

Northern Territory Government Program Evaluation Toolkit

Document title	Northern Territory Government Program Evaluation Toolkit
Contact details	Department of Treasury and Finance
Approved by	
Date approved	
Document review	Annually
TRM number	

Version	Date	Author	Changes made
1.0	April 2020	Assistant Director Program Evaluation	First version
2.0	August 2020	Assistant Director Program Evaluation	Incorporating feedback from the Northern Territory Government Program Evaluation Community of Practice

Acronyms	Full form
CMC	Department of the Chief Minister and Cabinet
DTF	Department of Treasury and Finance
PEU	Program Evaluation Unit
NHMRC	National Health and Medical Research Council
OCPE	Office of the Commissioner for Public Employment

Contents

How to use this toolkit	5
A whole of government approach to evaluation	6
Program evaluation framework.....	7
Treasurer’s Direction on Performance and Accountability	8
Program evaluation, organisational reviews and audits– what is the difference?.....	8
Integrating evaluation into program design and budget development	10
Evaluation as part of the program cycle.....	10
Evaluation as part of budget development.....	12
Evaluation overview.....	12
Evaluation work plan.....	12
Sunset clauses.....	12
Evaluating strategically	13
Developing the annual program master list.....	14
Prioritising evaluations for rolling schedule.....	14
Getting existing programs ready for evaluation	16
Stages for planning and commissioning an evaluation	19
1. Complete the evaluation overview	20
1.1. What is the program aiming to achieve?.....	20
1.2. How will the program achieve this?	20
1.3. What external factors may also influence the program’s outcomes?.....	21
1.4. How will the program’s success be measured?.....	21
1.5. What evaluations will be required in the first five years of the program?	21
1.6. Other evaluation-related resource requirements.....	22
2. Complete the evaluation work plan	23
2.1. Introduction.....	23
2.1.1. Who should develop the evaluation work plan?.....	23
2.2. Program overview.....	23
2.2.1. Needs assessment	24
2.3. Evaluation overview	24
2.4. Evaluation roles and responsibilities	24
2.5. Evaluation methodology	27
2.5.1. Program logic.....	27
2.5.2. Evaluation questions.....	36
2.5.3. Types of evaluation	39
2.5.4. Data matrix	45
2.6. Budget and resourcing	45
2.6.1. Evaluation on a shoestring.....	48
2.7. Stakeholder engagement.....	49

2.8. Ethical considerations	49
2.9. Evaluation risks	51
2.10. Reviewing the evaluation work plan.....	53
2.10.1. Technical review of the evaluation design.....	53
2.10.2. Review of the design by the evaluation management structure.....	53
2.10.3. Review the program logic.....	54
3. Engage the evaluation team.....	55
3.1. When to engage the evaluation team.....	55
3.2. Prepare the request for tender.....	55
3.3. Select the evaluation team.....	56
3.4. Prepare the contract.....	56
3.5. Support the evaluation team	57
4. Manage the implementation of the evaluation	58
4.1. Collecting the data	58
4.2. Analysing the data.....	58
4.2.1. Evidence synthesis	59
5. Guide production of a quality evaluation report.....	60
6. Disseminate results and support use of evaluation.....	61
6.1. Communicating evaluation results	61
6.2. Responding to and using evaluation findings.....	61
6.2.1. Program reality checks	62
6.2.2. De-implementation	63
6.3. Sharing lessons learned.....	63
Templates	64
Glossary	65

How to use this toolkit

This toolkit is for Northern Territory Government (Territory Government) employees who are responsible for policy development, program design and program evaluation. It has been compiled by the [Program Evaluation Unit \(PEU\)](#) at the Department of Treasury and Finance (DTF) as part of a whole of government approach to program evaluation. It provides guidance, resources and templates to help managers integrate evaluation planning into program design, get a program ready for evaluation, commission and manage an external evaluation or lead an internal evaluation. It will help evaluation teams (either internal or external) meet the expectations for evaluation under the [program evaluation framework](#).

The toolkit is structured to guide program managers and policy officers through completing two key documents: the [Evaluation overview](#) and the [Evaluation work plan](#). The key underlying resources are the [Manager's guide to evaluation](#) from the [BetterEvaluation](#) website and the [NSW Evaluation Toolkit](#).

There are many different approaches to evaluation and this toolkit is not intended to be exhaustive. It will be updated regularly in response to user feedback. If you would like to suggest a change or if you need further information on how to use the toolkit, please email DTF.ProgramEvaluation@nt.gov.au.

A whole of government approach to evaluation

The Territory Government is focused on improving evidence-based decision-making as part of [A plan for budget repair](#). This will be supported by a whole of government approach to evaluation in the Territory to help drive a culture of continuous improvement across government.

Program evaluation aims to improve government services to achieve better outcomes for Territorians. This is important because sometimes programs don't work in reality and may even cause unintended harm.¹ Program managers need to know whether their programs are helping people and whether implementing changes could help more people within the same budget. Without evaluation, there is a risk that poorly performing programs continue without change, slowing progress towards achieving desired outcomes and potentially wasting taxpayers' money.

When an evaluation shows a program is not working well, managers can [use the evaluation findings](#) to improve the program by either modifying the existing program or taking a new approach. Each evaluation is an opportunity to learn by either demonstrating what works well or what does not. Over time, evaluations build an evidence base of what works in the Territory and foster a culture of continuous improvement.

Central oversight is critical to developing a strategic whole of government approach to evaluation and strengthening evaluation culture.² A centralised approach to program evaluation supports:

- a consistent standard of evaluation across agencies
- an ability to identify systemic issues across government
- capacity to set strategic priorities for and identify gaps in evaluation
- accountability for multi agency and whole of government programs
- coordinated capability building, resourcing, data collection, reporting and evaluative effort
- a centralised repository of evaluations to enhance continuous learning and quality improvement.

Under the Territory's approach, evaluation activity will continue to be undertaken primarily by the agency delivering the program (this may include using external experts commissioned by the agency). This is necessary to maintain a close link between the evaluation and the program area with relevant subject matter knowledge and experience.

Evaluation activity will be overseen, coordinated and supported by the PEU within DTF, supported by the Department of the Chief Minister and Cabinet (CMC), the Office of the Commissioner for Public Employment (OCPE) and the Department of Corporate and Digital Development.

¹ For example, [Scared Straight](#) and other similar programs involve organised visits to prison by juvenile delinquents or children at risk for criminal behaviour. Programs are designed to deter participants from future offending through firsthand observation of prison life and interaction with adult inmates. A recent Cochrane review found that these programs fail to deter crime and actually lead to more offending behaviour. See Petrosino et. al., 2013, 'Scared Straight and other juvenile awareness programs for preventing juvenile delinquency', accessed October 2020, <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD002796.pub2/full>.

² Evaluation and learning from failure and success, ANZSOG, 2019.

Program evaluation framework

The [Program evaluation framework](#) integrates evaluation into the government's policy and budget development processes. The framework aims to improve transparency and accountability, and encourage better use of Territory Government funds, by:

- ensuring new programs and extensions to existing programs have identified goals and objectives that are achievable and measurable, or include actions to develop measurement as part of the program
- ensuring new programs and extensions to existing programs have an evaluation strategy
- applying [sunset provisions](#) to new programs (or extensions to existing programs), where the decision for further funding is informed by evaluation outcomes
- establishing a rolling schedule of evaluations to ensure existing programs are evaluated over time
- providing a clear mandate for agencies to evaluate their programs and target their investments
- outlining expected evaluation principles and standards
- providing government with clear advice about the costs and benefits of evaluation (including data collection and analysis) to help inform evaluation decisions
- establishing a protocol for policy and program officers to plan for evaluation across the program lifecycle (with a step-by-step guide in the program evaluation toolkit)
- establishing a tiered system of evaluations to ensure evaluation is proportionate to the cost, risk and complexity of a program
- describing how the Territory Government can build evaluation capability within the Northern Territory Public Sector and foster a culture of continuous improvement
- outlining how the Territory Government will measure progress in implementing the framework.

Territory Government agencies must use the framework and toolkit to help plan, commission and use evaluations. The framework and toolkit may also provide useful guidance for Territory Government service delivery partners and external evaluators of Territory Government programs.

The [program evaluation framework](#) is underpinned by 10 best practice evaluation principles:³

1. **Build evaluation into program design.** Plan the evaluation as part of program design to ensure clearly defined objectives and measurable outcomes prior to commencement.
2. **Base the evaluation on sound methodology.** Adopt a best practice evaluation methodology that is commensurate with the program's size, significance and risk.
3. **Allocate resources and time to evaluate.** Include provision for the required evaluation resources and timeframes when planning and budgeting for a program. Ensure evaluation findings are available when needed to support key decision points.
4. **Use the right mix of expertise and independence.** Use evaluators who are experienced and independent from program managers, but include program managers in evaluation planning.
5. **Ensure robust governance and oversight.** Establish governance processes to ensure programs are designed and evaluated in accordance with this framework, including meeting reporting requirements.

³ Adapted from the NSW Government Program Evaluation Guidelines.

6. **Be ethical in design and conduct.** Carefully consider the ethical implications of any evaluation activity, particularly collecting and using personal data, and any potential impacts on vulnerable groups.⁴
7. **Be informed and guided by relevant stakeholders.** Listen to stakeholders, including program participants, government and non-government staff involved in managing and delivering the program, and senior decision makers.
8. **Consider and use evaluation data meaningfully.** Include clear statements of findings, recommendations or key messages for consideration in evaluation reports. Use reports to inform decisions about program changes.
9. **Be transparent and open to scrutiny.** Disseminate key information to relevant stakeholders, including methodologies, assumptions, analyses and findings.
10. **Promote equity and inclusivity.** Harness the perspectives of vulnerable groups during evaluations, to enable fair and socially just outcomes.

Treasurer's Direction on Performance and Accountability

Treasurer's Directions are mandatory requirements that specify the practices and procedures that must be observed by Accountable Officers in the financial management of their agencies ([Financial Management Act 1995](#)). DTF is currently developing a Performance and Accountability Treasurer's Direction that will set out the minimum requirements for all Territory Government agencies for:

- planning objectives and actions
- managing or delivering services
- performance reporting
- reviewing and evaluating outcomes.

Guidance on performance and accountability will be provided to agencies to assist them in complying with the requirements of the Performance Accountability Treasurer's Direction.

Program evaluation, organisational reviews and audits— what is the difference?

For the purposes of the [program evaluation framework](#):

Evaluation is:

*"A systematic and objective process to make judgements about the merit or worth of one or more programs, usually in relation to their effectiveness, efficiency and appropriateness."*⁵

Monitoring is:

*"A management process to periodically report against planned targets or key performance indicators that, for the most part, is not concerned with questions about the purpose, merit or relevance of the program."*⁶

⁴ In some circumstances, formal review and approval from an ethics committee certified by the National Health and Medical Research Council may be required. See Ethical considerations for further information.

⁵ NSW Government Evaluation Framework August 2013. This definition applies more to outcome and impact evaluations rather than process evaluations that tend to focus more on monitoring.

⁶ NSW Government Evaluation Framework August 2013.

While there are a number of different approaches to evaluation,⁷ the [Program evaluation framework](#) is based on three types⁸ linked to a program's lifecycle:

1. **Process evaluation:** considers program design and initial implementation.
2. **Outcomes evaluation:** considers program implementation and short to medium-term outcomes.
3. **Impact evaluation:** considers medium to long-term outcomes, and whether the program contributed to the outcomes and represented value for money.

The different types of evaluation are covered in more detail in [section 2.5.3. Types of evaluation](#).

Program evaluation is most effective when it is complemented by other activities which collect information and assess performance, including:

- **Organisational reviews:** consider an agency's entire budget, ensure expenditure is aligned to government priorities and services are being provided efficiently. Implementation of a rolling schedule of organisational reviews was recommended in [A plan for budget repair](#). DTF is currently developing an Agency Organisational Review Framework.
- **Program reviews:** typically quick, operational assessments of a program to inform continuous improvement.⁹
- **Research:** closely related to evaluation, but can ask different types of questions that may not be related to the merit or worth of a program.¹⁰
- **External audits:** undertaken by an independent auditor. Reviews records supporting financial statements.¹¹
- **Internal audits:** undertaken by agencies. Reviews governance, risk management, and control process according to a risk-based need.¹²
- **Performance Management System (PMS) audits:** undertaken by the Northern Territory Auditor-General. Consider whether appropriate systems exist and are effective in enabling agencies to manage their outputs.¹³
- **Performance audit:** undertaken by an Auditor-General but not currently within the scope of the Northern Territory Auditor-General. Examine the economy, efficiency and effectiveness of government programs and organisations.^{14,15}
- **[Independent Commissioner Against Corruption](#) audit or review:** investigate practices, policies or procedures of a public body or public officer to identify whether improper conduct has occurred, is occurring or is at risk of occurring.

Further definitions are in the Glossary.

⁷ For information about other evaluation types, please see the BetterEvaluation website.

⁸ Further information on these three evaluation types are in sections 3.2.1 to 3.2.3.

⁹ <https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/evaluation-in-the-nsw-government/>.

¹⁰ <https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/evaluation-in-the-nsw-government/>.

¹¹ The Institute of Internal Auditors Australia, Internal Audit Essentials, 2018.

¹² *ibid* 11.

¹³ Northern Territory Auditor-General's Office Annual Report 2017-18.

¹⁴ Black, M., 2018, Strategic Review of the Northern Territory Auditor-General's Office.

¹⁵ A Guide to Conducting Performance Audits, Australian National Audit Office, 2017,

<https://www.anao.gov.au/work/corporate/guide-to-conducting-performance-audits>.

Integrating evaluation into program design and budget development

Evaluation as part of the program cycle

For the purposes of the program evaluation framework, a program is broadly defined as:

“A set of activities managed together over a sustained period of time that aim to deliver an outcome for a client or client group”.¹⁶

Essentially, programs deliver government functions.¹⁷ A program can also include related government spending on a single intended outcome.¹⁸ A program can be as broad as all government expenditure to reduce cost of living pressures, or as specific as a single social concession.¹⁹ Education, health and policing services are deemed functions, not programs. Further, as a general rule, the [Program evaluation framework](#) will not apply to infrastructure and information and communications technology projects (which are covered by separate review processes) or externally funded programs.

The term ‘program’ is sometimes used interchangeably with project, service, initiative, strategy or policy. In practice, programs vary in size, duration and structure, and may span multiple agencies. Whole of government programs can be large and significant strategies, action plans or frameworks that encompass multiple agencies and locations, and comprise many agency-level programs, sub-programs and projects (Figure 1). Regardless of program size, when designed and conducted well, evaluation can yield useful evidence about the effectiveness of programs.

A strategic approach to evaluation (see Evaluating strategically) includes evaluations at several levels. For example, evaluating at the whole of government level to identify how different components of a strategy work together to achieve outcomes, and evaluating at the project level to examine specific aspects of a program.

