Government of **Western Australia**
Department of **Treasury**

Program Evaluation Guide

December 2020

Program Evaluation Guide

Further information relating to this report may be obtained by emailing info@treasury.wa.gov.au or program.evaluation@treasury.wa.gov.au

Contents

Introduction 1
Definitions 3
Policy 3
Program 3
Evaluation 3
'SMART' result 4
Data Analytics and Service Redesign 4
Evaluation Community of Practice 4
Programs subject to a Sunset Clause 5
Why Evaluate? 6
'SMART' Results 7
Evaluation and the Policy Cycle 8
Evaluation as part of the Budget Cycle 10
Program Evaluation 11
Rigour, Utility, Feasibility and Ethics in Program Evaluation 13
Characteristics and Scale of Evaluation 14
Types of Evaluation 17
Evaluation Process 20
The Five Stages of an Evaluation 20
Stage 1: Scoping, Consultation and Agreement 21
Define the Program's Purpose 21
Define the Purpose for the Evaluation 21
Risk Identification and Analysis 22
Stakeholder Identification and Engagement 23
Program Logic Map 24
Evaluation Questions 29
Establish a Steering Committee to Oversee the Evaluation 31
Terms of Reference 32

Stage 2: Plan Evaluation 33
Resources to Conduct an Evaluation 33
Evaluation Plan 33
Stage 3: Conduct Evaluation 35
Evaluation Questions and Data Collection and Analysis 35
Data Collection 36
Secondary Data 38
Data Analysis 39
Quantitative Analysis 40
Qualitative Analysis 40
Stage 4: Report and Recommendations 42
Use Findings to Inform Decisions 42
Evaluation Reporting 43
Communicate the Findings 44
Stage 5: Implementation Plan 45
Monitor the Program 45
References and further information 47
Appendix A: Glossary of Evaluation Terms 49
Appendix B: Program intent and 'SMART' results 51
Appendix C: Common Evaluation and Data Collection Types 53
Appendix D: Framework for developing an Evaluation Plan 57
Appendix E: Program Evaluation Plan 61
Appendix F: Questions to ask before using Secondary Data 64
Appendix G: Evaluation Report Checklist 65

Introduction

The Western Australian Government is committed to delivering programs which provide value for money for the people of Western Australia. Evaluation is a key tool for ensuring efficient and effective delivery of government services through evidence-based policy and decision making across the public sector. The Data Analytics and Service Redesign unit (DASR) within Treasury supports evaluations in the public sector.

A consistent approach to program evaluation aims to improve programs and provide more rigorous evidence of program results.

DASR will assist in the:

- provision of consistent and transparent evaluation of government funded programs;
- provision of data analytics capabilities in assessing programs; and
- development of a culture of evaluation as part of core business activity across the Western Australian public sector.

This evaluation guide has been designed to provide consistency across evaluations, improve cost effectiveness of programs, promote accountability, and to provide a platform for continuous learning. It is designed for all personnel responsible for program development, implementation and evaluation within the sector.

This guide outlines:

- evaluation as a key component of the policy cycle and the role it plays as part of the Budget and performance management processes;
- key principles of good evaluation practice;
- a strategic approach to evaluation that prioritises evaluation, and scales evaluations based on the characteristics of different sizes and types of programs;
- different types of evaluation and how they might be used;
- how to conduct an evaluation; and
- the use of evaluation findings for learning and better decision making.

A culture of evaluation and continuous improvement will assist agencies when designing future programs. Historically, there has been a limited focus on the evaluation of program delivery and results. In an environment of constrained public finances, it is essential that public funds are spent on activities that provide the greatest economic and social return.

All programs should be evaluated periodically to see whether they are meeting or are on track to meet intended results and whether those results are being achieved efficiently. That is, are the limited resources delivering value for money for the people of Western Australia?

Agencies should take a strategic approach to evaluation that prioritises and focuses effort on those programs that are new, large or potentially involve a high level of risk.

In prioritising evaluation effort, key factors to consider include:

- materiality or the size of the program;
- risk to clients, stakeholders, the agency and Government;
- alignment with agency and government priorities;
- complexity of delivery or uncertainty about program results;
- evaluation options and their expected benefits and costs;
- external requirements for review (such as programs subject to a Sunset Clause); and
- past evaluation findings.

In this environment of competing priorities and limited resources, evaluation can greatly assist in the decision making of Government and departments. It is also an effective way to demonstrate value for money whilst strengthening accountability.

By providing information on a program's contribution to results, evaluation:

- supports accountability to Parliament and the people of Western Australia by helping the Government to credibly report on the results achieved with resources invested in programs;
- informs government decisions on resource allocation by:
  - supporting reviews of existing program spending, to help Ministers understand the ongoing relevance and performance of existing programs; and
  - providing objective information to help Ministers understand how new spending proposals fit with existing programs, to identify synergies, avoid waste through duplication and articulate intended results;
- supports Directors General/Chief Executive Officers by informing them about whether programs are producing the results that they were designed to produce, at the intended cost;

- supports policy and program improvements by helping to identify lessons learned and best practices[1] as well as simultaneously improving the quality and lowering the costs of service delivery[2];
- supports transparency through baseline data comparison to current status of a program; and
- increases an understanding of the intended and unintended results of a program both within and external to an agency.

Definitions

The language and terms used for evaluation are quite specific and it is important that the terms are used consistently throughout any evaluation. Providing a brief definition of terms used for interim results and the evaluation report will enable the reader to understand what is meant by each term.

It is important to have a shared understanding of what we mean by 'policy', 'program', 'evaluation' and 'SMART' results.

Policy

In this guide **'policy'** is defined as:

A statement of principle that articulates, and aligns with, legislative, regulatory or organisational requirements[3].

Program

In this guide a **'program'** is defined as:

A group of related activities (may be called a program, project, policy, intervention, initiative, strategy or service) undertaken by or for Government that intends to have a specific impact (that is, Government is choosing to do something to achieve a result).

Evaluation

In this guide **'evaluation'** is defined as:

1 Treasury Board of Canada Secretariat 2009, Policy on Evaluation.

2 McKinsey and Company discussion paper 2011, Better for Less: Improving Public Sector Performance on a Tight Budget.

3 Freeman, B 2013, *Revisiting the Policy Cycle,* Association of Tertiary Education Management, Developing Policy in Tertiary Institutions, (21 June 2013), Northern Metropolitan Institute of TAFE, Melbourne.

The systematic collection and analysis of information to enable judgements about a program's effectiveness, appropriateness and efficiency.

'SMART' result

In this guide a **'SMART'** result is defined as:

A describable and measurable change that is derived from a cause-and-effect relationship4. 'SMART' results are the same as outcomes and are defined as Specific, Measurable, Attainable, Relevant and Time-bound.

A full Glossary of Evaluation Terms can be found in *Appendix A*.

Data Analytics and Service Redesign

Part of the Government's commitment to efficient and effective service delivery has been the establishment of evaluation support within the Data Analytics and Service Redesign team within Treasury. DASR will assist in providing transparency around whether programs are *delivering value for money* for the people of Western Australia.

