management will enhance credibility, although close interaction between evaluators and policy and program staff can also be important for informing evaluation design. One option is for some parts of an evaluation to be undertaken externally (for example, data collection) and other parts undertaken internally by the agency. However, there should be functional independence between evaluators and those who make decisions about policy and program design and delivery.

The decision about whether the evaluation is undertaken internally or externally should be based on who is best placed to undertake the evaluation given its purpose, circumstances and the type of evaluation. Agencies should ensure that evaluators (whether in-house or external) have the requisite skills and experience to implement the Strategy's principles – including centring Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges (noting that evaluations can use sophisticated methods that require particular skills and expertise).

For some evaluations – such as those of place-based initiatives and community-led projects – Aboriginal and Torres Strait Islander leadership of the evaluation may be the most effective way to align with the overarching principle. However, these approaches will not be feasible or appropriate for all evaluations, including where the program has a broader mainstream focus, and/or experienced Aboriginal and Torres Strait Islander evaluators are not available or prepared to partner on an evaluation.

Where agencies commission external evaluators, it is important that commissioning practices address – and do not hinder – the need for Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges to be centred during evaluation (box 4). Done poorly, commissioning practices can limit the scope for external evaluators to centre Aboriginal and Torres Strait Islander people. On the other hand, good commissioning practices can improve the quality of evaluations, and potentially contribute to expanding the capacity of culturally capable evaluators, including Aboriginal and Torres Strait Islander evaluators.

Box 4 Centring Aboriginal and Torres Strait Islander people when commissioning evaluators

While both internal and external evaluations should centre Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges, many decisions that an evaluation commissioner makes (such as evaluation timelines, resourcing or scope) have the potential to affect how an evaluation is conducted. As such, commissioners should carefully consider how their commissioning practices may affect how Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges are centred. These include:

- ensuring that Aboriginal and Torres Strait Islander evaluators are able to bid for contracts by using open tender processes or procurement panels that include Aboriginal and Torres Strait Islander evaluators
- that an evaluator's experience in conducting research with Aboriginal and Torres Strait Islander people and their relationships with the communities involved are included as part of the selection criteria
- ensuring evaluators engage local researchers where necessary
- commissioners ensuring sufficient funding and time is allowed for appropriate engagement with Aboriginal and Torres Strait Islander people during an evaluation, and for conducting the evaluation in line with ethical research practices (including allowing time for ethics clearance processes if required)
- ensuring that contracts and agency approval processes enable findings to be communicated back to Aboriginal and Torres Strait Islander participants and communities
- commissioners and evaluators working together to ensure ongoing communication with Aboriginal and Torres Strait Islander people participating in the evaluation.

Evaluation in practice: Good practice commissioning in a small-scale evaluation of a mainstream program

A business coaching program is designed to help small to medium business enterprises expand. It is not targeted at Aboriginal and Torres Strait Islander businesses, and administrative records show that few Aboriginal and Torres Strait Islander businesses use the program. Program managers want to know how to improve the program for Aboriginal and Torres Strait Islander (and other) users.

Aboriginal and Torres Strait Islander business experts were asked to provide feedback on the evaluation plan, which noted that Aboriginal and Torres Strait Islander businesses had difficulty accessing the program. Their feedback was incorporated into the evaluation plan and it was decided that a small case study was the best way to identify challenges, while keeping evaluation costs low.

The agency decided to contract one evaluator to undertake the general evaluation and data analysis. A second evaluator, with experience working with Aboriginal and Torres Strait Islander people, was contracted to conduct the case study. This allowed the agency to draw on specialised skills for the evaluation team.

The commissioner and evaluators had regular catch-ups to ensure that findings were integrated and shared during the evaluation. The data analysis found that while Aboriginal and Torres Strait Islander businesses did equally as well as non-Indigenous businesses from accessing the program, smaller Aboriginal and Torres Strait Islander businesses were less likely to apply for the program. The case study then explored these issues in detail with two different businesses.

The evaluation contracts stipulated that a results interpretation session would be conducted to ensure that findings from both studies were brought together. Those who gave feedback on the evaluation plan were invited to the session. Each participating business was given a short summary of the findings and their data.

Questions to consider when deciding who will conduct the evaluation

- Will the evaluation be conducted by internal or external evaluators, or some combination of the two? Are potential evaluators objective and independent of those who make decisions about policy and program design and delivery?
- Should the evaluation be led by, or co-designed with, Aboriginal and/or Torres Strait Islander people? Would engaging an Aboriginal and Torres Strait Islander evaluator increase the relevance, cultural safety and credibility of the evaluation?
- Are evaluators culturally and technically competent? Do they have the necessary skills to conduct the evaluation with Aboriginal and Torres Strait Islander people effectively? Do they have existing partnerships with Aboriginal and Torres Strait Islander people?
- Who will be responsible for tendering, project management and quality assurance? Are the processes and criteria used to make tender decisions transparent? Are evaluators selected through an open and transparent process?
- Have the terms of reference/request for tender been shaped by Aboriginal and Torres Strait Islander people and community input – for example, in terms of the required level of engagement, ethics requirements, evaluation outcomes, and treatment of intellectual property?
- Do procurement processes provide opportunities for Aboriginal and Torres Strait Islander evaluators to bid for contracts?

Determining the purpose and scope of the evaluation

An important early step to a high-quality evaluation is being clear about its purpose and intended use (why the evaluation is being undertaken, who it is being undertaken for, and how it will be used). Answering these questions requires being clear about the policy or program's objectives and the levers that the policy will use to achieve its intended effects. The policy or program's logic model (section 1) will help evaluators identify the outcomes that need to be measured. If a program logic has not already been developed, this should be done as part of a detailed evaluation plan.

The evaluation's objective and questions should be developed with evaluation users and relevant stakeholders. Centring the perspectives, priorities and knowledges of Aboriginal and Torres Strait Islander people means that, for relevant policies and programs, Aboriginal and Torres Strait Islander people are engaged to help shape the evaluation questions so that the evaluation examines effects and outcomes that are important to Aboriginal and Torres Strait Islander people. Early engagement is key.

The scope of an evaluation should consider the length and timing of the evaluation, available resources, target groups and geographical areas to be covered by the evaluation.