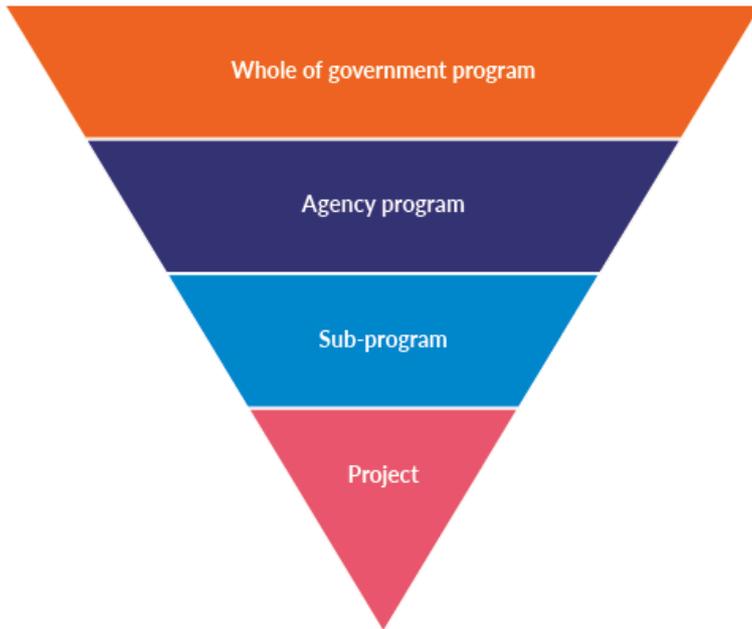
¹⁶ NSW Government Evaluation Framework August 2013.

¹⁷ Program Evaluation: Sunset Clauses, Agency Guide, Government of Western Australia, 2014.

¹⁸ Program Evaluation: Sunset Clauses, Agency Guide, Government of Western Australia, 2014.

¹⁹ Program Evaluation: Sunset Clauses, Agency Guide, Government of Western Australia, 2014.

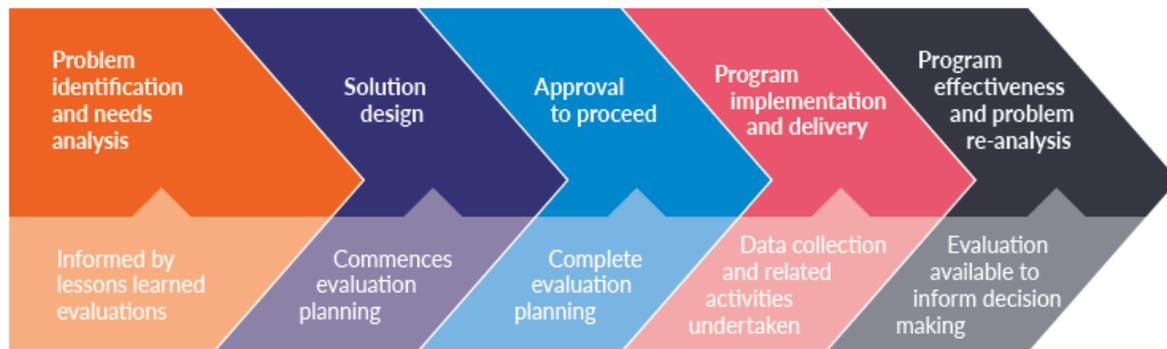
Figure 1: Program hierarchy



Integrating evaluation into the program lifecycle ensures cost-effective evaluation is delivered in time to support key decision making points (Figure 2). Planning for evaluation should start at the program design stage so all stakeholders understand the key performance indicators the program will be assessed against and how and when evaluation will occur. Early planning also ensures data requirements are identified prior to commencement and lessons learned from previous evaluations can be used effectively.

Figure 2: Integrating evaluation into the program lifecycle

Program cycle



Evaluation process

Further guidance on good practice policy development is available on the [APS Policy Hub](#).

Evaluation as part of budget development

Integrating evaluation into the budget process allows governments to make better use of resources.²⁰ The [program evaluation framework](#) integrates evaluation into the Territory Government's budget process through an evaluation overview as part of the Cabinet submission template,²¹ an evaluation work plan for approved programs, sunset clauses and a rolling schedule of evaluations.

Evaluation overview

An evaluation overview is required as part of the Cabinet submission process for programs requesting funding of \$1 million or more in a year. The overview should be a concise summary of the key outcomes the program is trying to achieve and how success will be measured (further info at [section 1. Complete the evaluation overview](#)).

A full evaluation work plan will be required within six months if the program is approved to proceed.

Evaluation work plan

The detailed evaluation work plan²² outlines future evaluation activity for a particular program over the next five years. The template requires agencies to consider:

- the program's theory of change (the program logic model)
- key evaluation questions, indicators, and data sources (the question bank and data matrix)
- appropriate types and timing of future evaluations (combined with the logic model and data matrix to form the program's evaluation work plan).

See [section 2. Complete the evaluation work plan](#) for further.

Sunset clauses

Programs subject to sunset clauses are funded for a finite period, with the decision for further funding (either wholly or in part) informed by an evaluation.

Currently, agencies' annual recurrent budgets represent an accumulation of funding decisions by a variety of governments over time. Without program evaluation, the Budget Review Subcommittee and Cabinet have little visibility of:

- the effectiveness of existing programs
- the applicability of existing programs to the current policy context
- whether delivery models for the existing program provide the best value for money.²³

²⁰ Building Better Policies, Chapter 6 Monitoring and Evaluation Systems and the Budget, World Bank, 2012.

²¹ In the absence of exceptional circumstances, all submissions seeking additional funding must be considered as part of the Budget development process under the Northern Territory Government's *Charter of Budget Discipline*.

²² The [Evaluation work plan template](#) has been adapted from the Evaluation strategy template with permission from the Evaluation Office at the Commonwealth Department of Industry, Science, Energy and Resources. Publicly available at the APS Policy Hub <https://www.policyhub.gov.au/resources/getting-evaluation-ready-process-diis>.

²³ Program Evaluation: Sunset Clauses, Agency Guide, Government of Western Australia, 2014.

A sunset clause is a built-in decision point for government. Unless otherwise directed by Cabinet, funding for new programs (or extensions of existing programs) that impact the Territory Government's operating balance by \$1 million or more in a year will be subject to an initial five-year sunset clause. This ensures that ongoing funding for programs is informed by evaluation.

Sunset clauses have been included in the Cabinet template and [handbook](#) and will be supported by a:

- Sunset clause guide for agencies (in development)
- Sunset clause guide for Treasury analysts (in development).

Evaluating strategically

Monitoring and evaluation requires the commitment of resources. If an evaluation does not provide decision-makers with meaningful information it reduces resources available for program implementation.²⁴ Therefore, it is necessary to balance the cost of evaluation and the risk of not evaluating, noting that sometimes monitoring will be sufficient. While outcome and impact evaluations are important, well-designed data collection, program monitoring and process evaluations can help refine programs over time with minimal cost.²⁵

Agencies and program managers will need to take a strategic approach in determining appropriate evaluation scopes, designs and resourcing requirements. For some programs, evaluation could simply involve routine assessment of activities and outputs built into program reporting, while for others evaluation will need to be more comprehensive and assess whether the program is appropriate, effective and efficient.²⁶

Although it is not feasible, cost effective or appropriate to fully evaluate all Territory Government programs, some level of monitoring and review should be considered for all programs.

For whole of government programs, or programs with multiple components, it may be necessary to evaluate components separately as well as collectively, considering questions such as:

- which program initiatives are providing the greatest impact
- which elements of program delivery are most effective in generating desired outcomes
- is greater impact achieved when specific strategies are combined into a package of initiatives
- in what contexts are mechanisms of change triggered to achieve desired outcomes?²⁷

Evaluations should aim to achieve the highest rigour for the lowest cost by:

- incorporating evaluation planning at the initial program design stage
- collecting the required data for monitoring and evaluation throughout program implementation and aligning this to existing data collections where possible
- using a tiered approach that prioritises evaluative effort (see *Prioritising evaluations for rolling schedule* for further information).

²⁴ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

²⁵ APS Review, *Evaluation and learning from failure and success*.

²⁶ Queensland Government Program Evaluation Guidelines.

²⁷ Queensland Government Program Evaluation Guidelines.

Developing the annual program master list

To balance evaluative effort against the potential benefit, agencies need to review their existing stock of programs and prioritise evaluations.

A good practice starting point is a program master list,²⁸ which is designed to capture all current Territory Government-funded programs to help prioritise evaluations. It identifies the extent to which existing programs have been evaluated and the proposed timing of any future evaluation.²⁹ Agencies are asked to complete the Program master list template annually as part of the Budget development process. Integrating the program master list into the Budget development process ensures new Cabinet submissions are considered within the context of existing government programs and the available evidence base from completed evaluations.

Agency activities that are not captured as part of the program master list will still be scrutinised as part of broader organisational reviews.

Prioritising evaluations for rolling schedule

To help manage and prioritise evaluations, agencies are required to prepare multi-year rolling evaluation schedules that are reviewed annually by the Budget Review Subcommittee of Cabinet. In addition to evaluating new programs in accordance with the approved evaluation overview, the schedules will be expected to include a list of existing programs³⁰ planned for evaluation, including the tier and expected evaluation timeframe.

The evaluation schedule for each agency should be aligned to agency corporate planning cycles and internal decision-making processes and should be developed in consultation with DTF and CMC.

A whole of government evaluation schedule will be compiled by DTF and submitted to the Budget Review Subcommittee of Cabinet for approval along with an annual summary of evaluation findings for the previous year. An example of an [evaluation schedule](#) is available from the Commonwealth Department of Industry, Innovation and Science.

Table 1 provides a guide to prioritising programs for the rolling schedule of evaluation. A best-fit approach should be used to categorise programs (that is, a program does not need to satisfy every characteristic to fall into a particular tier).

²⁸ In 2016, the New South Wales Auditor-General undertook a [performance audit of the NSW Government' program evaluation initiative](#). The audit set out the good practice model expected from each agency to prepare an evaluation schedule including a master list of all current agency programs with their tier ranking and linkage to government priorities. NSW Auditor-General's Report to Parliament, Implementation of the NSW Government's program evaluation initiative, 2016 accessed October 2020.

²⁹ APS Review, Evaluation and learning from failure and success; and 2016 NSW Auditor-General's Report to Parliament, Implementation of the NSW Government's program evaluation initiative.

³⁰ Evaluating existing programs can be complex and expensive, particularly where the data required to answer basic evaluation questions has not been collected. The section 'Getting existing programs ready for evaluation' has further guidance.

Table 1: A guide to program tiers, evaluation types and timing³¹

Tier	Characteristics of program	Evaluation type		
		1 year	2 years	3–5 years
4	<p>Priority: strategic priority for government</p> <p>Program accountability: Cabinet or Cabinet subcommittee</p> <p>Funding: significant government/agency funding</p> <p>Risk: high risk (either to government or the community)</p> <p>Scope: multiple government agencies and/or multiple external delivery partners</p> <p>Other factors: lack of evidence base, major external reporting requirements (for example, Commonwealth), innovative approach</p>	Process	Outcomes	Impact
3	<p>Priority: strategic priority for agency</p> <p>Program accountability: portfolio Minister(s)</p> <p>Funding: significant agency funding</p> <p>Risk: moderate to high risk</p> <p>Scope: multiple government agencies and/or external delivery partners</p> <p>Other factors: lack of evidence base, internal reporting and evaluation requirement</p>	Process	Outcomes	Impact or outcomes
2	<p>Priority: named in department agency strategic plan</p> <p>Program accountability: agency chief executive</p> <p>Funding: moderate agency funding</p> <p>Risk: low to moderate</p> <p>Scope: responsibility of single agency, may involve external delivery partners</p> <p>Other factors: limited evidence base, internal reporting and evaluation requirement</p>	Process		Outcomes
1	<p>Priority: low or emerging strategic priority for agency</p> <p>Program accountability: business unit within agency</p> <p>Funding: limited agency funding</p> <p>Risk: low</p> <p>Scope: single agency, may involve external delivery partners</p> <p>Other factors: local delivery similar to other successful programs</p>	Process		Process

³¹ Adapted from DIIS evaluation Strategy 2017–2021 and NSW Guidelines and WA Guidelines.

The appropriate evaluation types and timing will need to be determined on a case-by-case basis to ensure the overall evaluation approach is fit for purpose.

When prioritising evaluations, agencies should give priority to:

- Tier 3 and Tier 4 programs (as per the program tiering in Table 1)
- programs that have not previously been evaluated
- programs for which evaluation is required by Cabinet (for example, in line with an evaluation overview approved by Cabinet).

Tier 3 and Tier 4 programs should be prioritised for evaluation and would usually be expected to go through process, outcome and/or impact evaluations over the program lifecycle.

The prioritisation of Tier 1 and 2 programs is at the discretion of agencies but should be influenced by how they fit into higher tier programs (if applicable). In particular, agencies should consider evaluating small programs if they will be used to inform decisions about whether to roll out the program to a wider area and/or client group (such as a pilot or a trial) or will be used as evidence of another program's effectiveness.³²

Getting existing programs ready for evaluation

The [program evaluation framework](#) emphasises the importance of planning for evaluation and data capture at the program design stage. However, existing programs without an evaluation work plan should also be periodically reviewed because:

- the bulk of government spending relates to legacy programs
- the nature and outcomes of these programs may have evolved or drifted away from their initial rationale or purpose over time
- legacy programs have the potential to become embedded or institutionalised by the participants or community in ways that may have significantly affected their outcomes.

Evaluating programs that were not designed with evaluation in mind can be complex and expensive.³³ Completing an evaluation work plan (see [section 2. Complete the evaluation work plan](#)) can assist to get programs 'evaluation ready'. An important first step is clarifying what the program aims to achieve and how it tries to achieve this by developing a program logic (further information in [section 2.5.1. Program logic section](#)).³⁴ Developing a program logic for an existing program can be an uncomfortable process. Stakeholders may disagree about how a program works or even what it is aiming to achieve. The program logic may reveal that the program is not well formulated or that it includes dubious assumptions. To genuinely support learning and improvement, developing a program logic should not try to rationalise past program decisions. Instead, developing a program logic for an existing program should be an opportunity to question, debate and learn. This process can help agencies identify unnecessary activities and make space for more important ones.³⁵

³² Queensland Government Program Evaluation Guidelines

³³ C. Althaus, P. Bridgman, G. Davis, *The Australian Policy Handbook: A practical guide to the policy making process*, 6th edition, Sydney, Allen and Unwin, 2018

³⁴ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018

³⁵ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018

The program stage will also have implications for the evaluation design, see Table 2 from the BetterEvaluation website.³⁶

Table 2: Evaluation design at different program stages

Stage of program development	Consequence	Possible implication for the evaluation design
Not yet started	Can set up data collection from the beginning of implementation	Possible to gather baseline data as a point of comparison and also to establish comparison or control groups from the beginning
		Opportunity to build some data collection into administrative systems to reduce costs and increase coverage
	Period of data collection will be long	Need to develop robust data collection systems including quality control and storage
Part way through implementation	Cannot get baseline data unless this has already been set up	Will need to construct retrospective baseline data to estimate changes that have occurred
	Might be able to identify 'bright spots' where there seems to be more success and those with less success	Scope to do purposeful sampling and learn from particular successes and also cases that have failed to make much progress
Almost completed	Cannot get baseline data unless this has already been set up	Will need to construct retrospective baseline data to estimate changes that have occurred
	Depending on timeframes, some outcomes and impacts might already be evidenced	Opportunity to gather evidence of outcomes and impacts
Completed	Cannot get baseline data unless this has already been set up	Will need to construct retrospective baseline data to estimate changes that have occurred

³⁶ <https://www.betterevaluation.org/en/node/5294>

Stage of program development	Consequence	Possible implication for the evaluation design
	Depending on timeframes, some outcomes and impacts might already be evidenced	Opportunity to gather evidence of outcomes and impacts
	Cannot directly observe implementation	Will depend on existing data or retrospective recollections about implementation

Stages for planning and commissioning an evaluation

This toolkit is divided into six steps for planning and commissioning an evaluation. Table 3 shows how the numbered sections in the toolkit match these steps.