Evaluation Community of Practice

A consistent approach to evaluation will be enhanced by building evaluation capability across the public sector. A central feature of this has been the establishment of a whole-of-government evaluation Community of Practice

The Community of Practice, coordinated by DASR:

- promotes the role of evaluation in delivering public value for the people of Western Australia;
- creates an understanding of the strengths of evaluation, and its role in evidence-based decision making;
- provides closer alignment between evaluation and program design;
- creates a forum for the sharing of knowledge, expertise and experiences across agency boundaries; and
- improves skills and the sharing of resources.

4 Foreign Affairs, Trade and Development 2013, Results Based Management, Canada, p. 2.

Programs subject to a Sunset Clause

The introduction of Sunset Clauses is part of the Government's commitment to implement a systematic approach to program evaluation across the public sector.

From 1 January 2014, all new programs (or the extension of existing programs) that impact the State's net operating balance by $5 million or more in any one year will be subject to a Sunset Clause.

Programs subject to a Sunset Clause will have funding allocated to the program for a finite period of time (the default period is three years). The Business case will include an evaluation plan summary, which will specify (among other things) the program results, defined using 'SMART' criteria[5].

The continuation of a program subject to a Sunset Clause (and the associated resource allocations) beyond the agreed cessation date will be subject to consideration by Government, informed by a program evaluation, conducted in line with pre-agreed 'SMART' results.

The implementing agency must ensure ongoing performance evaluation occurs because at any point Government may seek information on the achievement, or movement towards the achievement of the program's interim or final results.

Ongoing performance evaluation:

- allows agencies to justify a program's continued funding or increased funding;
- provides an opportunity to revise the program structure relative to the movement toward or achievement of interim or final results; and
- affords the Government the opportunity to periodically evaluate program performance.

All programs subject to a Sunset Clause will have the following requirements:

- a pre-agreed funding cessation date;
- initial program funding submission (that is, the submission to the ERC seeking funding for a program in the first instance) including a program evaluation plan summary and 'SMART' results; and
- the lead agency will be responsible for program implementation and ongoing evaluation. This includes monitoring program performance on an ongoing basis against the 'SMART' results.

[5] 'SMART' criteria define the program results in Specific, Measurable, Attainable, Relevant and Time-bound terms. More information on 'SMART' criteria is on page 7.

An evaluation on a program subject to a Sunset Clause will seek to find out:
• the program's achievement or movement toward the pre-agreed interim and final results; • whether the program is efficiently delivering results or whether an alternative service delivery model may be a more efficient method of program delivery; and • the impact program cessation may have on stakeholders.

Why Evaluate?

Properly planned and implemented, evaluation can result in benefits for Government, agencies, public servants and the community. Table 1 outlines the benefits of evaluation for all stakeholders.

Table 1: Benefits of Evaluation for Stakeholders[6]

Stakeholder	**Potential benefits**
Government	• Information to assist decision making. • Improved ability to achieve government priorities. • Efficient resource allocation. • Highlights achievements and opportunities to strengthen performance. • Encourages greater public trust in government.
Agencies	• Stronger basis for informing government priorities and resource allocation. • Improved service delivery and client satisfaction. • Builds an agency's reputation for innovation and continuous improvement.
Public servants	• Develops new skills and broadens experience. • More opportunity to shape public policy. • Fosters a more dynamic and creative work environment. • Recognises and rewards efforts to improve performance.
Community	• Better government services. • Informative government reporting. • Transparent and accountable government. • Public monies used more efficiently. • Greater confidence in activities of government.

[6] Australian Capital Territory Government 2010, Evaluation Policy and Guidelines, (December 2010), Policy Division, ACT Chief Minister's Department, p. 5.

‘SMART’ Results

‘SMART’ results are the achieved outcomes, observed characteristics or consequences of a program. The results should align with the overarching objective of the program and measure program effectiveness, cost effectiveness, appropriateness and efficiency.

An effective result should meet the following criteria:

Criteria for ‘SMART’ Results[7]	
Criteria	**Description**
Specific	Clear and well defined.
Measurable	The need for concrete criteria for measuring progress and to know when it has been achieved.
Attainable	Is there a realistic path to achievement? Neither out of reach nor below standard performance.
Relevant	Choosing results that matter within the constraints of resources, knowledge and time. That is, results that will drive the program forward.
Time-bound	Reasonable timeframe to achieve the goal. A time-bound result is intended to establish a sense of urgency. For example, can data be collected to ensure that it aligns with the required reporting timelines?

There are a number of characteristics that distinguish a good program evaluation. Good program evaluations:

- are crafted to assess specific goals and results of a program. That is, they are focused on key issues that will inform decision making;
- capture results that are reliable, relevant and useful. They should include the program’s strengths and limitations;
- provide results that will stand up to scrutiny, that are rigorous and answer the key evaluation questions;
- are replicable and transparent. That is, someone else could conduct the same evaluation and get the same results; and
- are time-bound. That is, the evaluation is timed to fit in with departmental and whole-of-government planning and resource allocation processes.

[7] Wikipedia 2013, ‘SMART’ Criteria.

A good program evaluation will, in addition, address the program's appropriateness and alignment with Government priorities, roles and responsibilities. It will identify the program's performance in terms of achievement of expected results and efficiency. More information on these elements is provided in Table 2.

Table 2: Elements Addressed by a Good Program Evaluation	
A good program evaluation will address:	
Appropriateness	The extent to which the program continues to address a demonstrable need and is responsive to the needs of the community.
Alignment with Government Priorities	The extent to which the program results align with: • Government priorities; and • the agency's strategic plan.
Alignment with Government Roles and Responsibilities	Assessment of the role and responsibilities for the Government in delivering the program.
Performance including: **a. Effectiveness: Achievement of Expected Results**	Progress toward expected results including reference to performance milestones, program reach and design. There should also be a linkage between the outputs/activities and the results. In addition, a good evaluation will measure and capture unexpected results.
b. Efficiency	Resource utilisation in relation to the production of outputs and progress toward expected results.

Additional information on how to develop 'SMART' results (including examples):

- Program intent and 'SMART' results, *Appendix B*;
- Federation University, Australia, *Using 'SMART' Objectives*;
- Kean University, *'SMART' Objectives*; and
- New Zealand State Services Commission, *Better Public Services - Delivering Results*.

Evaluation and the Policy Cycle

Evaluation is a key component of the policy cycle. The policy cycle involves the process of conceiving of, developing, implementing and modifying public policies[8].

[8] Davis, G & Bridgman, P 2007, *Australian Policy Handbook*, 4th edn, Allen and Unwin, Sydney, p. 130-131.

A Policy is a statement of principle that articulates and aligns with legislative, regulatory or organisational requirements[9]. Policy development is often organic, iterative and irregular, due to the inherent complexity of public policy problems and the need to address competing interests.

The policy cycle provides a formal mechanism to incorporate evaluation planning and activity into the policy development process across government. Evaluation as a stage within the Australian Policy Cycle represents the formal review of practices and results associated with policy implementation at a particular point in time.