Questions to consider when determining the purpose and scope of the evaluation

- Why is the evaluation being undertaken? Who is it being undertaken for, and how will the evaluation findings be used?
- Does the program logic clearly state how the program outcomes are expected to be achieved and how they are linked to higher level outcomes?
- What are the evaluation questions? Have Aboriginal and Torres Strait Islander people, and others affected by the policy or program, had input into what the evaluation questions and evaluation approaches and methods should be?
- For mainstream programs, have the experiences of Aboriginal and Torres Strait Islander people been given sufficient consideration?

Choosing an appropriate evaluation type, approach and method

Evaluations are undertaken for a range of reasons, including to support evidence-informed policy development, learning by doing, and accountability. They can answer a broad range of questions, but there are three main questions they seek to answer.

- How well is the policy or program being delivered?
- What difference did the policy or program make?
- Do the benefits of the policy or program justify the costs?

Different evaluation types, approaches and methods can be used to evaluate policies and programs (table 1).

Table 1 Key evaluation terms

Evaluation type	The type of evaluation is defined by the evaluation question. The three main types of evaluation are process evaluations, impact evaluations and economic or value-for-money evaluations.
Evaluation approach	The approach used to answer evaluation questions. For example, an impact evaluation could use an experimental, quasi-experimental or non-experimental approach.
Evaluation method	Method refers to the way that information is collected and analysed to test theories and answer evaluation questions (for example, randomised controlled trials, propensity score matching or case studies).
Data collection	The collection of information to use in evaluation. This can be quantitative or qualitative. ^a

^a Different sources of data are discussed in box 7.
Source: adapted from HM Treasury (2020b, p. 7).

What type of evaluation?

There are three broad types of evaluation – process, impact and value-for-money. The evaluation type adopted will depend on the purpose of the evaluation and the questions that need to be answered (table 2).

Table 2 What type of evaluation? It will depend on the purpose and the questions you want answered

<i>Evaluation type</i>	<i>Questions</i>	<i>Purpose</i>
Formative or process evaluations – these evaluations are undertaken early in the development or implementation of a policy or program.	<ul style="list-style-type: none"> – How is the policy or program delivered? – Is the program being delivered as intended? – Is the policy or program appropriately targeted? – How effective has implementation been so far? – What are the policy or program's strengths and weaknesses? – Can the policy or program be improved to achieve better outcomes? 	These evaluations help us to better understand the mechanisms at play in successful and less successful policies and programs. They can help shape a policy or program to perform better. The evaluation can assist in improving an initiative as it is rolled out and can provide a baseline for future evaluations.
Summative, outcome or impact evaluations – these evaluations judge the overall merit, worth and impact of a policy or program.	<ul style="list-style-type: none"> – What difference did the policy or program make? – Has the policy or program achieved its objectives? Does the evidence support the theory? – Has the policy improved outcomes? If so, by how much? – Did the policy affect groups of users differently? 	These evaluations are undertaken for lesson-learning (they can be used to inform decisions about whether to expand, cease, replicate or scale up a program) and accountability.
Economic evaluations – these evaluations assess the net benefit of a policy or program.	<ul style="list-style-type: none"> – Do the benefits justify the costs, or was it worth it? – Are there alternative approaches that would result in lower costs for the same benefits? 	These evaluations quantify the value of policies and programs and can be used to compare options. They are undertaken for accountability and resource allocation decisions.

The type of evaluation chosen will also be influenced by the characteristics of the policy or program. Often the type of evaluation will depend on what stage the policy or program is at. Answering evaluation questions can involve a combination of evaluations. For instance, a process or formative evaluation could be undertaken in the early stages of implementing a new policy or program. Early evaluations can be useful to allow agile and adaptive policy development or help ensure that poor implementation practices do not derail what otherwise might be a good policy or program. Subsequently, an impact evaluation could be conducted once the policy or program has been operating for several years.

Other factors, such as the kind of policy or program being evaluated, the circumstances under which the policy or program is implemented and the resources available for the evaluation will also affect the type of evaluation. For some policies and programs, even after they have been fully implemented it may not be feasible or worthwhile to assess the impacts or net benefits and a process evaluation may be all that is appropriate. This might be the case when a program is relatively small and where it is difficult to discern the effects of the program from those of other programs with similar objectives.

There are a number of questions agencies should ask when deciding on a type of evaluation.

- How complex is the relationship between the policy or program and the outcome (simple relationships can be investigated using process evaluations)?
- How important is it to control for other factors affecting the outcome (if it is important to control for other factors this will point to an impact evaluation)?
- How significant are the potential outcomes of the policy or program in terms of its contribution to an overall policy objective?
- How significant is the policy or program in terms of resources?
- How is the policy or program being implemented (will it allow for a counterfactual)?¹

Evaluation approaches and methods

Once the purpose and type of an evaluation are decided, evaluation approaches and methods need to be chosen. The choice of approach(es) and methods should maximise the rigour and credibility of the evaluation given the questions the evaluation seeks to answer, the nature and scale of the policy or program, resourcing and data constraints.² The different approaches and methods have advantages and disadvantages. A combination of approaches and methods will often provide the most complete and useful answers to evaluation questions.

Rigour is required for all types of evaluations, approaches and methods, and requires:

- ensuring findings about cause and effect are well founded
- being clear about the populations, settings or circumstances to which evaluation results can be generalised
- that the indicators or metrics used accurately capture the outcomes of interest.

Experimental and quasi-experimental approaches infer the net impact of a policy or program through statistical comparison to a group or time period that has not been subjected to the policy or program. An experimental or quasi-experimental approach can provide an objective measure of the effect of a policy or program. These approaches are most suited to evaluations where there is a new program – or distinct change to an existing program – that is expected to have a known, measurable and relatively large impact, where data are available and extraneous factors can be controlled.

Experimental approaches, such as randomised controlled trials (RCTs), construct a control group through random assignment to assess the net impact of an initiative.³ Usually, the average effect of the treatment on the treated population is compared with the effect on a suitably chosen control group, although more rigorous analysis attempts to understand why some treatments have an effect, on whom, and for how long. Sometimes an evaluation will capture variations in the intensity of services across the population, to examine the impacts of different service levels.

¹ Adapted from HM Treasury (2011, p. 24).

² Table 4 in the *Magenta Book - Supplementary Guide: Handling Complexity in Policy Evaluation* sets out some evaluation questions and approaches for answering them (HM Treasury 2020a, pp. 46–47).

³ Some useful guides on experimental approaches to policy and program design and evaluation include: *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials* (Haynes et al. 2012); *Better Public Services Through Experimental Government* (Breckon and Mulgan 2015).