The evaluation planning that occurs in steps 1 and 2 starts during program design and may take a substantial amount of time. Sections 1 and 2 of the toolkit are structured to mirror the [Evaluation overview template](#) and the Evaluation work plan template to give step by step guidance through the templates.

Table 3: Stages for planning and commissioning an evaluation

Stage	Step	Title	Content
Program design	1	Complete the evaluation overview	How to concisely summarise what the program is aiming to achieve, how it will achieve this, external factors that may affect success, how the program's success will be measured and what evaluations will be required (and when) and identify other evaluation-related resource requirements.
Before the evaluation	2	Complete the evaluation work plan	How to: <ul style="list-style-type: none"> decide the appropriate evaluation methodology including program logic, key evaluation questions, data matrix and ethical considerations consider implementation including roles, responsibilities and resourcing identify risks.
	3	Engage the evaluation team	How to select the right evaluation team and apply the procurement governance policy.
During the evaluation	4	Manage the implementation of the evaluation	What the role of the evaluation manager is in overseeing the implementation of the evaluation work plan.
	5	Guide production of a quality evaluation report	How to structure an evaluation report, including succinct reporting of the evaluation findings. This step also outlines the minimum requirements of an evaluation report.
After the evaluation	6	Disseminate results and support use of evaluation	How to appropriately communicate evaluation results and respond to recommendations.

What's the difference between an evaluation overview and an evaluation work plan?

An evaluation overview briefly summarises a program's evaluation requirements as part of the Cabinet submission process (see [section 1. Complete the evaluation overview](#)). If the program is approved to proceed, an evaluation workplan details the evaluation requirements including a full program logic, evaluation questions and data matrix (see [section 2. Complete the evaluation work plan](#)).

1. Complete the evaluation overview

This section of the toolkit is structured to mirror the Evaluation overview template to give additional guidance, section by section.

An evaluation overview is completed during program design (see Stages for planning and commissioning an evaluation) and is part of the Cabinet submission template. It is mandatory for all new programs seeking additional funding of \$1 million or more in a year to complete an evaluation overview (refer to the [Cabinet handbook](#) and submission templates). It should be a concise summary of the key outcomes the program is trying to achieve, and how and when success will be measured. The overview should be completed in consultation with the agency's evaluation manager and/or the PEU at DTF.

When seeking Cabinet approval for new programs, agencies must:

- consider previous evaluations as part of the new policy/program design to ensure continuous learning and improvement
- identify the outcomes the program aims to influence
- identify the key performance indicators against which the success of the program will be measured
- identify data sources to monitor program effectiveness, including baseline data
- include sunset clauses
- include provision for evaluations as part of the initial funding request.

Evaluation overviews will be reviewed by CMC and DTF as part of normal Cabinet submission processes. If a program is approved to proceed, a full evaluation work plan (see [section 2. Complete the evaluation work plan](#)) is required and must be provided to DTF within six months of the approval.

1.1. What is the program aiming to achieve?

Explain why the program is needed and what it will aim to achieve. Consider how the program will impact future demand on government services if successful, and the time frame for this impact. This will need to be described in more detail in the evaluation work plan if the program is approved to proceed. Further guidance is at [section 2.2. Program overview](#).

1.2. How will the program achieve this?

Show how the program's inputs logically flow through activities to outputs and outcomes. Ideally, this is done using a program logic (see [section 2.5.1. Program logic](#)). If a program logic has not yet been developed for the program, the table below can be used as a starting point. A full program logic will be required as part of the evaluation work plan if the program is approved.

Table 4: How will the program achieve this?

Inputs	Activities	Outputs	Outcomes
The resources (funds, expertise, time) required for the delivery of activities.	The day-to-day tasks an organisation must undertake in order to provide a product or service.	The products or services generated by program activities; deliverables. The provision of outputs is typically under the control of the program and is related to the quantity and quality of program implementation.	The intended results of a program – the change it seeks to create. Outcomes can be affected by external factors.

1.3. What external factors may also influence the program's outcomes?

Explain how external factors could influence the program's outcomes, including economic, geographic, legislative, competitive and technological factors, and proposed risk management strategies.

List relevant programs with similar aims run by other agencies or governments, or non-government organisations to show how other programs may influence the outcomes of this program. See [section 2.5.1. Program logic](#) for further guidance.

1.4. How will the program's success be measured?

Show how the program's success will be measured for the activities, outputs and outcomes from Table 4. Indicators should be based on the SMART principles (specific, measurable, attainable, relevant and time-bound). Further guidance on selecting indicators can be found at the [BetterEvaluation website](#).

Table 5 should build on Table 4. A new row should be added to the table for each activity, output and outcome to identify an indicator, baseline, target and data source. Clearly identify and explain any gaps.

Table 5: How will the program's success be measured?

Program stage	Indicator	Baseline	Target	Data source

1.5. What evaluations will be required in the first five years of the program?

While there are many different types of evaluation, the [Program evaluation framework](#) is based on three main types of evaluation, linked to the program lifecycle:

- Process evaluation:** was the program implemented as expected? (Approximate timing ≤18 months.)
- Outcomes evaluation:** are the desired outcomes being achieved or on track to be achieved within the agreed timeframe? (Approximate timing >2 years.)

3. **Impact evaluation:** did the program influence the achievement of the desired outcomes, and was it value for money? (Approximate timing >3<5 years.)

Table 6 should list the evaluations required for the first five years of the program. All programs require a process evaluation within 18 months of commencement. The timing and need for an outcome and/or impact evaluation will depend the program's cost, strategic significance and degree of risk. As a general guide, only programs that are a high strategic priority would require all three evaluation types. Further guidance on evaluation types is at [section 2.5.3. Types of evaluation](#), and [section 2.6. Budget and resourcing](#) has guidance on estimating the budget for evaluations.

Table 6: What evaluations will be required in the first 5 years of the program?

Evaluation type	Yes/no	Timing	Estimated cost*
Process			
Outcome			
Impact			

1.6. Other evaluation-related resource requirements

Briefly address the following considerations:

- Is internal program data and/or analysis required?
- If data is required, is this data already being reported in the format required? If not, what resources are required to get the data?
- If analysis is required, what unit will undertake this work, have they been consulted in the preparation of the submission and do they have the skills and resources required? [Section 2.6. Budget and resourcing](#) has further guidance.

Further information on data collection is at [section 4. Manage the implementation of the evaluation](#).

2. Complete the evaluation work plan

This section of the toolkit is structured to mirror the [Evaluation work plan template](#) to give additional guidance, section by section.

2.1. Introduction

The evaluation work plan outlines future evaluation activity for the program, usually over a five-year period. It forms the basis of the request for tender (see [section 3.2. Prepare the request for tender](#)) to commission an external evaluation or to clarify requirements for an internal evaluation.

The work plan details roles and responsibilities, the evaluation methodology (including the program logic, the evaluation questions, evaluation types and data matrix) explains whether the evaluation will be external or internal, the budget and resourcing available, stakeholder communications, ethical considerations, evaluation risks and review. Where appropriate, key stakeholders should be included in these discussions. Specific methodologies for each of the evaluations in the evaluation work plan may need to be refined prior to the commencement of each evaluation.

The work plan also takes into account the strategic importance of the program and the expected level of resourcing for evaluation.

2.1.1. Who should develop the evaluation work plan?

The evaluation work plan is usually completed by the program manager with input from the agency's evaluation unit and/or DTF. For large and complex evaluations, it may be worth engaging an evaluation team at the development stage to help complete the evaluation work plan (see [section 2.5.3.4. External or internal evaluation](#) and [section 3.1. When to engage the evaluation team](#)).

While a program team may be able to develop designs for smaller scale evaluations, evaluation expertise may be required for more complex evaluations. Specialist expertise might be needed to:

- gather data about hard-to-measure outcomes or from hard-to-reach populations
- develop an evaluation design that adequately addresses causal attribution in outcome and/or impact evaluations
- advise on the feasibility of applying particular designs within the context of a program
- identify specific ethical and cultural issues.³⁷

If the program manager decides to use an external provider to develop the evaluation work plan, they should commission the design of the evaluation as a separate project as soon as the program is approved to proceed. This would be based on the evaluation overview in the Cabinet submission and would be used to complete the evaluation work plan. Subsequent requests for tender for each evaluation would draw on this work.

2.2. Program overview

The program overview needs to explain why the program is needed, what it is aiming to achieve and how it is expected to impact demand on future government services. It also briefly describes how the program operates (its funding and governance) and any sensitivities. It should be about half a page in length with any additional information in an appendix.

³⁷ <https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/4-manage-development-of-the-evaluation-design/>

2.2.1. Needs assessment

Ideally, a needs assessment will be undertaken as part of the program design. If a needs assessment has been carried out, please outline the findings in the program overview or a relevant appendix.

A needs assessment is a tool that is used for both designing programs and for conducting program evaluation.³⁸ It is a systematic method to determine who needs the program, how great the need is, characteristics of the target group, patterns of unmet needs, and what might work to meet the needs identified.³⁹

Key questions may include:

- What problems exist and how large or serious are they?
- What are the characteristics and needs of the target population?
- How are people affected as individuals?
- How is the community affected? What are the financial and social costs of the issue?
- Are other groups or agencies (including Commonwealth agencies) working to address the need?
- What are the opportunities for collaboration and shared funding?

Undertaking a systematic needs assessment is a transparent way of ensuring resources are used in the most effective way possible. The needs assessment should be reviewed during evaluation to assess whether the program is still needed.

Further information on needs assessment is available from the [Australian Institute of Family Studies](#), including [a guide](#) on how to do a needs assessment.

2.3. Evaluation overview

Include all the information from the evaluation overview section of the Cabinet submission to ensure the evaluation work plan has all the relevant information as part of a single document. You could attach the Cabinet submission evaluation overview as an appendix and refer to the attachment in this section.

2.4. Evaluation roles and responsibilities

There are many decisions to be made in an evaluation including:

- the focus of the evaluation (including the key evaluation questions)
- choosing the evaluator/evaluation team
- approving the evaluation design
- approving the evaluation report(s) and who can access them.

[BetterEvaluation's Manager's Guide to evaluation](#) encourages consideration of who will be involved in making these decisions, what their role will be and how the decisions will be made.⁴⁰

³⁸ NSW Evaluation Toolkit (step 1).

³⁹ NSW Evaluation Toolkit (step 1).

⁴⁰ BetterEvaluation Commissioner's Guide (step 1).

Who?

Contributors to involve in the decision-making process may include:

- the program manager within the agency
- an evaluation steering committee
- a technical advisory group or a number of individual technical advisors (including service providers)
- a community consultation committee or relevant people from the community.

What?

The role of each individual or group in relation to specific decisions can be categorised as follows:

- **to consult:** those whose opinions are sought (bilateral)
- **to recommend:** those who are responsible for putting forward a suitable answer to the decision.
- **to approve:** those who are authorised to approve a recommendation.
- **to inform:** those who are informed after the decision has been made (unilateral).

How?

One or more of the following processes may be employed in the decision-making process:

- Decisions made based on support from the majority. Where decisions may be contentious it is important to be clear about who is eligible to vote and whether proxy votes are allowed.
- Decisions made based on reaching a consensus. In practical terms, that can mean giving all decision makers the right to veto.
- Decisions made based on hierarchy (formal positions of authority).

Evaluation managers are often, but not always, the program manager. For large evaluations, the evaluation manager may be assisted by one or more staff members with specific responsibilities in the management process.

Table 7: Potential evaluation roles and responsibilities

Area/Committee	Responsibility
Program manager	<ul style="list-style-type: none"> • Educate the external evaluator(s) about: the program's objectives, operations and intended beneficiaries; the expectations about the evaluation and any relevant organisational background. • Provide input and/or collate feedback on the evaluation plan. • Specify reporting requirements in terms of progress in the implementation of the evaluation (including reporting of important challenges and their resolution or which potential issues need to be raised for decision making elsewhere). • Specify what is expected to be included in the formal evaluation report(s). • Keep the evaluator(s) apprised of any changes in the program's operations or evaluation context. • Provide regular updates on the evaluation process to all staff. • Monitor the implementation of the evaluation including completion of milestones/deliverables. • Facilitate program staff involvement in the evaluation, where relevant and agreed. • Serve as troubleshooter, resolving problems or locating help to resolve them.
[Program name] Evaluation Steering Committee	<ul style="list-style-type: none"> • Endorse the terms of reference and evaluation work plan. • Provide feedback on draft findings and recommendations and the draft evaluation report. • Chair of the Steering Committee to sign-off on the final evaluation report. • Draft the evaluation terms of reference and evaluation plan for the evaluation. • Conduct, manage, or advise on evaluation activity as required.
Evaluator	<ul style="list-style-type: none"> • Develop an evaluation plan, in conjunction with the program manager. • Provide monthly or quarterly progress reports on the implementation of the evaluation (written or in person). • Attend evaluation meetings. • Train data collectors on participant/case selection for sampling purposes, using data collection instruments, data quality assurance. • Ensure adherence to ethical standards adherence (for example, confidentiality of data) during all phases of the evaluation. • Oversee implementation of data collection such as: interviewing program staff and participants, conducting focus groups, observing service delivery activities, reviewing participant case records, developing data management procedures and tools (such as a database), coding and cleaning data, analysing data. • Write interim (quarterly, biannual, yearly) evaluation reports and the final evaluation report. • Present findings.

Area/Committee	Responsibility
DTF	<ul style="list-style-type: none"> Review the final evaluation plan

2.5. Evaluation methodology

The evaluation methodology should include a detailed program logic, key evaluation questions and a data matrix. It needs to consider what types of evaluations should be used and how key evaluation questions will be addressed.