Monitoring and Evaluation (although not formalised as stages) occur as an ongoing process throughout the policy cycle. According to Davis and Bridgeman et al., integrating evaluation into policy design and implementation adds rigour, and facilitates decisions made by a well-informed accountable decision maker[10].

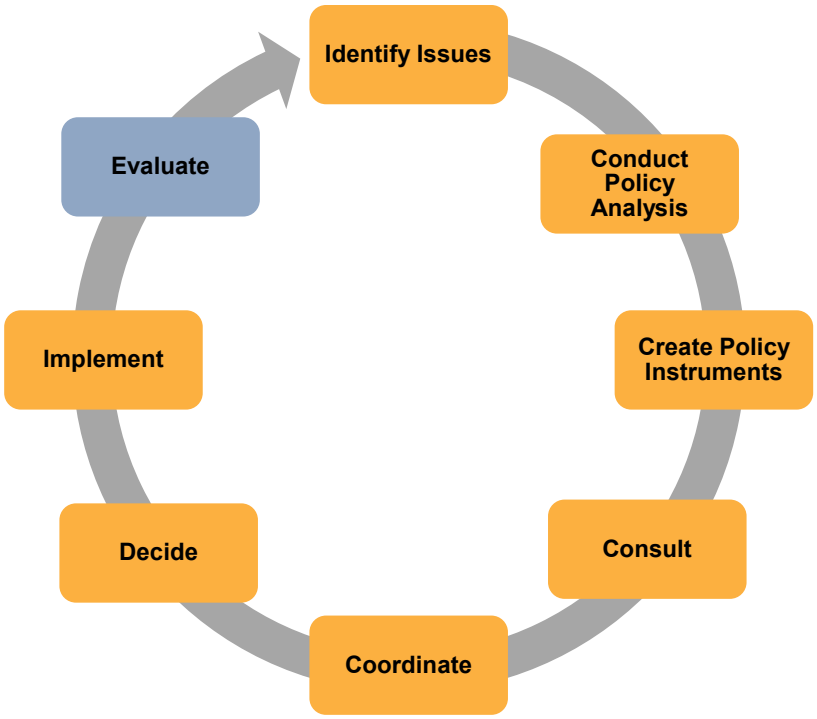
Evaluation helps determine the success of earlier steps in the policy development cycle. That is, whether the policy achieved the intended results and determines whether things can be done better in the future.

The policy cycle is not intended to encourage a process driven approach to policy development and implementation but rather to underscore the need for a planned strategic approach to evaluation. Evaluation as part of the policy cycle is outlined in Figure 1.

[9] Freeman, B 2013, Revisiting the Policy Cycle*, Association of Tertiary Education Management, Developing Policy in Tertiary Institutions*, (21 June 2013), University of Melbourne.

[10] Ibid. 8.

Figure 1: Evaluation and the Policy Development Cycle[11]



Source: Davis, G & Bridgman, P 2007, *Australian Policy Handbook*, 4th edn, Allen and Unwin, Sydney, p. 130-131.

Evaluation as part of the Budget Cycle

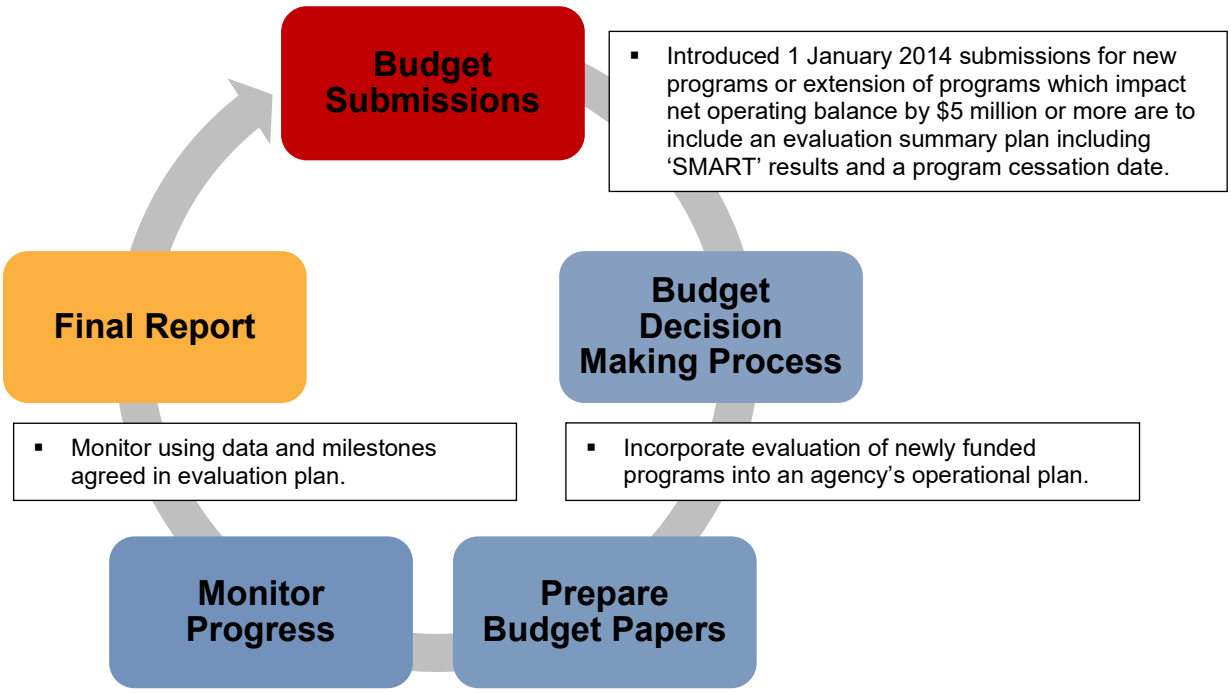
Where possible, evaluation activity should be aligned with the annual Budget process. This includes:

- supporting Budget initiatives with an evaluation of existing activities, particularly proposals to continue, extend or expand existing funding (for example, programs subject to the Sunset Clause policy);
- outlining the proposed approach for evaluating any initiatives submitted for consideration as part of the Budget process — with the evaluation approach appropriately designed to reflect the size, risk and priority of the proposed initiative;
- identifying appropriate resourcing requirements including the estimated cost of the evaluation where evaluation is deemed necessary or given a high priority by the Government; and
- incorporating the evaluation of newly funded initiatives into an agency's operational plan.

Figure 2 highlights the role of evaluation during the Budget cycle.

[11] Ibid. 9.

Figure 2: Evaluation and the Budget Cycle



Source: Department of Treasury, Western Australia 2013, Budget Cycle.

Program Evaluation

Evaluation is an integral part of the program lifecycle from design and piloting of a program through to implementation and ongoing mainstream delivery. All programs should be evaluated on a regular and systematic basis. For evaluation to be valuable to decision makers across government there needs to be consistency with its planning and execution through all stages during the life of the program.

Program evaluation should be:

- built into program design;
- methodologically rigorous, with appropriate scale and design;
- conducted with the right mix of expertise and independence;
- timely to support and influence decision making; and
- transparent and open to scrutiny.