Quasi-experimental approaches, such as propensity score matching and difference-in-differences analysis, also measure impacts, and can involve statistical analysis using pre-test and post-test results, but they do not involve random allocation between control and treatment groups. They are simpler to administer than RCTs and are less likely to be limited by practical and feasibility constraints, but are often considered weaker forms of evidence because of problems with internal validity and establishing the counterfactual.

Experimental and quasi-experimental approaches are not always feasible or appropriate for answering some evaluation questions (such as what the different impacts of a policy or program are on particular groups, or why a policy or program did or did not work). Where data are of poor quality, the use of experimental or quasi-experimental designs can produce misleading results. And for some policies, programs and circumstances it may not be ethical or feasible to provide different services to different groups to create a control or counterfactual for statistical analysis.

Qualitative, theory-based approaches can help answer some of the questions that experimental approaches cannot answer. Theory-based impact evaluations draw conclusions about the impact of a policy or program by testing whether the causal chains thought to bring about change are supported by evidence (they are concerned about the extent of the change and why change occurs).⁴ They are based on a well-defined theory of change and the theory is tested through various evidence sources. Some examples of theory-based approaches include realist evaluation (box 5) and contribution analysis.

Theory-based approaches are particularly suited when there are many policies and programs in place, an adaptive approach is being taken to implementing the policy or program, and there is no predicted outcome. Other approaches designed to support learning and adaptation, and/or to use in complex systems, include developmental, participatory and empowerment approaches (box 5). These approaches are not necessarily mutually exclusive — for instance, a developmental approach could incorporate participatory or collaborative elements.

4 Rigour in theory-based methods comes from: coherence of the theories; evidence that is specific enough to test the theories; triangulation of multiple sources; ruling out of alternative causes; critical reflection and opening up to peer review and external scrutiny (HM Treasury 2020b, pp. 36–37).

Box 5 Different evaluation approaches may be required

Developmental evaluation seeks to maximise the success of a policy or program through incremental review and improvement as it is rolled out, and relies on program participants' responses to shape learning and improvement. The evaluator is not an independent observer, but a core part of the program design team and involved in long term progress and outcomes monitoring. Developmental evaluation can be an iterative process — design a policy or program, trial it, evaluate it, modify it guided by the evaluation results, then trial and evaluate a second version, and so on.

Realist evaluation is a form of theory-based evaluation designed to help understand the circumstances under which some programs are more or less effective with different groups of people. It seeks to provide contextual explanations for why change occurs, while uncovering the diversity of participants' experiences. However, realist approaches can provide insights on how and why a program works and can identify areas for improvement, but do not seek to measure whether change has occurred compared with other scenarios. Findings are context specific and may not be generalisable.

Participatory evaluation approaches involve collaboration between the evaluator and stakeholders. The evaluator and the stakeholders jointly make decisions about the evaluation questions, how the evaluation is conducted, and how evaluation results are shared. Participatory approaches can be used at any stage of the evaluation process, can be practiced in various ways and can involve methods such as storytelling and participatory social mapping.

Empowerment evaluation is another collaborative approach that involves a set of principles that guide the evaluation at every stage. The model aims to create a sense of ownership to ensure program outcomes will benefit participants. The approach is designed to involve all stakeholders (evaluators, management, practitioners, participants and the community) in the evaluation process. Empowerment evaluation seeks to create positive change, as well as evaluating program outcomes. Capacity building is one of the key objectives of empowerment approaches. One criticism of empowerment evaluation is that the process is resource and time intensive. Another is that empowerment evaluation can be more a method of teaching people about evaluation than an evaluation itself.

Indigenous evaluation approaches place an Indigenous lens over the initiative being evaluated, as well as the evaluation process itself, to ensure that the values, experiences, knowledges and interests of Aboriginal and Torres Strait Islander people are recognised and upheld. While there is no one definitive Aboriginal and/or Torres Strait Islander approach, these approaches are generally principles- and strengths-based, and they promote self-determination (and the inclusion of Indigenous people in the evaluation process), Indigenous ways of knowing and culturally-relevant and respectful processes.

Participatory, developmental and realist evaluations are particularly suited to allowing Aboriginal and Torres Strait Islander knowledges, perspectives and world views to be incorporated into the design and delivery of evaluations. Culturally valid methods, such as yarning (storytelling), ganma (knowledge sharing) and dadirri (listening) can also be used to engage Aboriginal and Torres Strait Islander people throughout the evaluation process (box 6).

Box 6 Yarning and dadirri

Narratives and stories are a fundamental part of sharing Aboriginal and Torres Strait Islander ways of knowing.

Yarning is a qualitative research method that draws on Aboriginal and Torres Strait Islander people's cultural practice of storytelling as a way of conveying information and for relationship building (it can be one-on-one or in groups). It involves in-depth discussions to gain 'thick description' and offers a culturally safe place for Aboriginal and Torres Strait Islander people to freely talk about their experiences. Yarning positions the evaluator as a listener and learner in the data collection process and respects Aboriginal and Torres Strait Islander people as the authority of their knowledges. It involves reciprocal relationships and is a two-way process of learning and knowledge exchange.

Dadirri is an Aboriginal concept and practice of deep contemplation and listening. The word comes from the Ngan'gikurunggurr and Ngen'giwumirri languages (meaning 'listening'). Dadirri is known by different names in different Aboriginal cultures, but it typically translates as 'contemplation'. It requires researchers to continually be reflective of their relationships with others and as an approach is empowering as it 'enables working with indigenous people and allowing their voices to be heard' (West et al. 2012, p. 1585). It is an approach that:

- honours the integrity and fidelity of Aboriginal communities
- requires reciprocity in how researchers relate to, and behave with, Aboriginal people
- encourages researchers to reflect on their own assumptions or bias
- uses Aboriginal world views, in which learning is a process of listening and witnessing without judgement or prejudice, and of being responsible for self in relationship with others in the listening and learning process
- upholds Aboriginal world views so that the activity of learning introduces a responsibility to act with integrity and fidelity to what has been learnt (Atkinson 2002, p. 20).

Sources: Atkinson (2002, pp. 15, 20); Geia, Hayes and Usher (2013, p. 16); Leeson, Smith and Rynne (2016, p. 8); West et al. (2012, p. 1585).

Because narratives and stories are part of sharing Aboriginal and Torres Strait Islander ways of knowing, qualitative methods of collecting data can be preferred in some Aboriginal and Torres Strait Islander contexts. These methods can also counteract power differences (by strengthening the voices of those with less power).