2.5.1. Program logic

A program logic should illustrate how the program will work by linking program activities with intended outcomes. It visually represents the theory of change underpinning the program and describes how the program contributes to a chain of results flowing from the inputs and activities to short-term, intermediate and long-term outcomes.

Different terms are used for a program logic including program theory, logic model, theory of change, causal model, outcomes hierarchy, results chain, and intervention logic.⁴¹ Usually it is represented as a one page diagram. The diagrams and terms used with program logic may also vary – sometimes the diagrams are shown as a series of boxes, as a table, or as a series of results with activities occurring alongside them rather than just at the start. Some diagrams show the causal links from left to right, some from bottom to top. In all cases, a program logic needs to be more than just a list of activities with arrows to the intended outcomes. For some examples, see [section 2.5.1.4. Program logic library](#).

2.5.1.1. What is a program logic used for?

A program logic should show what needs to be measured in order to distinguish between implementation failure (not done right) and theory failure (done right but still did not work).⁴² A program logic:

- clarifies and communicates program intentions and outcomes
- demonstrates alignment between activities and objectives
- explains causal assumptions and tests if they are supported by evidence
- identifies relevant external factors that could influence outcomes (either positively or negatively)
- identifies key indicators to be monitored
- identifies gaps in available data and outlines mitigation measures
- clarifies the outcomes measurement horizon and identifies early indicators of progress or lack of progress in achieving results
- focuses evaluation questions.

A program logic underpins data collection by identifying a program's operating steps and defining what program managers should monitor and measure. A program logic also helps identify the components of the

⁴¹ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

⁴² S. C. Funnell, , P. J. Rogers, (20110209). Purposeful Program Theory: Effective Use of Theories of Change and Logic Models [VitalSource Bookshelf version].

program to be tracked as part of monitoring (outputs) versus those that should be assessed as part of an outcome or impact evaluation (outcomes).⁴³

Monitoring activities and outputs shows which program components are being well implemented and which could be improved.⁴⁴ A focus on measuring outcomes and impacts without a good monitoring system can result in wasted resources. For example, if a program aims to improve literacy in schools using particular books, it is important to monitor delivery and use of the books so that the program can be adjusted early if the books are not being delivered or used. If the program does not have a good monitoring system in place and waits three years before doing an outcome evaluation, this could be an expensive way of finding out that the books had not even been used. Further information on the importance of a good monitoring system as a way of keeping evaluation costs down is in the [Goldilocks toolkit](#).

A program logic illuminates the critical assumptions and predictions that must hold for key outcomes to occur, and suggests important areas for data collection.⁴⁵ It also helps prioritise data collection; for example, if there is no way to isolate external factors that influence the outcomes of the program, is it worth collecting the outcome data? Important considerations include the cost of collecting the outcome data and the conclusions that can reasonably be drawn from it. In some cases, a process evaluation may be sufficient but it is essential that the results of the process evaluation are not overstated.

2.5.1.2. Developing a program logic

Developing a program logic is a part analytical, part consultative process. Analytically, it should review the program settings to identify statements of activities, objectives, aims and intended outcomes. It should then refine and assemble these statements into a causal chain that shows how the activities are assumed to contribute to immediate outcomes, intermediate outcomes and ultimately to the longer term outcome. Consultatively, the process should involve working with a range of stakeholders to draw on their understanding of the outcomes and logic, and also encourage greater ownership of the program logic.⁴⁶

It is useful to think realistically about when a successful program will be able to achieve particular outputs and outcomes. For example, within a domestic violence context, a successful program may see an increase in reporting (due to increased awareness and/or availability of support) before reporting decreases. Where possible, estimated timing of indicators should be built into the program logic to help clarify what success looks like in different timeframes.⁴⁷

The [Evaluation work plan template](#) includes a suggested template for the program logic. This is an optional starting point rather than a mandatory structure, however all program logics should clearly identify assumptions and relevant external factors. 'Backcasting' starts with identifying long-term outcomes of a program and envisaging alternative futures, and then working backwards to determine the necessary steps towards achieving these outcomes. Unlike forecasting, which considers what is currently occurring and predicting future outcomes, the benefit of backcasting is that it allows stakeholders to brainstorm and consider alternative courses of action.⁴⁸ BetterEvaluation's direction on [developing programme theory / theory of change](#) may assist in determining the best approach.

⁴³ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

⁴⁴ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

⁴⁵ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

⁴⁶ <https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/1-develop-program-logic-and-review-needs/>.

⁴⁷ For example see Figure 11, page 66, <https://www.dfat.gov.au/sites/default/files/evawg-final-report-nov-19.pdf>.

⁴⁸ <https://www.betterevaluation.org/evaluation-options/backcasting>.

Table 8: Potential steps for developing a program logic⁴⁹

Task	Approach
Undertake situational analysis	The context of the problem, its causes and consequences. A good situation analysis will go beyond problems and deficits to identify strengths and potential opportunities.
Identify outcomes	<ol style="list-style-type: none"> 1. Prepare a list of possible outcomes. 2. Cluster the outcomes that are related. 3. Arrange the outcomes in a chain of 'if-then' statements to link the short term and long term outcomes. 4. Identify where a higher level outcome affects a lower level one (feedback loops). 5. Validate the outcomes chain with key stakeholders.
Identify outputs	The direct deliverables of a program. The products, goods or services that need to be provided to program participants to achieve the short-term outcomes.
Identify activities	The required actions to produce program outputs.
Identify inputs	The resources required to run the program.
Identify assumptions	<p>In every link between activity, output and outcome, many different assumptions are made that must hold for the program to work as expected. Making these assumptions explicit and identifying the most critical among them helps to figure out what testing and monitoring is needed to ensure the program works as planned. This includes assumptions about:</p> <ul style="list-style-type: none"> • the need for the program • how the program will work • whether program activities are likely to produce the intended results.
Consider external factors that also cause changes	<p>What besides the program could influence the intended outcome? Listing the most important external influences helps organisations better understand the counterfactual and clarify whether it will be possible to attribute a change in the outcome solely to the program.</p>
Identify risks and unintended consequences	The world around the program is unlikely to remain static; changes in external conditions pose unavoidable risks to any program. It is important to identify the most likely and potentially damaging risks and

⁴⁹ Adapted from S. C. Funnell, P. J. Rogers, (20110209). Purposeful Program Theory: Effective Use of Theories of Change and Logic Models [VitalSource Bookshelf version] and M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

Task	Approach
	develop a risk reduction or mitigation plan (see section 2.9. Evaluation risks).

Once a program logic is developed, it is useful to map existing data from an established program or previous evaluation onto the program logic to identify priority areas for additional data collection.

2.5.1.3. What does a good program logic look like?

There is no one way to represent a program logic – the test is whether it is a representation of the program's causal links, and whether it communicates effectively with the intended audience by making sense and helping them understand the program.⁵⁰ See examples in [Program Logic Library](#).

Table 9 provides guidance on the aspects required for a good program logic, and explains the different criteria for 'requires improvement', 'satisfactory' and 'good'. This may be refined over time in response to user feedback to ensure it is appropriate to a Territory Government context.

See [section 2.10. Reviewing the evaluation work plan](#) for suggestions on what to look for when reviewing an established program logic.

⁵⁰ <https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/1-develop-program-logic-and-review-needs/>.

Table 9: Program logic rubric⁵¹

Section of program logic	Requires improvement	Satisfactory	Good (includes all satisfactory criteria plus those listed below)
Overall	<ul style="list-style-type: none"> • The logic linking activities/outputs to outcomes is not convincing. • Arrows not well matched to timescale. • Theory of change ill-defined or not evidence-based. • Not comprehensive across the columns. • Some components incorrectly placed in columns. • Doesn't fit on one page. 	<ul style="list-style-type: none"> • Adequately represents the views of the main stakeholders: policy, program and Evaluation Unit. • The theory of change is clear and indicated by arrows. • The outcomes are realistic relative to the inputs and activities (not changing the world). • Uses active, not passive voice. • The focus is evaluative rather than promotional. • All components are in correct columns. • Outputs and/or outcomes are linked to activities. • The logic linking activities/outputs to outcomes is plausible. • Fits on one page. • Has been cleared/approved at GM level or other where appropriate. • Has been presented to PAC for noting. 	<ul style="list-style-type: none"> • The template has been adapted to a sensible extent to capture differences between programs. • A key is provided where useful/applicable. • Acronyms are explained. • Isn't cluttered, with a suitable level of detail. • The logic linking activities/outputs to outcomes is based on evidence.

⁵¹ DIIS Evaluation Strategy 2017-2021

Section of program logic	Requires improvement	Satisfactory	Good (includes all satisfactory criteria plus those listed below)
Inputs and participation	<ul style="list-style-type: none"> • Is either not comprehensive or is inaccurate in relation to inputs, stakeholders. • Omits staffing and or administered funding. • Lists government under participation (unless the program targets government as the beneficiary). 	<ul style="list-style-type: none"> • Guidelines for prioritising and scaling evaluation activity are used. 	<ul style="list-style-type: none"> • Priority programs are evaluated. • Evaluations use fit-for-purpose methodologies.
Activities and or outputs	<ul style="list-style-type: none"> • Too much detail on generic administration processes such as for granting programs. • Outputs are confused with or substitute for outcomes. • Activities don't link to outputs and outcomes. 	<ul style="list-style-type: none"> • Identifies who does what to whom. • Separates Commonwealth and participant activities as necessary. • Shows ordering of key activities and links to outcomes. • Activities/outputs are directly related to objectives and can be monitored and assessed. • Avoids too much detail on generic administration processes such as for granting programs. 	<ul style="list-style-type: none"> • Uses action verbs to identify activities. • Outcomes are informed by evidence and experience/lessons learnt.

Section of program logic	Requires improvement	Satisfactory	Good (includes all satisfactory criteria plus those listed below)
Outcomes	<ul style="list-style-type: none"> • Outcomes are not comprehensively identified. • Outputs are confused with outcomes. • No theory of change (no connecting links between boxes or every box connects to every other box). • Outcomes are aspirational and or not able to be assessed. • Simply restates policy objectives. • Doesn't consider short/medium/long-term outcomes. • Links between shorter and longer-term outcomes aren't convincing. • Outcomes are out of proportion to inputs. 	<ul style="list-style-type: none"> • Identification of outcomes is suitably comprehensive. • Articulates who the outcomes relate to (who is benefiting/being affected). • Uses evaluative, not promotional language. • Language is proportional increase and not just number. • Provides realistic timeframes for outcomes. • Uses SMART indicators.¹ Outcomes that can't be measured are clearly indicated. • Outcomes align with objectives. Outcomes are well connected with a logical flow from short-term to long-term. • Demonstrates logic links and clearly articulates anticipated changes. • Doesn't restate activities/outputs. • Links between shorter and longer-term outcomes are plausible. 	<ul style="list-style-type: none"> • Uses feedback loops if appropriate. • Marks external factors and assumptions in links. • Outcomes link backwards to outputs and activities. • Links such as between shorter and longer-term outcomes are based on evidence.

Section of program logic	Requires improvement	Satisfactory	Good (includes all satisfactory criteria plus those listed below)
External factors and assumptions	<ul style="list-style-type: none"> • Not included or not clearly identified. • Not supported by evidence. 	<ul style="list-style-type: none"> • Key external factors and assumptions identified. 	<ul style="list-style-type: none"> • Assumptions supported by evidence/theory of change and risks. • Informed by lessons learnt. • Assumptions comprehensively state the conditions required for the program to function effectively.

1. SMART Specific, Measurable, Attainable, Relevant and Time-bound.

Source: Department of Industry, Innovation and Science (2017)

2.5.1.4. Program logic library

Table 10: Program logic library

Program logic source	Comments
Australian Institute of Family Studies Blairtown example program logic	A hypothetical program aiming to ensure children reach appropriate developmental milestones. Includes assumptions and external factors.
Evidence-Based Programs and Practice in Children and Parenting Support Programs	A project supporting nine Children and Parenting Support services in regional and rural NSW to enhance their use of evidence-based programs and practice. Includes assumptions and external factors.
National Forum on Youth Violence Prevention	A program that aims to maximise the use of city partnerships and increase the effectiveness of federal agencies to reduce youth violence. Includes assumptions and external factors.
Australian Policy Service Policy Hub Evaluation Ready example program logic: Save Our Town	A hypothetical program aimed at stimulating private sector investment, population growth, and economic expansion and diversification to increase a region's viability. Includes assumptions and external factors.
University of Michigan Evaluation Resource Assistant example program logic	A hypothetical program aimed at reducing rates of child abuse and neglect. Does not include assumptions and external factors, however it is a good example of how to use a program logic to prioritise key evaluation questions and indicators.
The Goldilocks Toolkit Case Studies	Program logics and lessons learned from a range of international social programs including Acumen, GiveDirectly, Digital Green, Root Capital, Splash, TulaSalud, Women for Women International and One Acre Fund. The program logics do not include assumptions and external factors but the case studies provide examples of how to reduce evaluation costs through a good monitoring system.
Geo-Mapping for Energy & Minerals	Appendix A of this evaluation report by Natural Resources Canada has a program logic for a program improving regional geological mapping for responsible resource exploration and development. Does not include assumptions and external factors but does have an evaluation matrix to show how the evaluation questions will be addressed.
The Logic Model Guidebook: Better Strategies for Great Results	Program logics from a Community Leadership Academy (see page 10, and a marked up version on page 56) and a Health Improvement program (see page 39 and a marked up version on page 57).

2.5.2. Evaluation questions

Across the program cycle, evaluations need to include a range of questions that promote accountability for public funding and learning from program experiences. These questions need to align with the program logic and will form the basis of the terms of reference. Evaluation questions may be added to or amended closer to evaluation commencement to account for changes in policy context, key stakeholders, or performance indicators.

Different types of questions need different methods and designs to answer them. In evaluations there are four main types of questions: descriptive, action, causal and evaluative.

2.5.2.1. Descriptive questions

Descriptive questions ask about what has happened or how things are. For example:

- What were the resources used by the program directly and indirectly?
- What activities occurred?
- What changes were observed in conditions or in the participants?

Descriptive questions might relate to:

- Inputs – materials, staff.
- Processes – implementation, research projects.
- Outputs – for example, research publications.
- Outcomes – for example, changes in policy on the basis of research.
- Impacts – for example, improvements in agricultural production.

2.5.2.2. Action questions

Action questions ask about what should be done to respond to evaluation findings. For example:

- What changes should be made to address problems that have been identified?
- What should be retained or added to reinforce existing strengths?
- Should the program continue to be funded?

2.5.2.3. Causal questions

Causal questions ask about what has contributed to changes that have been observed. For example:

- What produced the outcomes and impacts?
- What was the contribution of the program to producing the changes that were observed?
- What other factors or programs contributed to the observed changes?

What is the difference between correlation and causation?