Every program evaluation will be different and there is no 'one size fits all' evaluation plan which can be used for all programs as they vary in size and structure. It can be useful to categorize programs into three levels; mega, macro and micro programs.

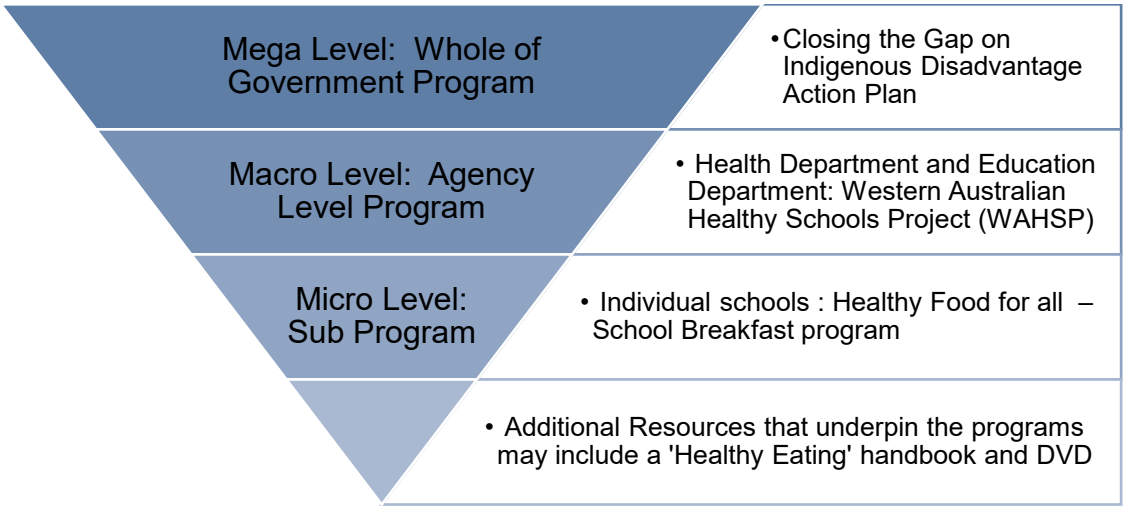
- Mega level whole of Government programs are typically multifaceted and are comprised of many sub-programs delivered by multiple agencies and/or in partnership with non-government organisations. They can be large and significant such as a whole of government approach to reducing Indigenous Disadvantage with the 'Closing the Gap on Indigenous Disadvantage Action Plan'. Evaluation planning at this level is likely to be in terms of overall economic or social impact.

- Macro level programs refer to agency specific programs such as the 'Western Australian Healthy Schools Project (WAHSP)'. This is a multi-agency initiative between the Departments of Education and Health developed for regional areas within Western Australia. At this level, programs provide specific interventions with a tighter link to results. Programs are often delivered at multiple sites by one or multiple Government departments or through alternative service providers.

- Micro level programs may be the responsibility of agency business units or individuals. An example is the 'Healthy Food for All – School Breakfast Program' run by Foodbank WA.

The level of a program can have implications for approaches to evaluation. There are often links between macro and micro programs with micro programs forming part of a macro level program. Evaluations of this type are known as multi-level multi-site evaluations. This poses a methodological challenge for evaluators who must be mindful of this when designing evaluations intended to be responsive to audiences at each program level[12]. Figure 3 provides examples of programs at the 'Mega', 'Macro' and 'Micro' level.

Figure 3: Examples of Programs at each Level



Rigour, Utility, Feasibility and Ethics in Program Evaluation

The design of all program evaluations must balance rigour, utility, feasibility and ethical standards.

- **'Rigour'** in evaluation refers to the quality of the evidence, and the validity and certainty around the findings. For results driven evaluations in particular, rigour includes assessing the extent that observed results were due to the program.
- **'Utility'** refers to the scope for evaluation users to actually use the findings, particularly when information is needed at a certain times to inform decisions.
- **'Feasibility'** refers to the practicalities of collecting evidence in relation to the maturity of the program, and to the availability of time, skills and relevant data.
- **'Ethics'** refers to reducing the risk of harm from the evaluation and also doing the evaluation in a culturally appropriate way[13]. The Australasian Evaluation Society has produced Guidelines for the Ethical Conduct of Evaluation. The Guidelines are designed

[12] Owen, J 2006, *Program Evaluation – Forms and Approaches*, 3rd edn, Allen & Unwin, NSW, p. 27-30.

[13] New South Wales Government, 2013, Principles and Standards for Evaluation, Department of Premier and Cabinet.

to be used as a framework for discussing ethical issues, and for helping people to recognise and resolve particular ethical issues that may arise during the course of an evaluation.

Additional information to provide a guide to resolving ethical issues:

- Yarbrough, D Shulha, L, Hopson, R et al. 2011, The Program Evaluation Standards, 3rd edn, Sage Publishers, Inc.
- Australian Government 2013, National Statement on Ethical Conduct in Human Research, National Health and Medical Research Council (will assist in research involving people); and
- Better Evaluation 2013, Ethical and Quality Evaluation Standards.

Characteristics and Scale of Evaluation

As government policies differ, careful consideration is needed on the appropriate level, form, frequency and scope of evaluation. An evaluation approach should be commensurate with the size and importance of the program. For example, more expertise and evidence from an evaluation may be required for programs which are of high strategic importance, large, complex and costly.

As a general principle, each evaluation needs to be tailored and scaled to fit according to a program's:

- scope/size/complexity;
- stakeholder impact/expectations;
- community prominence;
- stage (program design, implementation and delivery or closure);
- social/economic/environmental impacts;
- risk factors;
- value of the investment; and
- resources available[14].

Decisions need to be made about:

[14] Queensland Government 2000, *The Queensland Policy Handbook: Governing Queensland*, Office of the Director General, Department of the Premier and Cabinet, Brisbane.

- what information, data collection and evaluation methodology would provide the evidence to best inform decision makers? This includes decisions about the target groups for the evaluation, sample size, and timeline for data collection; and
- what methods and data will produce the best evidence to support the intended purpose of the evaluation?

The strength of findings, conclusions, and recommendations about program implementation and results depends on well-founded decisions regarding evaluation measurement and design[15]. In addition, the timing of an evaluation is important. If an evaluation cannot be completed in time to affect decisions to be made about the program, the evaluation will not be useful[16].

Table 3 illustrates how an evaluation may be scaled depending on the assessed priority of a particular program.

Table 3: Program Characteristics and Relative Scale of Evaluation

Risk	Program Characteristics	Scale of Evaluation
Low	• Limited investment of resources. • Low strategic priority. • Similar to previous programs that have been evaluated and found to be successful. • Responsibility of a single Minister or agency. • Simple design. • Not widely publicised.	Evaluate at agency Director General discretion. May be a less formal review process with limited data needs and few resources allocated. Regardless of intent to evaluate, it is recommended that the following program information be included to determine the value of the program: • program rationale, objectives; • governance arrangements including the Terms of Reference; • budget; • risks; • milestones; and • performance measures including 'SMART' results and Key Performance Indicators.
Moderate	• Moderate investment (relative to agency total expenses or under $5 million dollars per annum impact on State net operating balance). • Moderate to high risk.	Agency Director General discretion to evaluate and decide: • at what point to evaluate (may be summative only);

[15] Ibid. 6.