However, qualitative methods of collecting information can be resource and time intensive to conduct rigorously, and findings may not be generalisable to broader populations. Qualitative studies are most rigorous when processes are systematic, clearly documented, transparent and auditable, when hypotheses are systematically tested, and when findings are reviewed and validated by participants.

Using different data collection methods can result in a more complete story and enhance evaluation rigour. Employing triangulation across methods and sources of data (including combining quantitative and qualitative methods) can make each method and data source more valuable and strengthen validity. Surveys, for example, can ask a mix of scale questions (quantitative) and open-ended questions (qualitative). Evaluators can also ask different people about the same issue (for example, an evaluation of teaching methods could get feedback from students, teachers and parents).

Mainstream policies and programs

For evaluations of mainstream policies and programs, the choice of approaches and methods will depend on the characteristics of the policy or program being evaluated, the number or proportion of Aboriginal and Torres Strait Islander people affected by the policy or program and what is already known about the impacts of the policy or program on Aboriginal and Torres Strait Islander people. Options for assessing the impact of mainstream policies and programs on Aboriginal and Torres Strait Islander people include:

- examining policy or program outcomes for Aboriginal and Torres Strait Islander people and non-Indigenous people to determine if there are differential impacts
- using focus groups or case studies to examine particular issues affecting Aboriginal and Torres Strait Islander people more closely
- undertaking a targeted evaluation that looks specifically at the impact of a mainstream policy or program on Aboriginal and Torres Strait Islander people.

Evaluation in practice: Evaluating to understand why there are poorer outcomes for Aboriginal and Torres Strait Islander people

An evaluation of a large mainstream employment program found that Aboriginal and Torres Strait Islander people had poorer employment outcomes than non-Indigenous people when participating in the program. The agency responsible for delivering the program wanted to understand why this was the case and planned a further evaluation focusing specifically on outcomes for Aboriginal and Torres Strait Islander people.

As a first step, the evaluation used administrative data from program participants to examine outcomes for different groups of Aboriginal and Torres Strait Islander people. This analysis found that some of the difference in employment outcomes between Aboriginal and Torres Strait Islander participants and non-Indigenous participants was because of higher caring burdens — carers typically had poorer outcomes from the program regardless of their Indigenous status. This led to a recommendation that further support be provided for all carers so that they could participate fully in the program.

The quantitative analysis also found that Aboriginal and Torres Strait Islander people had disproportionately poor employment outcomes in some regions. The evaluation adopted a case study approach, choosing several regions where outcomes were particularly good and particularly poor to study in more detail. Focus groups were held with service providers and participants in selected regions to better understand regional differences and to collect examples of good practice to share across regions. Surveys were also conducted with participants in case study regions to supplement administrative data on employment outcomes. Data collected were used to explore possible unintended consequences (both positive and negative) of the program.

Questions to consider when choosing an appropriate evaluation type, approach and method

- Should the evaluation be a formative, process or impact evaluation (or some combination)? Is an economic evaluation required?
- What evaluation types, approaches and methods will best answer the evaluation questions? Is there merit in using different approaches for the evaluation questions?
- What qualitative and quantitative methods could be used?
- What evaluation approaches and methods are most appropriate for Aboriginal and Torres Strait Islander participants? Do the methods used draw on Aboriginal and Torres Strait Islander expertise, perspectives, methodologies and knowledges?

Identifying and collecting data

The questions, scope and approach will determine what data are needed for an evaluation. A variety of different data sources can be used for evaluation (box 7). In some cases, evaluations may be able to draw on existing data sources. In others, data will need to be collected as part of the evaluation process. Evaluations often rely on both new and existing data. Data for evaluation should be fit-for-purpose and the costs of collecting data for evaluation balanced against the benefits.

Box 7 Sources of data

There are many different sources of data that can be collected for an evaluation. They include:

- *Administrative data* – data that are collected as part of administering a policy or program. They typically include registrations, transactions and general record keeping. Administrative sources, however, only include data on individuals who interact with a policy or program. For example, they can identify who is receiving unemployment benefits, but not what happens to them when they leave benefits. This weakness may be addressed through data linkage where multiple datasets are linked together.
- *Monitoring data* – monitoring data are performance management data that are collected as a policy or program is delivered. They generally cover all aspects of a policy or program including inputs, processes, outputs and outcomes. A particularly useful feature of monitoring data is that they provide information throughout the life of a policy. However, care needs to be taken to ensure they are of high quality. If data collection is seen as not useful or an administrative burden, there is little incentive to collect high-quality data.
- *Survey data* – large-scale surveys can be used for evaluation, including the Census of Population and Housing or the Longitudinal Study of Indigenous Children. Using existing survey data will be less expensive and will place less burden on participants than collecting new data, but the timing, representativeness or specific questions asked can limit their usefulness for evaluation. Surveys can also be used to collect data as part of evaluation or monitoring.
- *Qualitative data* – qualitative data will most likely be collected as new data, although some monitoring data can be qualitative. A variety of methods may be employed to collect qualitative data including interviews, focus groups, case studies and observations.

Source: adapted from HM Treasury (2020b).

Planning early for data collection allows data requirements to be built into a policy or program's delivery, for baseline and counterfactual data to be collected, for sources of relevant existing data to be identified, and for data approvals to be obtained.

Data should reflect the lives and experiences of Aboriginal and Torres Strait Islander people and be useful for answering evaluation questions that are important to Aboriginal and Torres Strait Islander people. Engaging with Aboriginal and Torres Strait Islander people on decisions about data collection, management and use will: ensure that collected data reflect what is important to Aboriginal and Torres Strait Islander people; improve trust in data collection and governance arrangements; ensure that data are collected, managed, used and stored in an ethical and culturally appropriate way; and support Aboriginal and Torres Strait Islander data sovereignty. Given the costs of collecting new data, existing data should be used wherever possible. However, most evaluations will involve at least some new data collection. Where Aboriginal and Torres Strait Islander people are a small percentage of participants in the policy or program, oversampling may be necessary to collect enough responses for robust results.

Evaluation in practice: Collecting baseline data

An evaluation of a literacy program in primary schools intends to compare outcomes before and after the program to determine its impact. An evaluation plan, including a data plan, was developed before the program was implemented in schools. The program logic helped the agency to identify the intended outcomes, as well as indicators that could be used to track progress and provide context for a future evaluation.