Two variables are classified as correlated if both increase and decrease together (positively correlated) or if one increases and the other decreases (negatively correlated). Correlation analysis measures how close the relationship is between the two variables.

Causal questions need to investigate whether programs are causing the outcomes that are observed. Although there may be a strong correlation between two variables, for example the introduction of a new program and a particular outcome, this correlation does not necessarily mean the program is directly causing the outcome. See Box 1 for an example.

Box 1: Evaluating to improve resource allocations for family planning and fertility in Indonesia

Indonesia's innovative family planning efforts gained international recognition in the 1970s for their success in decreasing the country's fertility rates. The acclaim arose from two parallel phenomena: (1) fertility rates declined by 22% between 1970–1980, by 25% between 1981–1990, and a bit more moderately between 1991–1994; and (2) during the same period, the Indonesian government substantially increased resources allocated to family planning (particularly contraceptive subsidies).

Given that the two things happened concurrently, many concluded that increased investment in family planning had led to lower fertility rates. Unconvinced by the available evidence, a team of researchers evaluated the impact of family planning programs on fertility rates and found, contrary to what was generally believed, that family planning programs only had a moderate impact on fertility, with changes in women's status deemed to have a larger impact on fertility rates.

The researchers noted that before the start of the family planning program very few women of reproductive age had finished primary education. During the same period as the family planning program, however, the Indonesian government undertook a large-scale education program for girls. By the end of the program, women entering reproductive age had benefited from the additional education. When the oil boom brought economic expansion and increased demand for labour in Indonesia, the participation of educated women in the labour force increased significantly. As the value of women's time at work rose, so did the use of contraceptives. In the end, higher wages and empowerment explained 70% of the observed decline in fertility—more than the investment in family planning programs.

These evaluation results informed policy makers' subsequent resource allocation decisions: funding was reprogrammed away from contraception subsidies and towards programs that increased the enrolment of women in school. Although the ultimate goals of the two programs were similar, evaluation studies had shown that in the Indonesian context, lower fertility rates could be obtained more effectively by investing in education than by investing in family planning.⁵²

There are many designs and methods to answer causal questions but they usually involve one or more of these strategies:

(a) Compare results to an estimate of what would have happened if the program had not occurred (this is known as a counterfactual)

This might involve creating a control group (where people or sites are randomly assigned to either participate or not) or a comparison group (where those who participate are compared to others who are matched in various ways). Techniques include:

- **Randomised controlled trial (RCT):** a control group is compared to one or more treatment groups.
- **Matched comparison:** participants are each matched with a non participant on variables that are thought to be relevant. It can be difficult to adequately match on all relevant criteria.
- **Propensity score matching:** create a comparison group based on an analysis of the factors that influenced people's propensity to participate in the program.
- **Regression discontinuity:** compares the outcomes of individuals just below the cut-off point with those just above the cut-off point.⁵³

⁵² 2011, 'Impact Evaluation in Practice', World Bank.

⁵³ For a good example of a low-cost impact evaluation using regression discontinuity, see the [Root Capital Case Study in the Goldilocks Toolkit](#).

(b) Check for consistency of the evidence with the theory of how the intervention would contribute to the observed results

This can involve checking that intermediate outcomes have been achieved, using process tracing to check each causal link in the theory of change, identifying and following up anomalies that don't fit the pattern, and asking participants to describe how the changes came about. Techniques include:

- **Contribution analysis:** sets out the theory of change that is understood to produce the observed outcomes and impacts and then searches iteratively for evidence that will either support or challenge it.
- **Key informant attribution:** asks participants and other informed people about what they believe caused the impacts and gathers information about the details of the causal processes.
- **Qualitative comparative analysis:** compares different cases to identify the different combinations of factors that produce certain outcomes.
- **Process tracing:** a case-based approach to causal inference which focuses on the use of clues within a case (causal-process observations) to adjudicate between alternative possible explanations. It involves checking each step in the causal chain to see if the evidence supports, fails to support or rules out the theory that the program or project produced the observed impacts.
- **Qualitative impact assessment protocol:** combines information from relevant stakeholders, process tracing and contribution analysis, using interviews undertaken in a way to reduce biased narratives.

(c) Identify and rule out alternative explanations

This can involve a process to identify possible alternative explanations (perhaps involving interviews with program sceptics and critics, and drawing on previous research and evaluation, as well as interviews with participants) and then searching for evidence that can rule them out.

While technical expertise is needed to choose the appropriate option for answering causal questions, the program manager should be able to check there is an explicit approach being used, and seek technical review of its appropriateness for third parties where necessary.

Further guidance and options for measuring causal attribution can be found in the [UNICEF Impact evaluation series](#).

2.5.2.4. Evaluative questions

Evaluative questions ask whether an intervention can be considered a success, an improvement or the best option, and require a combination of explicit values as well as evidence. For example:

- In what ways and for whom was the program successful?
- Did the program provide value for money, taking into account all the costs incurred (not only the direct funding) and any negative outcomes?

Many evaluations do not make explicit how evaluative questions will be answered – what the criteria will be (the domains of performance), what the standard will be (the level of performance that will be considered adequate or good), or how different criteria will be weighted. A review of the design could check each of these in turn:

- Are there clear criteria for this evaluative question?
- Are there clear standards for judging the quality of performance on each criterion?

- Is there clarity about how to synthesize evidence across criteria? Is there a performance framework that explains what ‘how good’ or ‘how well’ or ‘how much’ mean in practice?⁵⁴ For example, is it better to have some improvement for everyone or big improvements for a few?
- Are the criteria, standards and approach to synthesis appropriate? What has been their source? Is further review of these needed? Who should be involved?

Ideally, an evaluation design will be explicit about these, including the source of these criteria and standards. The BetterEvaluation website has further information on Evaluation methods for assessing value for money⁵⁵ and Oxford Policy Management’s approach to assessing value for money⁵⁶ is useful for assessing value for money in complex interventions.

2.5.2.5. Key evaluation questions

To clarify the purpose and objectives of an evaluation, there should be a limited number of higher order key evaluation questions (roughly five to seven questions) addressing:

- **Appropriateness:** to what extent does the program address an identified need?
- **Effectiveness:** to what extent is the program achieving the intended outcomes, in the short, medium and long term?
- **Efficiency:** do the outcomes of the program represent value for money?

These key evaluation questions are high-level research topics that can be broken down into detailed sub questions, each addressing a particular aspect. The key evaluation questions are not yes or no questions.

Key evaluation questions often contain more than one type of evaluation question. For example, to answer “How effective has the program been?” requires answering:

- **Descriptive questions:** What changes have occurred?
- **Causal questions:** What contribution did the intervention make to these changes?
- **Evaluative questions:** How valuable were the changes in terms of the stated goals taking into account types of changes, level of change and distribution of changes?

A way to test the validity and scope of evaluation questions is to ask: when the evaluation has answered these questions, have we met the full purpose of the evaluation?⁵⁷

See also [section 4.2.1. Evidence synthesis](#).

2.5.3. Types of evaluation

While there are a number of different approaches to evaluation,⁵⁸ the [program evaluation framework](#) is based on three types,⁵⁹ linked to the program lifecycle:

1. **Process evaluation:** considers program design and initial implementation (≤18 months).

⁵⁴ Further guidance on developing performance frameworks including examples is on pg 6-11 of [Evaluation Building Blocks – A Guide by Kinnect Group](#)

⁵⁵ <https://www.betterevaluation.org/sites/default/files/Evaluating%20methods%20for%20assessing%20vfm%20-%20Farida%20Fleming.pdf>

⁵⁶ <https://www.betterevaluation.org/en/resources/opms-approach-assessing-value-money>

⁵⁷ <https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/2-develop-the-evaluation-brief/>.

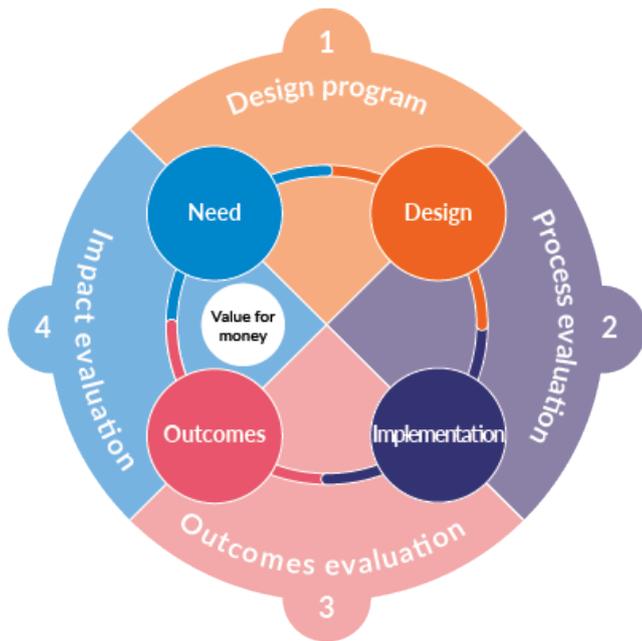
⁵⁸ For information about other evaluation types, please see the BetterEvaluation website.

⁵⁹ Further information on these three evaluation types are in sections 3.2.1 to 3.2.3.

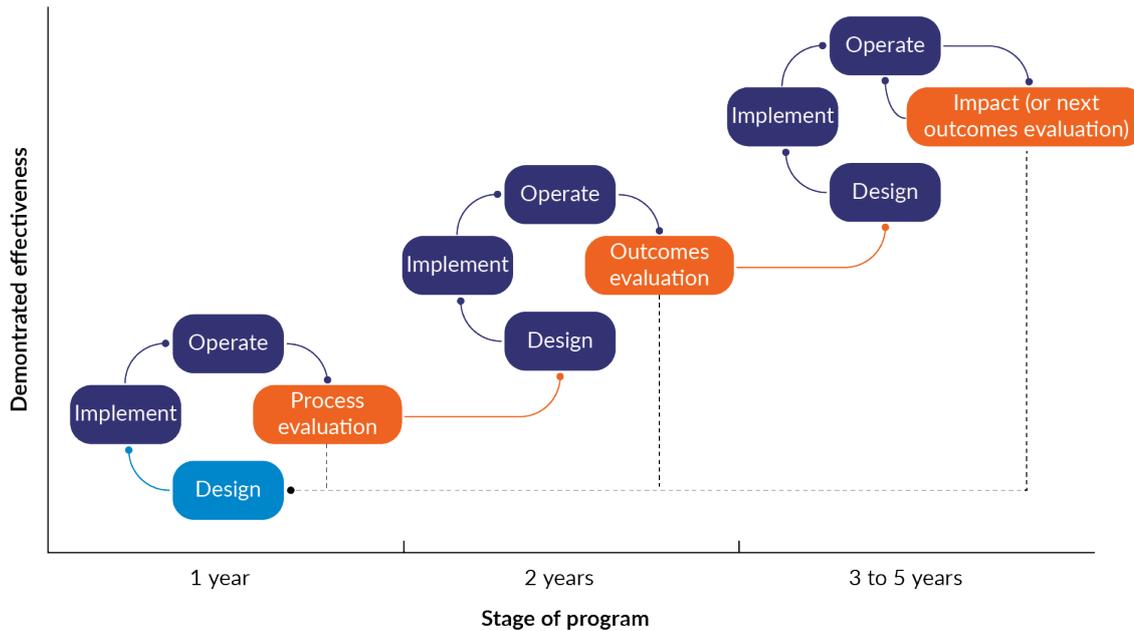
2. **Outcome evaluation:** considers program implementation (>2 years) and short to medium term outcomes.
3. **Impact evaluation:** considers medium to long term outcomes (>3 years), and whether the program contributed to the outcomes and represented value for money.

These three evaluation types address different questions at various stages of the program lifecycle, with each evaluation building on the evidence from the previous evaluation (Figure 3). Not all programs will require all three evaluation types. The evaluation overview, completed as part of the Cabinet submission process, will specify which evaluation types are necessary for each program. The different types of evaluation are used to build a clearer picture of program effectiveness as the program matures (Figure 4).

Figure 3: Different types of evaluations consider different aspects of the program⁶⁰



⁶⁰ Adapted from the Department of Industry, Innovation and Science Evaluation Strategy, 2017–2021.

Figure 4: Evaluations over the program lifecycle of a major program ⁶¹

2.5.3.1. Process evaluations

A process evaluation investigates whether the program is being implemented according to plan.⁶² This type of evaluation can help to differentiate ineffective programs from implementation failure (where the program has not been adequately implemented) and theory failure (where the program was adequately implemented but did not produce the intended impacts).⁶³ As an ongoing evaluative strategy, it can be used to continually improve programs by informing adjustments to delivery.⁶⁴

Process evaluations may be undertaken by the relevant program team, if they have appropriate capability.

A process evaluation will typically try to answer questions such as:

- Was the program implemented in accordance with the initial program design?
- Was the program rollout completed on time and within the approved budget?
- Are there any adjustments to the implementation approach that need to be made?
- Are more or different key performance indicators required?
- Is the right data being collected in an efficient way?

⁶¹ Adapted from NSW Government Evaluation Framework 2013.

⁶² BetterEvaluation Manager's Guide to Evaluation, accessed May 2020, <https://www.betterevaluation.org/en/node/5282>.

⁶³ Rogers, P. et al. (2015), Choosing appropriate designs and methods for impact evaluation, Office of the Chief Economist, Australian Government, Department of Industry, Science, Energy and Resources.

⁶⁴ NSW Government Evaluation Toolkit, accessed May 2020, <https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/2-develop-the-evaluation-brief/>.

2.5.3.2. Outcome evaluations

An outcome evaluation assesses progress in early to medium-term results that the program is aiming to achieve.⁶⁵ It is suited to programs at a business as usual stage in the program lifecycle and is usually externally commissioned.

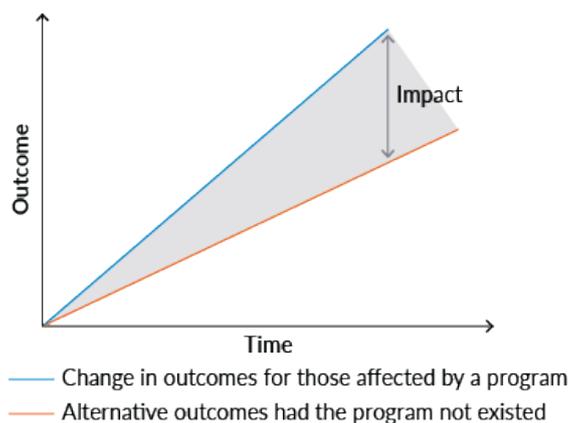
An outcome evaluation will typically try to answer questions such as:

- What early outcomes or indications of future outcomes are suggested by the data?
- Did the program have any unintended consequences, positive or negative? If so, what were those consequences? How and why did they occur?
- How ready is the program for an impact evaluation?