[16] Wholey, J et al. 2010, *Handbook of Practical Program Evaluation*, 3rd edn, Jossey-Bass, San Francisco, USA.

	• Responsibility of one or more Ministers or agencies. • Complicated design, with multiple elements (or sub-programs). • Involvement of external stakeholders or delivery partners. • Not recently reviewed (in the case of existing programs).	• whether to evaluate internally or contract out; • if contracted out, whether to manage contract through agency's evaluation unit (if applicable); • budget source and amount; • the need for a Steering Committee made up of internal and external stakeholders; and • production of evaluation plan and progress reports.
High	• Significant investment (relative to total government expenditure or total expenses for program has over $5 million dollars per annum impact on State net operating balance). • Resource intensive. • High priority (at whole of government 'mega' level). • Complex. • Possibly controversial. • May be innovative, a pilot, trial or need proof of concept. • May involve multiple delivery partners. • External reporting is an evaluation requirement.	• Formal evaluation mandatory. **For programs subject to a Sunset Clause:** • Evaluation plan agreed by Treasury (as per Sunset Clause requirements) with 'SMART' results and agency responsibility for data collection; • consider quarantining evaluation budget between 1-5% of total program costs; and • evaluation by agency or consultants. **For priority area evaluations:** • Director General Steering Committee involvement; • peer review, perhaps by evaluation expert in another jurisdiction; • consultation with responsible Ministers and DGs/CEOs; • support and advice from DPC/Treasury; and • results reported including an implementation plan to relevant Ministers and Government through an evaluation report.

Types of Evaluation

There are three broad types of evaluations which may be used at various stages of the program lifecycle. These include formative, process and summative evaluations which are conducted before, during or after a program has been implemented, depending on the purpose.

For example:

- at the design stage of a program a pre-program implementation (formative) evaluation is used to inform decisions to proceed and support budget priority decisions;
- during-program delivery (process) evaluations are used to monitor program performance and identify possible improvements; and
- at the closure stage of a program, a post-program implementation (summative) evaluation is used to assess achievement of program results, inform and improve policy.

There are many possible *approaches to program evaluation*, some of which are specified by government policies and processes. Evaluation approaches are generally developed to address specific evaluation questions or challenges.

Table 4 describes some formative, process and summative types of evaluation. It illustrates different approaches to evaluation and the types of questions that may be addressed with each approach.

Table 4: Approaches to Evaluation and Key Questions[17]

Type	Description	Example questions to be addressed
Formative/ Developmental	Formative evaluations provide information on how a program may be developed (new programs) or improved (both new and existing programs). Examples include: ***Needs Assessment*** – to determine who needs the program, the scale of need and what may meet the need.	• What is the problem to be solved? • What are the characteristics and needs of the target population? • Where is the problem? • What is the most appropriate plan of action to address the problem?

[17] Adapted from: New South Wales Government 2013, Evaluation Framework, Department of Premier and Cabinet, p. 7.

	Program Logic Map – to ensure a clear picture of how and why the program will produce the expected results. **Business case**: To define the program, delivery methods, the target population and the possible results. ***Evaluability Assessment*** – to determine whether an evaluation is feasible and how stakeholders can help shape its usefulness. This is useful if implementation has commenced without an evaluation plan. ***Cost-Benefit Analysis*** – compares the cost of the program with the dollar value of most of the program's benefits. Allows for program decisions to be analysed.	• Is government intervention appropriate? • How would we measure/recognise success?
Process	Process evaluations help to differentiate ineffective programs from failures of implementation. ***Process Evaluations*** – measure what is achieved by the program and investigates how the program is delivered. Can be used to improve programs by informing adjustments to delivery. Alternative delivery solutions may be investigated. Example includes: **Collection of Descriptive Statistics** – (age, race, marital status, education, income, number of children etc) to identify who benefits most from the program. **Process outcomes** – description of the status or condition of participants after they participate in the program. **Process evaluation tools** – include: process logs, attendance sheets, mailing list, telephone call log and participant satisfaction survey.	• How is the program being implemented? • Are the activities being delivered as intended? • Are program participants being reached as intended? • Could activities be delivered by an alternative method or means?

	Case Study – a method for developing a complete understanding of a process, program, event or activity. A common element of a case study is systematic and detailed data collection from multiple sources, particularly first hand experiences[18].	
Summative/ Impact	Summative evaluation reports when the program has been running long enough to produce results. Examples include: ***Outcome Evaluation*** – determines whether the program has caused the intended effect on the target results. ***Cost Effectiveness Analysis*** – is a technique that relates the costs of a program to its key results or benefits. Cost Effectiveness Analysis is often used for Health programs. ***Cost-Benefit Analysis*** – allows for program decisions to be analysed. Both Cost Effectiveness and Cost-Benefit Analysis address the questions of efficiency by standardising results in terms of their dollar value to answer questions of value for money.	• What are the net effects of the program? • Is the program achieving/has it achieved the intended results? • To what extent can changes be attributed to the program? • Is the program the best use of the resources that it costs? • What would be the impact of cessation? • What alternative service delivery options, if any, can be considered?

Additional information on evaluation and data collection types including data uses, limitations and benefits: *Appendix C*.

[18] Wholey, J et al. 2010, *Handbook of Practical Program Evaluation*, 3rd edn, Jossey-Bass, San Francisco, USA.

Evaluation Process

For evaluation to be valuable to decision makers across government there needs to be consistency between its planning and execution.

The process of every evaluation includes five stages which form the basis of the design and implementation of evaluation activity (see Figure 4). Every program will have different evaluation requirements. For most evaluations (as with any project), the early stages (planning) are critical and may need a substantial investment in time.

The Five Stages of an Evaluation

Stage 1: Scoping, Consultation and Agreement

Understanding the underlying assumptions and rationale of the program through the use of a Program Logic Map will assist in developing the program evaluation plan. Risk identification and analysis along with stakeholder identification and engagement are in addition, essential components of this stage. Defining the purpose of the evaluation will direct resources and focus the evaluation to collect data which will answer the key evaluation questions.

Stage 2: Plan Evaluation

This stage is characterised by the creation of an evaluation plan and a formal terms of reference document to specify how the evaluation will be conducted.

Stage 3: Conduct Evaluation

The monitoring of key milestones throughout the program through data collection and analysis is key to conducting the evaluation.

Stage 4: Report and Recommendations

Use evaluation findings for learning and decision making. Timely communication of the evaluation findings to inform decisions are a key element of this stage.

Stage 5: Implementation Plan

An implementation plan is a guide for further developing the program. It should include a timeline of actions or activities as a result of the evaluation.

Stage 1: Scoping, Consultation and Agreement

Define the Program's Purpose

The first step to planning any program evaluation is to clearly describe how the program is intended to work. This includes a clear description of why the program is needed, outlining the goals, objectives and results. In some cases, it may be important to review the issues the program aims to address, through a needs assessment.