Some of the indicators were already collected through school testing surveys conducted each year. When rolling out the program in schools, the agency made sure that participants were asked for consent to use data from their regular school testing survey in the evaluation. Participants also completed a survey prior to starting the program to collect baseline data on the key indicators that the agency expected the program to influence but that were not included in the regular survey.

When an outcome evaluation was conducted several years later, the evaluators were able to access the earlier data to use as a baseline. The agency had clearly documented the data sources and collection methods so that the same methods could be replicated to collect data from participants after they had participated in the program.

Evaluation in practice: Oversampling to improve data quality

An evaluation of a mainstream health program wants to know about outcomes for Aboriginal and Torres Strait Islander participants, who make up around 5 per cent of all program participants. Data on outcomes will be collected via interviews with participants when they are discharged from hospital, and follow-up surveys sent to participants six months later.

In previous evaluations of similar programs, evaluators found that response rates for Aboriginal and Torres Strait Islander participants were lower than those for non-Indigenous participants, particularly for the follow-up survey. However, response rates were higher for Aboriginal and Torres Strait Islander participants who were referred to the program by an Aboriginal community controlled health service compared to those referred by a mainstream health service.

To ensure there are sufficient survey responses from Aboriginal and Torres Strait Islander participants, evaluators plan to oversample – this means including more Aboriginal and Torres Strait Islander people in the interviews and surveys than would be the case if respondents were chosen randomly. The number of Aboriginal and Torres Strait Islander people targeted for interviews will be decided based on response rates to previous surveys, and the need to be able to break down the sample of Aboriginal and Torres Strait Islander people by key demographic characteristics (including gender, age and geographical location).

To encourage higher response rates, evaluators will work with Aboriginal community controlled health services to design the interview and survey questions, and will provide additional training to interviewers on culturally safe data collection techniques. More information will be provided to participants on the benefits of responding to the survey, including being able to tailor services to better meet their needs. The evaluation team will also produce summary reports of program statistics by region so that service providers and community organisations can see how their region compares with others.

Questions to consider when identifying and collecting data

- What data are needed to answer the evaluation questions? What data are needed to produce credible results? Will monitoring data be useful for evaluation, or can it be modified to improve its usefulness? Can multiple sources of data be used to improve the rigour of results?
- Have Aboriginal and Torres Strait Islander people been engaged in planning, collecting and managing data used in the evaluation?
- Do outcome measures align with the intended objectives of the policy or program? Do outcome measures reflect what is important to Aboriginal and Torres Strait Islander people?
- Are data needed on policy or program inputs, the context in which the policy or program is being delivered or any unintended consequences of the policy or program?
- When should data on outcomes be collected?
- What existing data are available to be used? Is approval needed to use them?
- What additional data should be collected? How will data be collected? Do the data collection methods consider the needs, rights and cultural safety of respondents? Are the benefits from collecting additional data likely to exceed the costs of doing so?
- Are collected data of high quality? What are the limitations of the data? How can these be overcome?
- Are there clear processes in place for access to and release of data?

Ensuring ethical evaluation

Ethics in evaluation is concerned with good conduct – the conduct of evaluators, the integrity of the evaluation process, and the protection of participants. It also requires that Aboriginal and Torres Strait Islander peoples' values and cultural practices are respected during evaluations.

Evaluations should be conducted ethically, although not all require formal ethics review (box 8). However, all evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people undertaken by Australian Government agencies should be subject to a systematic assessment of ethical risks, which should be included in evaluation reports. Ethical conduct during the evaluation should also be clearly documented, and should reflect Aboriginal and Torres Strait Islander perspectives. This highlights the need for agencies to establish appropriate governance arrangements, where Aboriginal and Torres Strait Islander people share in decision making around how ethical risks are assessed.

32

Evaluations conducted under the Strategy should follow established guidelines for research with Aboriginal and Torres Strait Islander people, such as those produced by the Australian Institute for Aboriginal and Torres Strait Islander Studies and the National Health and Medical Research Council listed in the further reading section.

Agencies should ensure that evaluation budgets and timeframes are sufficient for evaluators to fulfil ethical requirements. This means allowing enough time for ethics risk assessment and formal ethics review, if required, as well as for evaluators to undertake meaningful engagement with Aboriginal and Torres Strait Islander people, build capability and report back to evaluation participants.

Box 8 When is formal ethics review necessary?

Research, including evaluation, can have harmful consequences for Aboriginal and Torres Strait Islander people, so centring Aboriginal and Torres Strait Islander people in evaluation processes is fundamental to ensuring ethical evaluation practices.

Formal review by an ethics committee will be required for some, but not all, evaluations conducted under the Strategy. The need for a formal review will depend on the scope and design of the evaluation and the risk of harmful consequences. However, even where formal ethics review is not required, agencies and evaluators should still be guided by ethical practices and research quality may be improved by getting advice from an ethics committee with experience in Aboriginal and Torres Strait Islander research.

Evaluations with features such as detailed non-standard data collection, comparisons between different groups, or experimental approaches (such as RCTs) are more likely to require formal ethics review. Where formal ethics review is required, it should be undertaken by an ethics committee with expertise in Aboriginal and Torres Strait Islander research and guided by recognised ethical guidelines for research involving Aboriginal and Torres Strait Islander people.

All evaluation projects should include a systematic assessment of ethical risks. Decisions on whether to request formal review by an ethics committee should be clearly documented. Ultimate responsibility for deciding whether formal ethics review is required should lie with agencies rather than with commissioned evaluators. However, agencies should take advice from commissioned evaluators, and ethical assessments should reflect Aboriginal and Torres Strait Islander perspectives.

Questions to consider about ensuring ethical evaluation

- Has the evaluation design and method been critiqued against Aboriginal and Torres Strait Islander ethics frameworks for research and evaluation?
- Has there been a systematic assessment of ethical risks associated with the evaluation? Do these assessments reflect the perspectives of Aboriginal and Torres Strait Islander people? Is formal review by an ethics committee required?
- If ethics review is required, does the ethics review body have expertise appropriate for evaluation involving Aboriginal and Torres Strait Islander people?
- Have sufficient time and resources been allowed for evaluators to meet ethical requirements?
- In what ways can the evaluation methods and processes be adapted to adopt ethical principles?
- How are participants' contributions of information, knowledges and time being respectfully and appropriately recognised and valued by evaluators and evaluation commissioners?
- How will the intellectual property from the evaluation be treated? Has this been negotiated with communities involved?
- How will consent be obtained from participants?
- Are there any confidentiality or privacy concerns? How will they be addressed?
- Are there any ethics or research protocols that need to be followed in the communities involved in the evaluation?