There is an important distinction between measuring outcomes, which is a description of the factual, and using a counterfactual to attribute observed outcomes to the intervention.⁶⁶ A good outcome evaluation should consider whether the program has contributed to the outcome, noting this becomes easier over time and is therefore more of a focus in impact evaluations (see Figure 5).

Figure 5 illustrates how the impact of a program is measured by the change in outcomes for those affected by a program (blue line) compared to the alternative outcomes had the program not existed (pink line). As impact generally increases over time, it tends to be measured later in the program lifecycle.

Figure 5: Program impact over time



2.5.3.3. Impact evaluations

An impact evaluation builds on an outcome evaluation to assess longer-term results.⁶⁷ It must test whether the program has made a difference by comparing what would have happened in the absence of the program⁶⁸ (further guidance in [section 2.5.2.3. Causal questions](#)). In situations where it is not possible or appropriate to undertake a rigorous impact evaluation, it may be better to monitor, learn and improve,⁶⁹ though process and/or outcome evaluations.

⁶⁵ BetterEvaluation Manager's Guide to Evaluation, accessed May 2020,

<https://www.betterevaluation.org/en/node/5282>.

⁶⁶ <http://www.oecd.org/dac/evaluation/dcdndep/37671602.pdf>.

⁶⁷ https://www.betterevaluation.org/en/themes_overview.

⁶⁸ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

⁶⁹ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

As impact is the change in outcomes compared to the alternative outcomes had the program not existed,⁷⁰ it is usually easier to measure impact later in the lifecycle of the program (see Figure 5). These evaluations commonly occur at least three years after program implementation. However, the appropriate timing for measuring impact will depend on the program and needs to be decided on case-by-case basis.⁷¹

An impact evaluation will typically try to answer questions such as:

- Were the intended outcomes achieved as set out in the program's aims and objectives?
- Have other investments influenced the attainment of the program's aims and objectives? If so, in what way?
- Did the program contribute to achieving the outcomes as anticipated? If so, to what extent?
- Were there any unintended consequences?
- What would have been the situation had the program not been implemented?
- To what extent did the benefits of the program outweigh the costs?
- Did the program represent good value for money?
- Was the program delivered cost-effectively?

Impact evaluations are usually externally commissioned due to their complexity, and are generally reserved for high-risk and complex programs due to their cost. The design options for an impact evaluation need significant investment in preparation and early data collection. It is important that impact evaluation is addressed as part of the integrated monitoring and evaluation approach outlined in the evaluation work plan. This will ensure that data from other monitoring and the process and outcome evaluations can be used as needed.⁷² Equity concerns may require an impact evaluation to go beyond simple average impacts to identify for whom and in what ways the program has impacted outcomes (further guidance in [section 2.8. Ethical considerations](#)).⁷³

Impact evaluations usually include a value-for-money assessment to determine whether the benefits of the program outweighed the costs and whether the outcomes could have been achieved more efficiently through program efficiencies or a different approach.⁷⁴ Value for money in this context is broader than a cost benefit analysis, it is a question of how well resources have been used and whether the use is justified (further guidance in [section 2.5.2.4. Evaluative questions](#)).

Further information on impact evaluations is available from [BetterEvaluation](#) and the [UNICEF Impact evaluation series](#).

2.5.3.4. External or internal evaluation

Evaluations can be commissioned externally to an appropriate consultant or academic evaluator, conducted internally by agency staff, or conducted using a hybrid model of an internal evaluator supported by an external evaluator. Each method is detailed below:

- **External evaluator(s):** one evaluator serves as team leader and is supported by program staff
- **Internal evaluator(s):** one evaluator serves as team leader and is supported by program staff

⁷⁰ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

⁷¹ UNICEF Impact Evaluation Series, webinar 5, RCT's, <https://www.unicef-irc.org/KM/IE/impact.php>.

⁷² Adapted from https://www.betterevaluation.org/en/themes/impact_evaluation.

⁷³ https://www.betterevaluation.org/en/themes/impact_evaluation.

⁷⁴ Value for money can also be considered in earlier evaluations, if there is sufficient evidence. As the evidence for a program accumulates, so will the expectation for an assessment of value for money.

- **Hybrid model:** An internal evaluator serves as team leader and is supported by other internal evaluators and program staff, as well as external evaluator(s).

If an external evaluator is hired to conduct the evaluation, the program manager and other agency staff still need to be involved in the evaluation process. Program staff are not only primary users of the evaluation findings but are also involved in other evaluation-related tasks (such as providing access to records or educating the evaluator about the program). Be realistic about the amount of time needed for this involvement so staff schedules do not get overburdened.

The decision to conduct an evaluation internally or commission an external evaluation is usually a decision for the agency's accountable officer. However, as a general best-practice guide, outcome or impact evaluations of high tier programs should be externally evaluated. It is advisable to engage an external evaluator/evaluation team when:

- the scope and/or complexity of the evaluation requires expertise that is not internally available
- a program or project is politically sensitive and impartiality is a key concern
- internal staff resources are scarce and timeframes are particularly pressing (that is, there is little flexibility in terms of evaluation timing).

Table 11: Internal vs external evaluators outlines the tradeoffs between internal and external evaluators.

Table 10: Internal vs external evaluators⁷⁵

Component	Internal evaluator(s)	External evaluator(s)
Perspective	May be more familiar with the community, issues and constraints, data sources and resources associated with the project/program (they have an insider's perspective).	May bring a fresh perspective, insight, broader experience, and recent state-of-the-art knowledge (they have an outsider's perspective).
Knowledge and skills	Are familiar with the substance and context of research for development programming.	May possess knowledge and skills that internal evaluators are lacking. However it may be difficult to find evaluators who understand the specifics of research for development programming.
Buy-in	May be more familiar with the project/program staff and may be perceived as less threatening. In some contexts, may be seen as too close and participants may be unwilling to provide honest feedback.	May be perceived as intrusive or a threat to the project/program (perceived as an adversary) Alternatively, it may be considered impartial and participants may be more comfortable providing honest feedback.
Stake in the evaluation	May be perceived as having an agenda/stake in the evaluation.	Can serve more easily as an arbitrator or facilitator between stakeholders as perceived as neutral.

⁷⁵ BetterEvaluation Commissioner's Guide <https://www.betterevaluation.org/en/node/5284>

Component	Internal evaluator(s)	External evaluator(s)
Credibility	May be perceived as biased as 'too close' to the subject matter, which may reduce the credibility of the evaluation hindering its use.	May provide a view of the project/program that is considered more objective and give the findings more credibility and potential for use.
Resources	May use considerable staff time, which is always in limited supply, especially when their time is not solely dedicated to the evaluation.	May be more costly and still involve substantial management/staff time from the commissioning organisation.
Follow-up/Use of evaluation findings	More opportunity and authority to follow up on recommendations of the evaluation.	Contracts often end with the delivery of the final product, typically the final evaluation report, which limits or prohibits follow-up. As outsiders, do not have authority to require appropriate follow-up or action.

2.5.4. Data matrix

A data matrix outlines the sources and types of data that will need to be collected by the program team as part of the monitoring, as well as by the evaluator at the time of the evaluation, to ensure that the evaluation questions can be answered. The data matrix should indicate which evaluations will address which questions. Each evaluation does not need to address all the evaluation questions, however, all questions should be addressed over the entire evaluation plan (process, outcome and impact evaluations). See [section 2.5.2. Evaluation questions](#) for further information on which questions are addressed in the different types of evaluations.

A data matrix can also help focus data collection to ensure that only relevant data is collected. Data collection has real costs in terms of staff time and resources, as well as time it asks of respondents. It is important to weigh the costs and benefits of data collection activities to find the right balance.⁷⁶

2.6. Budget and resourcing

When designing a program, it is important to develop an estimate of the resources that are available for evaluation and what will be required to do the evaluation well.

The resources needed for an evaluation include:

- existing data
- funding to engage an external evaluator, evaluation team or for specific tasks to be undertaken and for materials and travel
- time, expertise and willingness to be involved of staff, program partners, technical experts and the wider community, whether as part of the evaluation team, evaluation governance and/or relevant people and data sources.

⁷⁶ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

When considering data availability, look carefully at the quality of existing data and what format it is in. Also clarify the skills and availability of any people who will need to be involved in the evaluation.

There are a few ways to estimate the budget for an external evaluation:⁷⁷

1. **Calculating a percentage of the program or project budget – sometimes 1–5%:** This is a crude rule of thumb approach. Large government programs with simple evaluation requirements may be around 1%; smaller government programs with more complex evaluations – for example, detailed testing and documentation of an innovation – may be around 5%.
2. **Developing an estimate of days needed and then multiplying by the average daily rate of an external evaluator:** This can be useful for simple evaluations, especially those using a small team and a standardised methodology such as a few days of document review, a brief field visit for interviews and then a short period for report write up.
3. **Using the average budget for evaluations of a similar type and scope:** This can be a useful starting point for budget allocation providing that the amounts have been shown to be adequate (see Table 12).
4. **Developing a draft design and then costing it, including collection and analysis of primary data:** This can be done as a separate project before the actual evaluation is contracted but will usually require staff with prior evaluation experience.

Estimate the costs of collecting and analysing the data, as well as the project management and reporting time needed. Allow time to secure resources (for example, including them in an annual or project budget, or seeking someone with particular expertise). If ongoing evaluation input is needed consider a staged approach to funding.

Table 112: Estimated costs for evaluation services⁷⁸

Evaluation services	Scale of the program	Estimated cost
Design and planning for evaluation		
Capability training for internal evaluation teams (one day workshop)	any scale	\$5,000–10,000
Facilitate internal development of a program logic and the outcomes to be targeted by the recovery program	any scale	\$5,000–10,000
Evaluation of needs/needs analysis for evaluation design	large scale	\$20,000–30,000
	small or mid scale	\$10,000–20,000
Developing outcome indicators and a plan for measuring and monitoring progress toward outcomes (including planning workshop)	any	\$5,000–15,000
Developing a plan for measuring and monitoring progress toward outcomes, including development of indicators	large scale	\$15,000–20,000
	small or mid scale	\$10,000–15,000

⁷⁷ Adapted from the BetterEvaluation Commissioner's Guide <https://www.betterevaluation.org/en/node/5286>.

⁷⁸ Personal communication from Dr George Argyrous (Manager, Education and Research, Institute for Public Policy and Governance, University of Technology Sydney) based on evaluation costs in New South Wales.

Evaluation services	Scale of the program	Estimated cost
Preparing an evaluation plan for a full outcome evaluation	any	\$25,000–35,000
Supporting an internal team to develop an evaluation plan for a full outcome evaluation (providing advice, reviewing documents, providing material and resources, small workshops)	any	\$5,000–15,000
Providing ongoing evaluation support and advice to an internal evaluation team	any	\$10,000–20,000
Conducting Process and/or outcome evaluation		
Process evaluation	large scale	\$65,000–90,000
	small or mid scale	\$50,000–70,000
Support and advice for an internally-led process review	any	\$20,000–30,000
interim evaluation of program process and progress toward outcomes	large or mid scale	\$70,000–100,000
	small scale	\$50,000–70,000
Conduct a full outcome evaluation	small scale	\$50,000–80,000
Conduct a full outcome evaluation with multiple components	large scale	Over \$175,000
	mid scale	\$80,000–175,000
Outcome evaluation of a component of a larger scale program (for example, social wellbeing, business recovery)	large or mid scale	\$50,000–125,000
	small scale	\$30,000–50,000

Identify the resources that can be used for the evaluation, including – potentially – the following:

- funding to engage external individuals or organisations to design and/or conduct the evaluation or review the design and the final report
- staff time to either conduct the evaluation or to manage an external contractor
- time and goodwill of other stakeholders who will be involved in the evaluation – such as partner organisations, community members
- existing data.

2.6.1. Evaluation on a shoestring

If the resources required for the evaluation are more than the resources available, additional resources will need to be found and/or strategies used to reduce the resources required.⁷⁹ A hybrid approach to evaluation (where an evaluation is delivered using internal resources with support from specialist providers) can help keep evaluation costs down and build internal capability.⁸⁰ Careful targeting of the evaluation within the context of existing evidence can also help keep the costs of evaluation down.⁸¹

It is not feasible or appropriate to try to evaluate every aspect of a program. As such, evaluations need scope boundaries and a focus on key issues. For example:

- a program evaluation might look at implementation in the past three years, rather than since commencement
- a program evaluation could look at performance in particular regions or sites rather than across the whole Territory
- an outcome evaluation may focus on outcomes at particular levels of the program logic or for particular components of the program
- a process evaluation may focus on the activities of particular stakeholders, such as frontline staff, or interagency coordination.⁸²

Table 12: Possible options for reducing evaluation costs⁸³

Cost reduction options	Possible implications	How to manage the risks
Reduce the number of key evaluation questions	Evaluation may no longer meet the needs of the primary intended users	Carefully prioritise the key evaluation questions Review whether the evaluation is still worth doing
Reduce sample sizes	Reduced accuracy of estimates	Check these will still be sufficiently credible and useful through data rehearsal (mock-ups of tables and graphs showing the type of data the evaluation could produce)
Make more use of existing data	May mean that insufficiently accurate or relevant data are used; cost savings may be minimal if data are not readily accessible	This is only appropriate when the relevance, quality and accessibility of the existing data is adequate – need to check this is the case before committing to use

⁷⁹ BetterEvaluation Commissioner's Guide <https://www.betterevaluation.org/en/node/5286>

⁸⁰ Australian Public Sector Evaluation Network Budgeting for Evaluation Research Project <https://www.aes.asn.au/special-interest-groups/apsen>

⁸¹ In the TulaSalud Case Study (part of the Goldilocks Toolkit), Innovation Poverty Action noted that the efficacy of the program's practices were documented in medical research. Therefore, they recommended that the evaluation should focus on the training of community health workers and their ability to use the system because this was more relevant and less burdensome operationally than an assessment of the platform on health outcomes.

https://www.poverty-action.org/sites/default/files/publications/Goldilocks-Toolkit-TulaSalud-Case-Study_1.pdf

⁸² <https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/2-develop-the-evaluation-brief>

⁸³ BetterEvaluation Commissioners Guide <https://www.betterevaluation.org/en/node/5296>.

Cost reduction options	Possible implications	How to manage the risks
Embed data collection in program implementation	Might lead to a reduction in data quality	Ensure staff are trained and motivated to collect data properly and have sufficient time and equipment to do so
Use fewer waves of data collection, including possibly retrospectively created baselines	Will increase the risk of inaccurate data	Check that retrospective baselines will be sufficiently accurate and that less frequent information on progress will be sufficient to inform decisions

2.7. Stakeholder engagement

Key stakeholders in an evaluation are likely to include senior management in the agency, program managers, program partners, service providers, program participants and peak interest groups (for example, representing industries, program beneficiaries).