A needs assessment and description of the program will provide a greater understanding of the program and allow the evaluation to be tailored to the program's original scope.

A good understanding of the program will help to identify why the evaluation is needed, how the evaluation will be used and by whom.

Evaluation planning is vital when developing a program's business case (particularly for large multifaceted programs or those subject to a Sunset Clause). This is to ensure that data is collected for example, before program implementation (for baseline comparison) and on an ongoing basis. The collection and monitoring of data will track and potentially facilitate the improvement of the program.

Define the Purpose for the Evaluation

The method of evaluation, communication, timing and use of results largely depends on the intended purpose of the evaluation.

It is critical to have a clear understanding of the evaluation's purpose to enable a decision about what you want to evaluate.

Program evaluations are usually conducted to:

1. refine an existing program with information to help redesign or improve a program; and
2. measure progress and improve accountability through regular monitoring of progress toward program results to report to key stakeholders.

Without a clear direction on what is to be evaluated, the evaluation may waste resources (both time and money), will lack direction and the resulting information may not be as useful as it could be.

Questions that may assist to identify program intent, rationale and purpose of the evaluation include:

- what are the assumptions behind the program?
- what is the rationale of the program? This includes the program's overarching goal and intended results; and
- what is its underlying logic? What issue is the program trying to address? Who are the recipients of the service? What outputs are being delivered in terms of quantity, quality, timeliness and costs?
- who will use the results of the evaluation (audience)?
- how will the evaluation be used? Will the evaluation be used to:
  - identify cost savings (as part of budget measures) to reallocate resources?
  - improve and inform policy with evidence-based information?
  - improve accountability and transparency?
  - inform budget deliberations?

Risk Identification and Analysis

A risk is an uncertain event that may have a positive or negative impact on a program and its evaluation. Unexpected events can affect a program's implementation and therefore its evaluation.

Risk identification allows the program evaluation team to proactively approach potential or real problems, rather than defensively respond to problems after they occur.

There are many ways evaluators may determine risks. Some *methods* include brainstorming, focus groups, experience judgement, flow charts, Strengths, Weakness, Opportunities and Threats (SWOT) analysis and scenario building.

Risks may be categorised under broad headings[19]. Table 5 provides examples of questions which may help to identify risks under each category.

Table 5: Risk Categories and Identification Questions

Risks	Example questions to determine risk factors
People	Will team members be available when needed? Can they work together?
Technology and data collection	Are data collection and analysis tools available?
Organisational and program	Do all stakeholders agree on the program's objectives and purpose?
Funding/finance	Has there been a budget set aside for the evaluation? Are there enough funds to complete the evaluation?
Law or contract	Are there any regulatory or ethical issues that may impact the evaluation?
Physical and/or environmental	Are evaluators likely to get hurt while conducting the evaluation?

It may also be useful to consult relevant historical information from previous similar program evaluations that may include lessons learned which describe problems and their resolution.

Stakeholder Identification and Engagement

Identifying and engaging with the key program stakeholders to clarify the evaluation's purpose is critical to the success of every program evaluation[20].

Consideration of stakeholder engagement(including the fostering of input, participation and power sharing among those with an investment in the conduct of an evaluation and its findings) is an important first step when planning an evaluation.

If stakeholders are involved in the planning stages of the evaluation (including the development of the evaluation questions), not only will it help to identify what they want to know, they will more likely commit to the evaluation and use the results of the evaluation for decision making.[21]

Stakeholders may be classed as external or internal stakeholders.

---

[19] Adapted from: New Horizons Classroom Learning 2009, *Project Management Fundamentals*, 2nd edn, p. 26-27.

[20] Wholey, J et al. 2010 *Handbook of Practical Program Evaluation*, 3rd edn, Jossey-Bass, San Francisco, USA, p. 40.

[21] Preskill, H and Jones, N 2009, A Practical Guide for Engaging Stakeholders in Developing Evaluation Questions, Robert Wood Johnson Foundation Evaluation Series.

External stakeholders for an evaluation include the program's target audience (that is, the users or beneficiaries of the program), partners involved in developing and delivering the program, ethics committees and the general public.

Internal stakeholders include the Government, Agency(s), Board of Directors or Governance bodies and potential data custodians.

Stakeholder Analysis is a tool used to identify the key people associated with a program and its evaluation.

Additional information on Stakeholder Analysis:

Svendsen, A 1998, *The Stakeholder Strategy, Profiting from Collaborative Business Relationships*, Berrett-Koehler Publishers, USA.

Program Logic Map

Program Logic is a tool to describe how a program is intended to work. A Program Logic Map unpacks the logic underpinning the program being evaluated – not the evaluation itself. Program Logic identifies what the program creators were thinking when they designed the program. Program Logic helps program teams consider how results can best be achieved, to articulate a clear narrative for the choice of initiative and will later help with monitoring, evaluation and the reporting on progress. Program Logic is often represented in diagrammatic form and show a sequence of expected consequences that show the pathway to change.

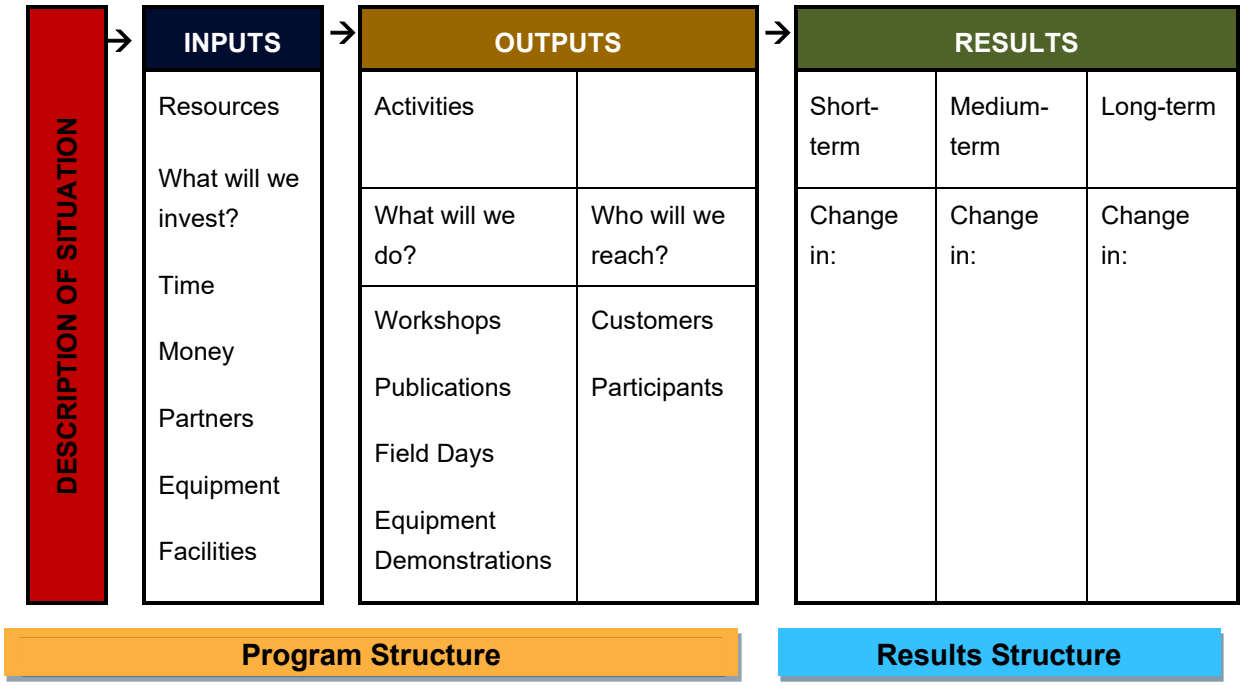
A Program Logic Map describes the logical causal links between the program's resources, activities, outputs and short/medium/long-term results (see Figure 5).