Interpreting and verifying the results

Interpreting results in an evaluation involves assessing the evidence (and its limitations) to give the evidence meaning and answer the evaluation questions. Ideally, during the planning stage when considering what approach and data to use, thought should be given to how the evidence will be used to make an assessment, and how Aboriginal and Torres Strait Islander people can contribute to interpreting evaluation results.

The interpretation of evidence should be systematic, logical and clearly articulated. Any limitations should be acknowledged with specific reference to how they may affect the validity and generalisability of findings. Alternative explanations for observed outcomes should be explored.

Interpreting results with Aboriginal and Torres Strait Islander people allows for interpretations and conclusions to incorporate Aboriginal and Torres Strait Islander knowledges and perspectives. This improves the credibility and usefulness of findings for policy and program designers and service providers, as well as for Aboriginal and Torres Strait Islander evaluation users. When interpreting the results, there may not always be consensus, and where there are differing views these should be reflected in the evaluation report.

Evaluation in practice: Workshopping evaluation findings with participants

An evaluation of a health program operating in urban areas collected quantitative and qualitative data via surveys, focus groups with program participants, and interviews with service providers.

Evaluators held workshops with service providers and evaluation participants to present preliminary results and to seek clarification and feedback on the key findings emerging from the evaluation. The workshop participants provided further context on the preliminary findings, which allowed the evaluators to better understand the reasons behind some conflicting results. The evaluators received useful feedback on what was important to participants, allowing the final evaluation report to be tailored to users' needs. The workshops also allowed service providers to get timely access to information that they could use to improve program delivery.

Whether an evaluation is conducted internally or using external evaluators, a key part of managing an evaluation is ensuring quality throughout the process. Ongoing quality assurance measures (such as peer review and steering committees) can help ensure that evaluation design, planning and delivery are properly conducted and conform to professional, ethical and analytical standards. Peer review, for example, allows experts independent of the policy or program and the evaluation to assess whether the evaluation questions, design and conduct were fit-for-purpose and that ethical requirements have been met. They can be conducted internally (by someone not connected to the policy or evaluation) or externally, by someone who is an expert in the subject area or in the evaluation approach or method used.

Questions to consider when interpreting and verifying the results

- Do the findings answer the evaluation questions? Are the findings supported by data and evidence? What limitations have been identified? Do they call into question the validity of the findings?
- Have alternative explanations been explored? Has any disagreement been acknowledged?
- Have Aboriginal and Torres Strait Islander people had opportunities to interpret the findings, or to provide feedback on the findings? Do Aboriginal and Torres Strait Islander people have a similar interpretation of the evidence and data to the evaluator?
- Are the data collection methods, assumptions and value judgements clearly documented and any limitations noted?
- How well might the findings apply to other policies, programs or contexts?
- What quality assurance processes are in place?

Further reading on evaluation planning, design and conduct

Indigenous Evaluation Strategy, Background Paper, chapters 4, 5, 7, 8 and 9.

Australian Evaluation Society nd, 'Resources', <https://www.aes.asn.au/resources.html>.

Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) 2020, *AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research*.

— 2020, *A Guide to Applying the AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research*.

Befani, B. 2016, *Choosing Appropriate Evaluation Methods: A Tool for Assessment and Selection*, Bond, London.

Better Evaluation nd, 'Evaluation practice in Aboriginal and Torres Strait Islander settings', <https://www.betterevaluation.org/en/themes/evaluation-practice-aboriginal-and-torres-strait-islander-settings>.

Better Evaluation nd, 'Manager's guide to evaluation', https://www.betterevaluation.org/managers_guide.

Centre for Epidemiology and Evidence 2019, *Commissioning Economic Evaluations: A Guide, Evidence and Evaluation Guidance Series*. Population and Public Health Division, NSW Ministry of Health, Sydney.

Department of the Prime Minister and Cabinet 2018, *Indigenous Advancement Strategy Evaluation Framework*.

HM Treasury 2012, *Quality in Policy Impact Evaluation: Understanding the Effects of Policy from other Influences (Supplementary Magenta Book Guidance)*.

— 2012, *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence (Supplementary Magenta Book Guidance)*.

— 2020, *Handling Complexity in Policy Evaluation, Magenta Book 2020 Supplementary Guide*.

— 2020, *The Magenta Book: Central Government Guidance on Evaluation*.

Hunt, J., 2013. *Engaging with Indigenous Australia – Exploring the Conditions for Effective Relationships with Aboriginal and Torres Strait Islander Communities*, Issues paper 5, Closing The Gap Clearinghouse.

Katz, I., Newton, B. J., Shona, B., and Raven, M. 2016, *Evaluation Theories and Approaches; Relevance for Aboriginal Contexts*, Social Policy Research Centre, UNSW, Sydney.

Kelagher, M., Luke, J., Ferdinand, A., Chamravi, D., Ewen, S. and Paradies, Y. 2018, *An Evaluation Framework to Improve Aboriginal and Torres Strait Islander Health*, The Lowitja Institute, Melbourne.

National Health and Medical Research Council 2018, *Keeping Research on Track II*.

— 2020, *Ethical Guidelines for Research with Aboriginal and Torres Strait Islander Peoples*.

Rogers, P., Hawkins, A., McDonald, B., Macfarlan, A., Milne, C. 2015, *Choosing Appropriate Designs and Methods for Impact Evaluation*, Department of Industry, Innovation and Science, Canberra.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., Befani, B. 2012, *Broadening the Range of Designs and Methods for Impact Evaluations*, DFID Working Paper 38, Department for International Development, London.

Stern, E., 2015, *Impact Evaluation: A Guide for Commissioners and Managers*, Bond, London.

World Health Organisation 2013, *WHO Evaluation Practice Handbook*, Geneva.

Section 4

Reporting and using evaluation findings

36

The aim of all evaluations conducted under the Strategy should be to influence decisions about policy and program design and implementation. This includes both making adjustments to existing policies or programs and adding to the evidence base to better inform future policy and program design. To do so, evaluation findings need be shared publicly in forms that evaluation users – including Aboriginal and Torres Strait Islander people – find useful. Transparency increases evaluation quality, accountability and opportunities for learning from findings.

To be useful, evaluation reports should be logical, clear and complete. That is, evaluation reports should:

- explain the policy or program being evaluated and the evaluation questions
- outline the methodology and data used
- present evidence to support findings
- clearly articulate limitations.

All evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people conducted under the Strategy should include a short plain English summary of the evaluation report (box 9).

Evaluations of mainstream policies and programs under the Strategy should report on the impacts and effectiveness of the policy or program for Aboriginal and Torres Strait Islander people.

Box 9 Plain English summaries of evaluation reports

To make evaluation findings accessible to all evaluation users, evaluation reports should include a short plain English summary, including details such as the:

- main findings of the evaluation (and depending on timing, the agency's response to the main findings)
- methodology and data used
- extent of engagement with Aboriginal and Torres Strait Islander people, organisations and communities and others affected by the policy or program during the evaluation
- whether a formal ethics review was undertaken and measures taken to ensure that the evaluation meets ethical standards
- limitations of the evaluation
- time and cost to complete the evaluation.

All evaluation reports should be published on agency websites, shared on the online Indigenous Evaluation Clearinghouse and disseminated to evaluation users. Where evaluation reports cannot be made public (such as in cases where there are concerns about privacy or culturally sensitive information) the reasons for not publishing should be included in a published summary.

Evaluation in practice: Using and sharing evaluation results when there are sensitivities about publishing a full evaluation report

An evaluation of a small-scale pilot program run in a remote community is undertaken to assess its effectiveness and whether there are broader lessons for a mainstream agency on how to deliver services to remote communities. Because the program, and the community, are small, publishing the evaluation's findings in full could inadvertently identify program participants and staff of the service provider involved.

The agency uses the unpublished evaluation findings to work with the service provider in the community where the pilot program took place to address issues identified during the evaluation. This includes developing a training program for the provider's staff that builds capability in areas identified as lacking by program participants. And the agency puts together an information sheet of the findings of the pilot that can be shared with service providers in other remote areas.

The agency also publishes a summary of the evaluation. The summary outlines the methods used during the evaluation, the limitations of the evaluation and a summary of the findings. The agency's response to the evaluation findings is also published on its website.

Collecting, disseminating and synthesising evaluation findings — collectively referred to as 'knowledge translation' — is a critical part of improving the use of evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people. Evaluation results are more likely to be used if they are tailored for different users, including Aboriginal and Torres Strait Islander people, who have contributed to the evaluation and/or are affected by the policy or program. Synthesising results — aggregating and reviewing findings across evaluations — can help to build a stronger and more compelling evidence base. Dissemination methods can include verbal feedback or information sessions, and summaries in languages other than English if appropriate. A knowledge translation plan should be agreed on with Aboriginal and Torres Strait Islander people, organisations and communities at the evaluation planning stage.

For transparency and accountability, agencies should publish a management response (it could be included as part of the evaluation report, or published separately). The management response should include an explanation about what was learnt, what was changed in response to the evaluation findings, and details of any further action that will be taken.

Evaluation in practice: Presenting evaluation findings for different audiences

A mainstream education program designed to assist secondary school students to remain in school until completion was evaluated to measure its impact and examine how effectively it was implemented in schools. As well as producing an evaluation report that documents the evaluation methods, data, findings and limitations in detail, the agency commissioned the evaluators to produce a range of fact sheets targeted at different audiences, including parents, teachers and students. The fact sheets highlighted examples of good practices that had been identified during the evaluation as contributing to better outcomes.

The findings of the evaluation were presented to agency staff and managers at a seminar. The seminar focused on both the detailed findings of the evaluation for the implementation and design of the program, as well as broader lessons that may be relevant to other program areas.

The evaluation was part of a broader program of research commissioned and conducted by the agency. The evaluation findings were incorporated into a synthesis report on factors affecting school retention that is published on the agency's online portal for sharing evidence and presented at conferences attended by teachers and education administrators.

Questions to consider when reporting and using evaluation findings

- Are evaluation methods and data described in detail in the report?
- Are the limitations of the evaluation analysis, data and results clearly documented in the report?
- Is the basis for the evaluation findings transparent?
- Is ethical conduct during the evaluation clearly documented?
- Are participants' contributions recognised in evaluation reports?
- Have any conflicts of interest, and how such conflicts were managed during the evaluation, been disclosed?
- Are differential effects observed across different populations? Does the evaluation report discuss the impacts and effectiveness of the policy or program for Aboriginal and Torres Strait Islander people?
- Is there a plain English summary of the evaluation findings?
- How will the findings be shared and disseminated? Are reports published and easy to find?
- Is appropriate attention given to addressing concerns about privacy, confidentiality and culturally-sensitive information?
- Will it be necessary to present the findings in different formats for different users?
- Have results been communicated back to the Aboriginal and Torres Strait Islander communities or organisations engaged in the evaluation?
- Are there mechanisms in place for knowledge translation? Have results been translated/interpreted in a culturally-sensitive manner?
- Has the evaluation report been submitted to the Indigenous Evaluation Clearinghouse?
- Has a management response to the findings of the evaluation been published?

Further reading on reporting and using evaluation findings

Indigenous Evaluation Strategy, Background Paper, chapter 7.

Better Evaluation nd, 'Report and support use', https://www.betterevaluation.org/rainbow_framework/report_support_use.

National Health and Medical Research Council 2018, *Keeping Research on Track II*.

Section 5

Building capability and a culture of evaluation

The Strategy will be most effective in an environment where agencies want to know how their policies and programs are performing, and are prepared to experiment, share learnings and use evaluation results in policy making. Building evaluation capability – including cultural capability – is also important for producing high-quality evaluations with credible results.

40

In practice, this means that evaluative thinking – a form of critical thinking where evaluative questions are asked as part of everyday business – is embedded within agencies. Agency staff need to be able to plan for and manage evaluations, and interpret and use findings. They also need to be able to commission external evaluators, including knowing how to select an evaluator, manage an evaluation and assess its quality. Agency staff need evaluation and cultural capability to ensure that commissioning processes facilitate high-quality evaluations that centre Aboriginal and Torres Strait Islander people. Agencies should take steps to strengthen their cultural capability at both the individual and organisational level, including developing strong and sustainable relationships with Aboriginal and Torres Strait Islander people.

Evaluation in practice: Building evaluation capability in a small agency

A small agency wants to build its evaluation capability so staff have the skills to implement the Indigenous Evaluation Strategy. Because the agency is small, it is unlikely to conduct evaluations itself, so capability building efforts are focused on improving the agency's skills at managing and commissioning evaluations.

The agency contracts an external evaluator to develop templates and checklists for program managers to use when planning evaluations and to provide training on evaluation for managers and staff. The training involves a coaching approach where program managers bring their own program logic or evaluation plan to the session and other program managers provide feedback.