Involving stakeholders during evaluation planning and implementation can add value by:

- providing perspectives on what will be considered a credible, high quality and useful evaluation
- contributing to the program logic and framing of key evaluation questions
- facilitating quality data collection
- helping to make sense of the data that has been collected
- increasing the utilization of the evaluation's findings by building knowledge about and support for the evaluation.⁸⁴

Once the evaluation is completed, stakeholders need to be informed of any lessons learned and recommendations (see [section 6.1. Communicating evaluation results for further information](#)).

It can be useful to map significant stakeholders and their actual or likely questions. See the [Remote Engagement and Coordination Strategy](#) for specific guidance within a Territory Government context. There is also useful guidance and templates in the [Remote Engagement and Coordination Online Toolkit](#).

2.8. Ethical considerations

All evaluations should take into account appropriate ethical considerations. Program managers should undertake an assessment on ethical risk against guidelines such as those produced by the [National Health and Medical Research Council](#) (NHMRC) and the [Australian Institute for Aboriginal and Torres Strait Islander Studies](#) (AIATSIS), to determine if formal ethics review processes are required. The Australasian Evaluation Society has produced [Guidelines for the Ethical Conduct of Evaluation](#) which members are obliged to abide by.

⁸⁴ BetterEvaluation Commissioner's Guide
https://www.betterevaluation.org/en/rainbow_framework/manage/understand_engage_stakeholders

Specific information on ethical conduct for evaluation in Aboriginal and Torres Strait Islander settings can be found in the Productivity Commission's [A Guide to Evaluation Under the Indigenous Evaluation Strategy](#), BetterEvaluation's [Ethical Protocol for evaluation in Aboriginal and Torres Strait Islander settings](#) and NHMRC's [Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: Guidelines for researchers and stakeholders](#). Other relevant Territory Government resources include Northern Territory Health's [Aboriginal Cultural Security Framework](#) and the [Remote Engagement and Coordination Strategy](#).

It is important to consider the appropriate timeframes and budget required, acknowledging that there are often additional resource implications associated with evaluations that have specific ethical and cultural requirements.

Key ethical questions to consider⁸⁵

Conduct of the evaluators:

- Do evaluators possess the appropriate knowledge, abilities and skills to undertake the proposed tasks?
- Have evaluators disclosed any potential conflicts of interest?
- Are all evaluators fully informed of what is expected in terms of their ethical and cultural safety responsibilities and adherence to protocols?

Integrity of the evaluation process:

- Does the evaluation design ensure data is valid, reliable and appropriate?
- Have the limitations and strengths of the evaluation been identified?
- How will the source of evaluative judgements be identified and presented?
- How will results be reported and communicated in a way that all stakeholders can easily understand?
- How will the evaluation findings be utilised and how will they impact implementation?

Respect and protection of participants:

- How will participants be engaged throughout the evaluation process?
- How will participants' contributions of information, knowledge and time be respectfully recognised?
- Are there potential effects of inequalities related to race, age, gender, sexual orientation, physical or intellectual ability, religion, socioeconomic or ethnic background that need to be taken into account when analysing the data?
- How will participants' informed consent be obtained?
- What confidentiality arrangements have been put in place?
- Will the evaluation involve interviews or focus groups that may raise potential trauma?

When is external ethics review required?

NHMRC identifies triggers for ethical review, including:

- comparison of cohorts where the activity potentially infringes the privacy or professional reputation of participants, providers or organisations
- secondary use of data – using data or analysis from quality assurance (QA) or evaluation activities for another purpose

⁸⁵ Adapted from the Productivity Commissioner's Draft Indigenous Evaluation Strategy <https://www.pc.gov.au/inquiries/completed/indigenous-evaluation/draft>.

- gathering information about the participant/s beyond that collected routinely. Information may include biospecimens or additional investigations
- testing of non-standard (innovative) protocols or equipment
- comparison of cohorts
- randomisation or the use of control groups or placebos
- targeted analysis of data involving minority/vulnerable groups whose data is to be separated out of data collected or analysed as part of the main QA/evaluation activity.

“Where one or more of the triggers above apply, the guidance provided in the [National Statement on Ethical Conduct in Human Research, 2007 \(National Statement\)](#) should be followed.”⁸⁶

2.9. Evaluation risks

This section should articulate the risks or limitations that the evaluation faces, not the risks of the program in general. If significant mitigatable risks are identified, the risk assessment plan will help program managers to implement appropriate controls.

In terms of risks associated with the accuracy of the program logic, one way to combat potential overconfidence and realistically assess risk is to imagine program failure and then think through how that failure would happen.⁸⁷ It may also be useful to review previous evaluations from a similar program to identify lessons learned and how they may apply to this evaluation.

Risk categories may include: stakeholder engagement and support, technology, data, funding, timeframes, regulatory or ethical issues, physical or environmental issues.

Table 134: Example risk assessment plan

Description	Consequence	Analysis*			Current control
		Likelihood	Consequence	Risk rating	
Poor stakeholder participation in research	The evaluation would lack descriptive information about perceptions	Possible	Moderate	Moderate	A variety of intercept surveys, focus groups, telephone, internet-based surveys and information conversation methods will be used to encourage maximum stakeholder participation

*Use the Likelihood and consequence rating matrix in the [Evaluation work plan template](#).

⁸⁶ <https://www.nhmrc.gov.au/about-us/resources/ethical-considerations-quality-assurance-and-evaluation-activities>

⁸⁷ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

Choose one of the following to define the likelihood of the risk occurring.

Table 15: Risk likelihood challenges

Rating	Description
Rare	may only occur in exceptional circumstances
Unlikely	is not expected to occur
Possible	could occur at some time
Likely	would probably occur in most circumstances
Almost certain	is expected to occur in most circumstances

Choose one of the following to define the consequence if the risk occurs.

Table 16: Risk consequence categories

Rating	Description
Negligible	the consequences are dealt with by routine operations
Low	impacts on a limited aspect of the activity
Moderate	moderate impact on the achievement of goals/objectives
High	high impact on the achievement of goals/objectives
Extreme	significant impact on the achievement of goals/objectives

Use the likelihood and risk rating to determine the overall risk rating. Those that are high or extreme are likely to require closer monitoring than those that are moderate or low.

Table 17: Overall risk rating matrix

		Consequence				
		Negligible	Low	Moderate	High	Extreme
Likelihood	Almost certain	Moderate	Moderate	High	Extreme	Extreme
	Likely	Low	Moderate	High	High	Extreme
	Possible	Low	Moderate	Moderate	High	Extreme
	Unlikely	Low	Low	Moderate	High	High
	Rare	Low	Low	Moderate	Moderate	High

2.10. Reviewing the evaluation work plan

The evaluation plan should be developed well in advance of the start of the first evaluation (ideally, before program implementation) to allow for review by relevant stakeholders, making necessary changes, obtaining ethical approval (where required) and pilot testing data collection instruments (as needed).

The evaluation work plan needs to be submitted to DTF within six months of program approval. As programs may change over time, the evaluation work plan should be considered a 'living document'. It should be reviewed periodically or in response to significant program events by the program manager. DTF should be provided with updated versions in a timely manner.

Prior to and throughout the implementation of the evaluation, it is important to review the evaluation work plan to determine whether it:

- is consistent with the available evaluation resources and agreed evaluation objectives
- focuses on the most important types of information to know ('need to know' rather than 'nice to know')
- does not place undue burden on project/program staff or participants
- is ethical and culturally appropriate.

Reviewers could include: the DTF, project/program staff, internal or external evaluation experts, project/program participants, and relevant community members.

2.10.1. Technical review of the evaluation design

Before finalising the design, it can be helpful to have a technical review by one or more independent evaluators. It may be necessary to involve more than one reviewer in order to provide expert advice on the specific methods proposed, including specific indicators and measures to be used. Ensure that the reviewer is experienced in using a range of methods and designs, and well briefed on the program context, to ensure they can provide situation-specific advice.

2.10.2. Review of the design by the evaluation management structure

In addition to being considered technically sound by experts, the evaluation design should be seen as credible by those who are expected to use it. Formal organisational review and endorsement of the design by an evaluation steering committee can assist in building credibility with users.

Undertake data rehearsal of possible findings with the primary intended users where possible. This is a powerful strategy for checking the appropriateness of the design by presenting mock-ups of tables, graphs and quotes that the design might produce. It is best to produce at least two different versions – one that would show the program working well and one that would show it not working.

Ideally, the primary intended users of the evaluation will review both designs and either confirm suitability or request amendments to make the potential findings more relevant and credible.⁸⁸

⁸⁸ BetterEvaluation Commissioner's Guide https://www.betterevaluation.org/en/commissioners_guide/step5 and Patton, MQ (2011) Essentials of Utilization-Focused Evaluation.

2.10.3. Review the program logic

When reviewing the program logic, the following questions should be addressed:

- What evidence was the basis for its development? What additional evidence should be used in the review?
- Whose perspective formed its basis? To what extent and in what ways were the perspectives of intended beneficiaries and partner organisations included?
- Were there different views about what the intended outcomes and impacts were and/or how these might be brought about?
- Has there been more recent research and evaluation on similar projects and programs which could inform the program logic?

3. Engage the evaluation team

This step will explain how to select the right evaluation team and apply the procurement governance policy.

3.1. When to engage the evaluation team

The timing of engaging the evaluation team will depend on size and complexity of the evaluation. For large and complex evaluations, it may be worth engaging an evaluation team at the development stage to help complete the [Evaluation work plan template](#). For smaller evaluations, the program team may prepare the evaluation work plan and then engage evaluation teams for the evaluations, as required.

The commissioning agency should consider expertise, resourcing and cost when determining engagement timing.

3.2. Prepare the request for tender

The request for tender (RFT) outlines the purpose of the evaluation, roles and responsibilities of the evaluator and contractual requirements.

It should include:

- background
- purpose, objectives and rationale for the evaluation
- intended uses of the evaluation
- key evaluation questions
- methodology
- roles and responsibilities of evaluators and stakeholders
- reporting requirements
- timelines.

Steps for the development of the request for tender:⁸⁹

1. Clarify who will provide direct input, review and approve the RFT.
2. Decide whether the evaluation will be conducted internally, externally or by a mixed team of evaluators.
3. Determine the selection criteria for the external evaluator(s). What will be 'essential' and what will be considered 'desirable'?
4. Clarify if the evaluation design will be included in the RFT. Select from these two options:
 - a. the RFT clearly states the methodology that the evaluator will need to use
 - b. the RFT requests the external evaluator to design the evaluation and select the most appropriate methodology. It is important to specify in the RFT the values and principles that need to be upheld.
5. Draft the RFT
6. Obtain feedback from relevant stakeholders

⁸⁹ https://www.betterevaluation.org/en/commissioners_guide/step3.

7. Obtain sign-off from senior management

Territory Government [procurement framework](#) must be followed when commissioning a program evaluation. A panel contract for commissioning professional services is currently under development including RFT templates appropriate for program evaluation.

3.3. Select the evaluation team

The selection of the evaluator or evaluation team should be a transparent process.

The expertise required for the evaluation will be dependent on the type of evaluation. Table 18 below lists broad essential evaluator characteristics.

Table 148: Essential evaluator characters or qualities matched to the main purpose of the evaluation⁹⁰

Main purpose of the evaluation	Essential evaluator qualities
Accountability: Emphasis on determining the worth or merit of a project/program.	<ul style="list-style-type: none"> Should possess qualitative and quantitative expertise and experience. Independence and credibility is of central importance.
Learning: Emphasis on facilitating project/program improvements.	<ul style="list-style-type: none"> Must be reflective, familiar and comfortable with concepts of adult education and organisational learning, and willing and able to take the role of facilitator. Should possess qualitative and quantitative expertise and experience.
Innovation: Emphasis on facilitating the design of new projects/programs based on what works.	<ul style="list-style-type: none"> Should be a strong leader but also a team player. Should possess good analytical skills.

The Territory Government guide to procurement sourcing provides information on the assessment and negotiation process. The qualities across the team should also include a good mix of:

- quantitative and qualitative research skills
- multidisciplinary skills (for example, economic, demographic, environmental, sociological)
- thematic and contextual knowledge and experience
- gender and cultural balance
- language skills.

3.4. Prepare the contract

The contract should clarify:

- who will perform the evaluation tasks
- the level of contact and expectations around communication between evaluator and agency
- specific milestones, deliverables and timeframes
- agreed total cost

⁹⁰ BetterEvaluation Commissioner's Guide (step 2) <https://www.betterevaluation.org/en/node/5285>

- who owns the evaluation information and who it will be released to (including data, reports, other publications)
- legal issues such as amendments to contract or conditions for terminating the contract.

For further Territory Government specific guidance, see the guide to contract management and related documents.

3.5. Support the evaluation team

Ensure the evaluation team has clear guidance around evaluation logistics. If logistics are poorly thought out or underfunded, even the best evaluation team will not be successful.

Critical logistical planning may include: negotiation of site visit dates, objectives and on the ground requirements (such as staff availability, access to documents/data etc.); necessary notification of officials or community leaders to ensure access and cooperation/collaboration; requirements for transportation, lodging, food, office space and other facilities; requirements for translators or other specific services (such as security).

The commissioning organisation may take on the logistics for the evaluation fully or partially, regardless of whether internal or external evaluators are used. In either case, a designated evaluation logistician should be identified to ensure everything runs smoothly and to problem-solve where needed.⁹¹

⁹¹ https://www.betterevaluation.org/en/commissioners_guide/step7

4. Manage the implementation of the evaluation

4.1. Collecting the data

The manager of an internal or external evaluation team will need to ensure the [collection of data](#) follows the appropriate methodology and timeframes set out in section 5 of the work plan (see [section 2.5. Evaluation methodology](#)). It is the responsibility of the manager to respond to any issues that arise and communicate often and early with the team.⁹²

The evaluation team should have expertise in data collection and be able to provide advice on issues including data quality. It is important to ensure data quality checking is undertaken when the data is collected, such as checking datasets for completeness, spot checking survey responses, and random checks for online systems.⁹³

Collecting baseline data will allow for meaningful comparisons. Ideally, baseline data will be collected prior to program implementation. For programs that have already commenced, baseline data will need to be collected retrospectively. In the absence of available baseline data, it may be appropriate to use benchmarking against best practice research or similar existing programs.

4.2. Analysing the data

Data analysis is a crucial step in the evaluation process. It involves sorting the data and looking for patterns to create insights. Quality analyses use the most appropriate methods for purpose and present the data in a meaningful way. Further information is on the Better-Evaluation website⁹⁴.

Analysing both qualitative and quantitative data strengthens the evaluation by balancing the limitations of both data types. It is important to ensure that these are combined so that qualitative data complements and provides explanation for the quantitative data.

Data analysis should look for evidence that proves causation rather than correlation between the program and impacts. There are various tools and approaches to check causal attribution including [contribution analysis](#), which offers managers and evaluators a step-by-step approach to drawing conclusions about whether the program has contributed to particular outcomes. In addition, impact evaluations should rule out possible alternative explanations. Refer to [section 2.5.2. Evaluation questions](#) for further guidance on answering evaluation questions.