Program Logic Maps are narrative and graphical depictions of processes that communicate the underlying assumptions upon which the program is expected to lead to a specific result. A Program Logic Map shows a series of consequences, not just a sequence of events.

The four main elements of a Program Logic Map include:

1. Description of situation (context): The context is the situation that necessitated the development of the program. That is, why is the program needed? Is the understanding of the context correct?
2. Inputs: Resources of time, money, partners, equipment and facilities required to implement the program and address the situation.
3. Outputs: The activities of the program. This includes an outline of how the activities will be delivered and who will be the target audience of the program.
4. Results: The results describe the changes that will occur over the short, medium and long term as a result of the program. They are described using 'SMART' criteria.

Figure 5: Four main elements of a Program Logic Map



A Program Logic Map will:

- assist all program stakeholders (planners, implementers, funders, participants etc.) to come to a common understanding of what a program is intended to achieve;
- diagrammatically un-package complicated programs on one page;
- clarify and communicate the nature of the activity;
- question the strength of the assumptions underpinning the activity;
- guide the planning of the activity's implementation;
- inform the development and focus of the evaluation; and
- interpret the findings.

A Program Logic Map may be built through asking the following questions:

- what is the current situation that we intend to impact?
- what will it look like when we achieve the desired situation or results?
- what behaviours need to change for results to be achieved?

The creation of a Program Logic Map requires back mapping. That is, to start with the intended long-term results and map backwards or downwards (depending on the format of the map) to think through what prior conditions are required to achieve results at each level. Eventually, the specific activities as a precursor to the results are considered.

When a Program Logic Map is created in a participatory manner, it helps groups of stakeholders to come to consensus and a common understanding about the activities of the program[22].

Mapping results achievement in this way enables:

- the identification of results at each stage of the process; and
- the analysis of risks which might interfere with the attainment of the results.

A Program Logic Map is a useful tool for cross agency initiatives where multiple agencies have a role, as the results chain clearly articulates each agency's input and contributions to the achievement of the overall government objectives.

Figure 6 provides an example of how the results of a Juvenile Detention Education Program are illustrated using a Program Logic Map.

Additional information on how to create a Program Logic Map:

- Wholey, J et al. 2010, *Handbook of Practical Program Evaluation*, Jossey-Bass, San Francisco, p. 56-79.
- Better Evaluation, Program Logic Model.
- New South Wales Government 2006, What you do and why – An agency guide to defining results and services: Program Logic. Department of Treasury.
- Victorian Government, Evaluation Toolbox: Program Logic, Department of Sustainability and Environment.
- New South Wales Government 2004, Outcome Hierarchy and Program Logic, Department of Environment and Conservation, p. 12, 44-45.
- W.K. Kellogg Foundation 2006, Logic Model Development Guide.

[22] Clear Horizon WA 2013, *Introduction to Monitoring and Evaluation Frameworks*, Learning and Development Program Workbook, WA Institute of Public Administration, p. 18.

Figure 6: Program Logic Map for a Juvenile Detention Education Program

DESCRIPTION OF SITUATION/NEED FOR PROGRAM	INPUTS	OUTPUTS	RESULTS		
Program Result: Reduce recidivism rates of juveniles in detention.	**Resources** • $5 million p.a. sought to contract teaching services from the Dept. of Education. • $5 million is sought for additional security to oversee additional juvenile visits to schools. • $10 million to upgrade facilities at schools to accommodate the extra students. • Extra teachers to assess students' educational level. • Supply of teaching services already available for use. • Additional security from private sector to oversee the increased number of school visits by juveniles.	**Program participants:** Low risk Juvenile detention inmates.	**Short-term (first 12 months)**	**Medium-term (2 years)**	**Long-term results (after 3 years)**
Current juvenile detention education programs have led to poor educational results for juveniles in detention, which are proven to be linked to higher recidivism rates. Rationale: • Align juvenile detention education program with school curriculum. • Greater engagement with education sector to determine juvenile's level of education. • Low risk juveniles to be taught at school sites instead of detention centres.		**Activities** • Juveniles attend mainstream school. • Assessment of student's educational level. • Tailoring the educational environment to the student's educational level. • Juvenile's engagement in internal educational program will be recorded by teachers and tracked by the pre-existing Dept. of Corrective Services unit.	• Return to formal education is 10% higher for participants in comparison to non-participants. • Improved mental health of participants measured by a 10% reduction in doctor visits for mental health issues as a proportion of participants.	• Recidivism rates for program participants are 50% lower than for non-program participants. • Number of program participants increases to 75% (an increase of 5% per annum).	• Total recidivism rates for all juveniles in detention falls from current 75% at program start to 50% after three years. • Well-being of juveniles' has increased. • Juveniles' ability to secure job after detention has increased by 5%.

• High risk juveniles will remain on-site. • 'Ex-students' will be given the opportunity to return to a detention-site.	• Transport costs. • Minor upgrades to schools to accommodate juvenile detention students.	• Information collation between the Dept. of Education and the Dept. of Corrective Services is required.			
PROGRAM STRUCTURE			**RESULTS STRUCTURE**		

Note: Recidivism is the repeated or habitual lapse into crime.

Evaluation Questions

Evaluation questions are developed from the Project Logic Map. Once the purpose of the program and evaluation has been identified, the development of *evaluation questions* is essential.

Evaluation questions focus the evaluation and drive decisions on appropriate data collection and analysis while also helping to determine the type of evaluation (formative, process or summative) required.

As different types of evaluation provide different information and support different decisions, it is important to plan upfront what questions need to be answered, how they will be answered, and by when. Asking questions relevant to stakeholders will maximise the use of the evaluation. One of the most important questions to be answered is: what do the stakeholders want to know about the program?

Evaluations are usually undertaken to provide answers to a wide range of questions. Typically, evaluation questions focus on issues of effectiveness, appropriateness and efficiency and will assist in the assessment of the overall performance and worth of a program. Some questions may be used to measure more than one criterion.

- Effectiveness refers to the extent to which a program's intended results have been achieved.
- Appropriateness compares the environmental context or need on the one hand with the objectives and strategies of the program to determine whether the latter are relevant to the former.
- Efficiency refers to the extent to which activities, outputs (products, services) and the desired results are achieved with the lowest possible use of resources/inputs (financial, people, time).