The agency works on building cultural capability, first by requiring staff involved in commissioning evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people to complete the Australian Institute for Aboriginal and Torres Strait Islander Studies' Core Cultural Learning foundation course. The agency also engages an Aboriginal and Torres Strait Islander organisation to provide coaching to staff to ensure that evaluation plans and program logics are culturally appropriate.

Program managers and staff involved in evaluation are also encouraged to seek out other opportunities to develop their evaluation or cultural capability skills by engaging with communities of practice, professional organisations, Aboriginal and Torres Strait Islander community controlled organisations and other evaluators within the Australian Public Service.

Transparency is vital for improving agencies' evaluation culture. As well as publishing evaluation reports, other documents about evaluation policies — including frameworks, strategies and rolling Three Year Evaluation Forward Work Plans — should be published on agency websites so it is easy for stakeholders to see agencies' commitments on evaluation.

Embedding evaluation in the policy cycle is key to building evaluation culture and capability. Adopting measures discussed earlier in this guide — such as building on existing evidence when designing new policies and programs, and planning early for evaluation — can improve the relevance, credibility and usefulness of evaluation to decision makers.

But it is also critical that evaluation findings feed upward to decision makers and translate to action — be it the establishment of a new or reformed initiative, the ending of a policy or program, or (in most cases) incremental change in light of the lessons learned. Evaluation findings should also inform the strategic priority setting of government, including decisions about the best and most effective allocation of resources.

Strengthening capability among Aboriginal and Torres Strait Islander people, organisations and communities

Adopting the Strategy's principles-based approach when conducting evaluations is expected to increase demand for culturally capable evaluators. There should also be more opportunities for Aboriginal and Torres Strait Islander people to be engaged in, participate in, and provide advice on evaluation design, data collection and the interpretation of evaluation results. After all, agencies will only be able to effectively centre Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges in evaluation if Aboriginal and Torres Strait Islander people, organisations and communities have the time and resources to engage, partner or lead in evaluation.

While many Aboriginal and Torres Strait Islander people, communities and organisations are already well placed to do this, agencies can support and strengthen Aboriginal and Torres Strait Islander capability by:

- developing relationships with Aboriginal and Torres Strait Islander people and organisations that can provide insights on priorities, perspectives and knowledges
- ensuring adequate resources, time and space are allocated for Aboriginal and Torres Strait Islander people and organisations to effectively engage in evaluations and evaluation processes
- working with Aboriginal and Torres Strait Islander people to build evaluation capability, including engaging Aboriginal and Torres Strait Islander evaluators and community researchers and providing training and development opportunities
- building agencies' cultural capability — both at an organisational and individual level — to work effectively with Aboriginal and Torres Strait Islander people.

Questions to consider when building capability and a culture of evaluation

- Does the agency value evaluative thinking?
- Does the agency have an evaluation policy or framework? Is it publicly available? Does the agency have a central unit to plan, and share lessons from, evaluation?
- Are there processes in place for incorporating evaluation findings into decisions about policy or program changes? How are evaluation findings shared within the agency?
- Does the agency have processes in place to build the capability to use and respond to evaluation findings? Do evaluation teams have the capacity, technical and cultural capability, and experience to deliver high-quality evaluation?
- Does the agency provide opportunities or encourage staff to pursue opportunities to strengthen their technical and cultural capability? Do evaluation teams have the capability to incorporate Aboriginal and Torres Strait Islander knowledges into their evaluative thinking, including by asking questions that matter for Aboriginal and Torres Strait Islander people, and collecting and using evidence in culturally safe ways?
- Does the agency have processes in place to build capability among Aboriginal and Torres Strait Islander evaluators, organisations and communities?
- Is the agency open to review and feedback from Aboriginal and Torres Strait Islander people and organisations about its evaluation practices?
- Is it clear what the findings will be used for and what decisions they will feed into?

Further reading on building capability and a culture of evaluation

Indigenous Evaluation Strategy, Background Paper, chapters 7 and 8.

Better Evaluation nd, 'Strengthen evaluation capacity', https://www.betterevaluation.org/rainbow-framework/manage/strengthen_evaluation_capacity.

References

- Atkinson, J. 2002, 'Dadirri: Listening to one another', *Trauma Trails: Recreating Song Lines*, ProQuest Ebook Central, pp. 6–22.
- Breckon, J. and Mulgan, G. 2015, *Better Public Services Through Experimental Government*, Alliance for Useful Evidence, United Kingdom.
- Geia, L.K., Hayes, B. and Usher, K. 2013, 'Yarning/Aboriginal storytelling: Towards an understanding of an Indigenous perspective and its implications for research practice', *Contemporary Nurse*, vol. 46, no. 1, pp. 13–17.
- Haynes, L., Service, O., Goldacre, B. and Torgerson, D. 2012, *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*, Behavioural Insights Team, Cabinet Office, United Kingdom.
- HM Treasury 2011, *The Magenta Book: Guidance for Evaluation*, London.
- 2020a, *Magenta Book 2020 - Supplementary Guide: Handling Complexity in Policy Evaluation*, London.
- 2020b, *The Magenta Book: Central Government Guidance on Evaluation*, London.
- Joint Council on Closing the Gap 2020, *National Agreement on Closing the Gap*, Canberra.
- Leeson, S., Smith, C. and Rynne, J. 2016, 'Yarning and appreciative inquiry: The use of culturally appropriate and respectful research methods when working with Aboriginal and Torres Strait Islander women in Australian prisons', *Methodological Innovations*, vol. 9, pp. 1–17.
- West, R., Stewart, L., Foster, K. and Usher, K. 2012, 'Through a Critical Lens: Indigenist Research and the Dadirri Method', *Qualitative Health Research*, <https://journals.sagepub.com/doi/10.1177/1049732312457596> (accessed 8 October 2020).

About the artist

44



River of Knowledge by Luke Penrith



Luke Penrith's ancestry is connected through the Wiradjuri, Wotjobaluk, the Yuin and the Gumbaynggirr Nation. His passion is mentoring and nurturing Indigenous Australian job seekers and supporting Aboriginal Businesses. Lore, culture and heritage are paramount to Luke. His art reflects what he sees, hears and can smell and touch; he is a modern contemporary Aboriginal Artist living in Brungle NSW. Luke's bloodlines are connected through the rivers, the mountains, the coastline and the plains.

www.pc.gov.au/ies