Box 2: Data visualisation

Data visualisation is the process of representing data in a way that is clear and easy to understand. The most appropriate type of graph or visualisation will depend on the nature of the variables; for example relational, comparative, time based. For further information on data visualisation options, see [Visualise data](#) on the BetterEvaluation website.

It is the responsibility of the evaluation manager to ensure data is presented in a user friendly way. It can be useful to request the evaluation team to present preliminary findings to a broader team to highlight any inconsistencies or errors.⁹⁵

⁹² https://www.betterevaluation.org/en/commissioners_guide/step7

⁹³ NSW Evaluation Toolkit <https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/6-manage-implementation-of-the-workplan-including-production-of-reports>

⁹⁴ https://www.betterevaluation.org/en/rainbow_framework/describe/analyse_data

⁹⁵ [https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/6-manage-implementation-of-the-workplan-including-production-of-reports/.](https://www.dpc.nsw.gov.au/tools-and-resources/evaluation-toolkit/6-manage-implementation-of-the-workplan-including-production-of-reports/)

4.2.1. Evidence synthesis

The evaluator should present an overall conclusion by synthesising the data and placing a value judgement on the results. There are various approaches for summarising evidence as outlined in Table 19.

Table 159: Common techniques in evidence synthesis⁹⁶

Technique	Description
Cost benefit analysis	Compares cost to benefits in monetary units.
Cost effectiveness analysis	Compares costs to the outcomes in terms of a standardised unit (for example, additional years of schooling, years of life saved).
Cost utility analysis	A particular type of cost effectiveness analysis that expresses benefits in terms of a standard unit such as Quality Adjusted Life Years.
Multi-criteria analysis	A systematic process which considers monetary impacts, material costs, time savings and project sustainability as well as the social and environmental impacts that may be quantified but not so easily valued.
Numeric weighting	Numeric scales to rate performance against each evaluation criterion to result in a total score.
Qualitative weight and sum	Qualitative ratings (such as symbols) to identify performance in terms of essential, important and unimportant criteria.
Rubrics	A descriptive scale for rating performance that incorporates performance across a number of criteria (can be qualitative and quantitative).

Evaluation findings often involve some form of extrapolation or generalisation of the data. This may involve making generalisations about the future. For example, if the evaluation finds a program is working well based on current data, is the success likely to continue.

Other evaluations such as pilot programs require recommendations to be made based on scaling up to a wider population or scope. The findings need to clearly outline the situation to which results will be generalised.

⁹⁶ Adapted from BetterEvaluation Website

https://www.betterevaluation.org/en/rainbow_framework/synthesise/synthesise_data_single_evaluation

5. Guide production of a quality evaluation report

A written report is often the final deliverable of the program evaluation. It should be discussed early in the evaluation process including likely/expected content, intended use and user/s. The target audience may range from government departments, funders, program stakeholders and the general public.

The report format, style and content will depend on the primary intended users, however, general principles should be followed for a succinct and meaningful report that promotes understanding, including:

- the executive summary contains a condensed version of the evaluation and explicitly answers the key evaluation questions
- the use of explicit evaluative language rather than 'value-neutral' language that simply describes the results
- the use of clear and simple data visualisation
- the findings section should use the key evaluation questions as subheadings
- large evaluative rubrics should be included in full in an annex and summarised in the main body of the report.⁹⁷

⁹⁷ https://www.betterevaluation.org/en/commissioners_guide/step8.

6. Disseminate results and support use of evaluation

6.1. Communicating evaluation results

Evaluation findings may be used for different purposes and should be received by key stakeholders (as outlined in [section 2.7. Stakeholder engagement](#)) as well as others who may not have a direct 'stake' in the evaluation but would benefit from the lessons learned.⁹⁸

A common approach to communicating evaluation findings is through a report, which may be internally circulated or published on the agency's website. Either a full report or summary needs to reach the intended audiences in a way that is relevant and meaningful to them.

Depending on the audience and budget, communicating results could also include:

- presentations at forums and conferences
- developing a short video
- sharing stories, photos or drawings
- creating posters or infographics.⁹⁹

[A plan for budget repair](#) noted that publicly releasing evaluation outcomes could increase trust in government and demonstrate a commitment to improving services and outcomes. If an evaluation has adverse findings, it is important to remember that it is better to find out and take corrective action rather than assume a program is working.

The communication strategy should build on the stakeholder engagement plan (see [section 2.7. Stakeholder engagement](#)) and articulate how any lessons learned and recommendations will be communicated to stakeholders. Ultimately, the decision to publicly release evaluation findings rests with the relevant agency Minister(s).

6.2. Responding to and using evaluation findings

For each evaluation report, the relevant agency should prepare a written response to the recommendations. The response might agree, partially agree or disagree with a recommendation and should provide an explanation for any partial acceptance or rejection of a recommendation.

Where recommendations have been accepted, or partially accepted, key follow-up actions should be identified, with a time frame specified and the responsible unit named. It is important to identify an individual to coordinate the overall management response and an agreed deadline by which comments must be provided (usually within two months of receiving the final evaluation report). The management response must also be provided to the DTF. Further guidance on preparing management responses is available on the BetterEvaluation website.¹⁰⁰

The [Program evaluation framework](#) outlines the responsibilities for ensuring evaluation findings are used, with roles for both central and lines agencies.

The [Cabinet handbook](#) states that agencies are expected to incorporate lessons learned from previous evaluations into program and policy design. This includes relevant evaluations from other states, territories or countries, as appropriate.

⁹⁸ BetterEvaluation, Commissioner's Guide (step 9).

⁹⁹ https://www.betterevaluation.org/en/rainbow_framework/report_support_use.

¹⁰⁰ BetterEvaluation, Commissioner's Guide, <https://www.betterevaluation.org/en/node/5303>.

CMC, OCPE and DTF have a role promoting the use of evaluation in government decision making. The PEU within DTF will prepare an annual whole of government summary of evaluation for the Budget Review Subcommittee. This will include:

- a list of the evaluations that have been completed in the previous year, including a status update on recommendation responses
- a list of any evaluations that were scheduled but did not take place
- an updated rolling schedule of evaluations for the next four years.

For the evaluations that have been completed in the previous year, the PEU will outline the recommendations from each and note whether or not they have been actioned. This will ensure that agencies respond to the recommendations from each evaluation and help close the loop between evaluation planning and evaluation use.

6.2.1. Program reality checks

There are sometimes gaps between the expectations of program proponents and the realities of program implementation. The Northern Territory Ombudsman has cautioned that:

“Government must be steadfast in its support of new approaches, recognising the realities discussed in the following table – realities that are often overlooked in the turmoil of spontaneous reaction to newsworthy events.”¹⁰¹

Box 3, sourced from the Ombudsman’s 2017 report Women in Prison II, outlines program reality checks – many of which are relevant to all programs.

Box 3: Program reality checks from the Ombudsman's 2017 report Women in Prison II¹⁰²

- No program solves every problem.
- Anyone can point to a theoretical gap or snag.
- ‘Better’ is a big step forward. Don’t expect a panacea.
- No program gets it right from the start.
- No battle plan survives contact with the enemy. Improving a program over time due to experience is a positive step, not a concession of failure.
- We all make mistakes, prisoners and staff.
- Individual failings may make a juicy story but they don’t mean a program is failing.
- Expect the best programs to be challenging and expect people to falter from time to time.
- No program works overnight.
- Monitor and evaluate but don’t dismiss a program until it has sufficient time to work.
- Don’t expect results today.

¹⁰¹ Women in Prison II – Alice Springs Women’s Correctional Facility, Volume 1 (May 2017), page 5, https://www.ombudsman.nt.gov.au/sites/default/files/downloads/women_in_prison_ii_aswcf_report_vol1-final_26.05.17.pdf.

¹⁰² Women in Prison II – Alice Springs Women’s Correctional Facility, Volume 1 (May 2017), page 5, https://www.ombudsman.nt.gov.au/sites/default/files/downloads/women_in_prison_ii_aswcf_report_vol1-final_26.05.17.pdf.

These program reality checks are consistent with the learning mindset that is encouraged in the [Program evaluation framework](#).

6.2.2. De-implementation

Keeping the program reality checks in mind, an evaluation may sometimes prompt an agency to consider partly or entirely de-implementing a program.

While detailed guidance on de-implementation is outside the scope of this toolkit, the Department of Education has developed a [De-implementation guide](#). This guide aims to assist Territory schools and Department of Education business units to reverse, reduce, replace, or rethink programs that are not evidence based. Although it has been written within an education context, the DoE De-implementation guide may help agencies consider what thoughtful de-implementation looks like in their own context.

6.3. Sharing lessons learned

DTF is responsible for maintaining a register of all completed Territory Government evaluations. Internal reports and recommendations are encouraged to be shared so that lessons learned, whether positive or negative, can help improve future programs and strengthen policy design across agencies. Visit the DTF website to view the [library of publicly available reports](#).

Templates

- [Evaluation overview template](#)
- [Evaluation work plan template](#)
- Evaluation terms of reference template (in development)
- Evaluation report template (in development)

Glossary

Term	Definition
Activities	The day-to-day tasks an organisation must undertake in order to provide a product or service. ¹⁰³
Appropriateness	The extent to which a program is suitable for achieving stated objectives in a given context. ¹⁰⁴
Assumptions	The conditions that have to hold for a certain part of a program to work as expected. ¹⁰⁵
Baseline	Information collected before or at the start of a program that provides a basis for planning and/or assessing subsequent program progress and outcomes. ¹⁰⁶
Benchmark	A reference point or standard against which performance can be assessed.
Cost benefit analysis	Compares the costs and benefits of a program in monetary terms. The difference between the present value of benefits and the present value of costs is referred to as the net present value. The program option with the highest net present value represents the most economically viable option. ¹⁰⁷
Cost effectiveness analysis	Compares the quantifiable relative costs (in dollars) and outcomes (effects) of two or more courses of action. Cost effectiveness analysis is distinct from cost benefit analysis and should be used when the benefits of a program cannot be easily quantified in monetary terms. The costs are compared with outcomes measured in natural units – for example, per life saved, per year of life gained. This process is used to identify the lowest cost means of achieving that outcome. ¹⁰⁸
Counterfactual	How individuals or communities would have fared had a program of policy not occurred (or occurred differently). ¹⁰⁹
Effectiveness	The extent to which a program achieves its objectives.
Efficiency	The extent to which a program is delivered at the lowest possible cost (technical efficiency), to the areas of greatest need (allocative efficiency) and/or continues to improve over time by finding better or lower cost ways to deliver outcomes (dynamic efficiency). ¹¹⁰
Equity	The extent to which a program meets the individual needs of participants. It can be distinguished from equality where participants are treated equally. ¹¹¹

¹⁰³ Goldilocks Toolkit, Guiding Your Program to Build Theory of Change, Innovation for Poverty Action, 2016, <https://www.poverty-action.org/publication/goldilocks-deep-dive-guiding-your-program-build-theory-change>.

¹⁰⁴ QLD Government Program Evaluation Guidelines 2020.

¹⁰⁵ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

¹⁰⁶ NSW Government Program Evaluation Guidelines.

¹⁰⁷ NSW Government Program Evaluation Guidelines.

¹⁰⁸ NSW Government Program Evaluation Guidelines.

¹⁰⁹ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

¹¹⁰ QLD Government Program Evaluation Guidelines.

¹¹¹ QLD Government Program Evaluation Guidelines 2020.

Term	Definition
Evaluation	A systematic and objective process to make judgements about the merit or worth of one or more programs, usually in relation to their effectiveness, efficiency and appropriateness. ¹¹²
Impact	The change in outcomes for those affected by a program compared to the alternative outcomes had the program not existed. ¹¹³ May also refer to longer term outcomes.
Impact evaluation	Assesses the longer-term results and impact of a program – positive or negative, intended and unintended, direct and indirect. ¹¹⁴
Inputs	The resources (funds, expertise, time) required for the delivery of activities to achieve outputs.
Meta-analysis	Comparing and combining results of many studies.
Meta-evaluation	The evaluation of an evaluation to judge its quality.
Monitoring	Tracks whether a program is being implemented as intended and whether participants are using the program as anticipated.
Objectives	Clear, measurable statements of what the program or evaluation aims to achieve. ¹¹⁵
Outcome evaluation	Assesses progress in the early and medium-term results that the program is aiming to achieve. ¹¹⁶
Outcomes	The intended (and unintended) results of program outputs. ¹¹⁷
Outputs	The products or services generated by program activities – deliverables. The provision of outputs is typically under the control of the program and is related to the quantity and quality of program implementation. ¹¹⁸
Process evaluation	Investigates whether the program was implemented according to plan. ¹¹⁹
Program	A set of activities managed together over a sustained period of time that aim to deliver an outcome for a client or client group. ¹²⁰

¹¹² NSW Government Evaluation Framework August 2013. This definition applies more to outcome and impact evaluations rather than process evaluations that tend to focus more on monitoring.

¹¹³ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

¹¹⁴ BetterEvaluation https://www.betterevaluation.org/en/themes/impact_evaluation.

¹¹⁵ QLD Government Evaluation Framework 2020.

¹¹⁶ BetterEvaluation Manager's Guide to Evaluation, accessed May 2020, <https://www.betterevaluation.org/en/node/5282>.

¹¹⁷ M. K. Gugerty, D. Karlan, *The Goldilocks Challenge: Right Fit Evidence for the Social Sector*, New York, Oxford University Press, 2018.

¹¹⁸ Goldilocks Toolkit, Guiding Your Program to Build Theory of Change, Innovation for Poverty Action, 2016, <https://www.poverty-action.org/publication/goldilocks-deep-dive-guiding-your-program-build-theory-change>.

¹¹⁹ Better Evaluation Manager's Guide to Evaluation, accessed May 2020, <https://www.betterevaluation.org/en/node/5282>.

¹²⁰ NSW Government Evaluation Framework August 2013.

Term	Definition
Program logic	Describes how the program contributes to a chain of results and visually represents how inputs and activities link to intended outcomes.
Qualitative data	Provides an understanding of social situations, people's values, perceptions and motivations. Generally presented in narrative, descriptive form.
Quantitative data	Measured on a numerical scale (for example, how many, how much, or how often). Generally presented using tables, charts and graphs.
Research	Closely related to evaluation, but can ask different types of questions that may not be related to the merit or worth of a program. ¹²¹
Sustainability	The capacity of a program to continue to deliver results into the future. Considers the social, economic, political, institutional and other conditions surrounding a program. ¹²²
Value for money	Value for money is achieved when the maximum benefit is obtained from a program within the resources available to the agency. It may not always mean the 'highest quality' program is selected. A lower cost option may be appropriate when the agency has limited funds. ¹²³

¹²¹ NSW Evaluation Toolkit.

¹²² NSW Program Evaluation Guidelines.

¹²³ NSW Program Evaluation Guidelines.