Table 6 provides examples of questions to measure effectiveness, appropriateness and efficiency. Note: Some questions may be used to measure more than one criterion.

Table 6: Example Questions to Measure Effectiveness, Appropriateness and Efficiency
To measure effectiveness:
Does the program meet its stated goals/objectives?
Is the program producing intended/unintended results?
What changes would help the program better target stakeholder needs?
Does this program meet Government objectives/goals?
What is the cost/benefit of the program?
Could reallocating resources from lower priority programs fund it?
To measure appropriateness:
Is there an ongoing community need for the program?
Should the current program be maintained, expanded, or discontinued?
Is the program aligned with Government priorities?
Does the program represent a legitimate role for government?
Has the program been implemented as planned?
Is there a case to roll out the program to a wider geographical area or population group?
Should the government continue to fund the program, or is there a better alternative service provider?
Does the program take sufficient account of emerging trends and new developments?
To measure efficiency:
Is the cost of the program commensurate with the perceived benefit to stakeholders?
How do the unit costs compare with those of like activities in other programs or jurisdictions?
Can resources be allocated more efficiently by modifying a particular program or a mix of programs to achieve the same result?
Do the results of the program represent value for money?
Could others provide the services more efficiently?

A specific, measurable, attainable, relevant and time-bound 'SMART' result answers the question (provides the evidence) how will I know it?

Table 7 outlines the role evaluation questions play in establishing whether results have been achieved. It provides an example of an evaluation question in relation to a program aimed at reducing road trauma.

Table 7: Hypothetical Example of the Assessment of a Road Trauma Reduction Program through an Evaluation Question

Evaluation Question	How will I know it? (the 'SMART' result)
Is the cost of the road safety program commensurate with the perceived benefit to stakeholders?	• Reduction in road trauma by 10% compared to base data 12 months after implementation of the road safety program. • A 10% reduction in hospital costs related to road trauma 12 months after road safety program implementation compared to the base costs. • 5% reduction in death associated with road trauma 12 months after program implementation.

Establish a Steering Committee to Oversee the Evaluation

For large, costly or complex program evaluations, the establishment of a Steering Committee will assist in overseeing and providing guidance on the evaluation.

To support the Steering Committee, working groups may be formed to scope the evaluation, develop the Terms of Reference and provide a supporting business case for Steering Committee endorsement. The working group may oversee the development of the evaluation plan as well as the evaluation. The role of the program's lead agency will be to collect and provide the data for the evaluation as required by the evaluation plan.

Terms of Reference

The Terms of Reference for an evaluation is a formal agreement amongst key stakeholders that defines all aspects of how an evaluation will be conducted. This is regardless of whether the evaluation is conducted internally or externally.

The Terms of Reference agreement:

- defines the objectives and the scope of the evaluation;
- outlines the responsibilities of the evaluation team;
- provides a clear description of the resources available to conduct the evaluation, including the budget;
- describes the purpose and timing of the evaluation, including the key questions to be answered;
- outlines the structure of the evaluation report that will present the evaluation findings; and
- serves as a basis for a contractual arrangement with one or more evaluators and sets the parameters against which success of the evaluation can be measured.

Additional information on Terms of Reference: Better Evaluation, Writing Terms of Reference: A How to Guide.

Stage 2: Plan Evaluation

Resources to Conduct an Evaluation

In planning a program, consideration needs to be given to the human, material and other resources needed for program implementation and for its evaluation.

Considering *evaluation resources* allows for an evaluation component to be included in the program's budget. The evaluation budget should include a rate for the evaluator, travel costs, costs of materials, and costs for any additional expertise or specialised services.

Consider using an *Evaluation Budget Matrix* to specify various items that need to be costed as individual line items. Developing an Evaluation Budget Matrix involves thinking through the cost implications of the evaluation design.

Some organisations have a policy of setting a certain percentage of the total program budget for evaluation. Common budget estimates range between 1-5% of program costs. The budget required for a program evaluation varies due to a number of factors including the role of the evaluator, the scope and size of the program and its total cost, risk and priority to Government and any other special requirements[23].

Evaluation Plan

For evaluation to be valuable to decision makers across government and to be sensitive to the audience at each level of a program (mega, macro and micro) there needs to be consistency between its planning and delivery.

An evaluation plan maps out the strategic evaluation methodology for a particular program.

An evaluation plan incorporates the overall purpose for the evaluation, including the main evaluation projects and reports over the period of the program, and related activities such as the data collection techniques unique to the particular evaluation and the process used to build capacity for evaluation.

Along with the Terms of Reference, a detailed evaluation plan should be developed during the design phase of a program.

For programs subject to a *Sunset Clause,* the evaluation plan must accompany the initial funding submission.

[23] Australian Government 2013, What Resources and Skills are Required to Conduct the Evaluation? Office for Learning and Teaching.

Evaluation plans are organic and likely to evolve during the evaluation process.

Principles for developing an evaluation plan:

- develop the plan early as part of the design of the program and in alignment with the Program Logic Map;
- identify key questions to be answered, the material and data needed to answer those questions and the methods for collecting and analysing the data. Review existing evaluation frameworks and results for similar programs to help define the evaluation questions and identify reliable sources of data; and
- engage with decision makers/stakeholders when developing the evaluation questions to help ensure that the evaluation fulfils its purpose and delivers results that are useful to all groups.

While evaluation plans can have a variety of formats, it has been found that certain headings are helpful to provide a logical plan that will be useful to stakeholders. The scale, complexity and cost of a program will dictate the depth and complexity of the evaluation plan.

All evaluation plans should include:

1. program intent and rationale;
2. identification of key stakeholders;
3. risk identification and analysis;
4. Program Logic Map;
5. key evaluation questions;
6. Evaluation methodology and data collection methods
7. 'SMART' results and how they will be measured; and an
8. evaluation budget.

Additional information on a framework for developing an evaluation plan: Appendix D. An evaluation plan template: Appendix E.

Stage 3: Conduct Evaluation

Responsibility for conducting the evaluation is an important consideration. A trade off may need to be made between using external or internal resources to evaluate the program.

The program evaluator should be capable, competent and independent of the direct program delivered. The evaluator does not always need to be external to the department or Government, but simply external to direct involvement in the delivery of the program in order to maintain objectivity.

If an external evaluator is to be contracted, it is important to decide who within the organisation will be responsible for facilitating the process. For programs subject to a Sunset Clause, Treasury and/or the Department of the Premier and Cabinet may require briefings from the external evaluator on methodology, assumptions and the final report. Contracts with such evaluators need to consider this requirement.

Every member of the evaluation team (whether the team is internally or externally managed) should possess distinct skill sets and contribute collectively through the life of the evaluation to achieve the common goal. Having a team with appropriate backgrounds, capabilities and experiences is essential. It is possible that members of an evaluation team may belong to different functional teams and work together for only the life of the particular evaluation.

Evaluation Questions and Data Collection and Analysis[24]

Evaluation questions guide the evaluation design and inform the data management techniques employed.

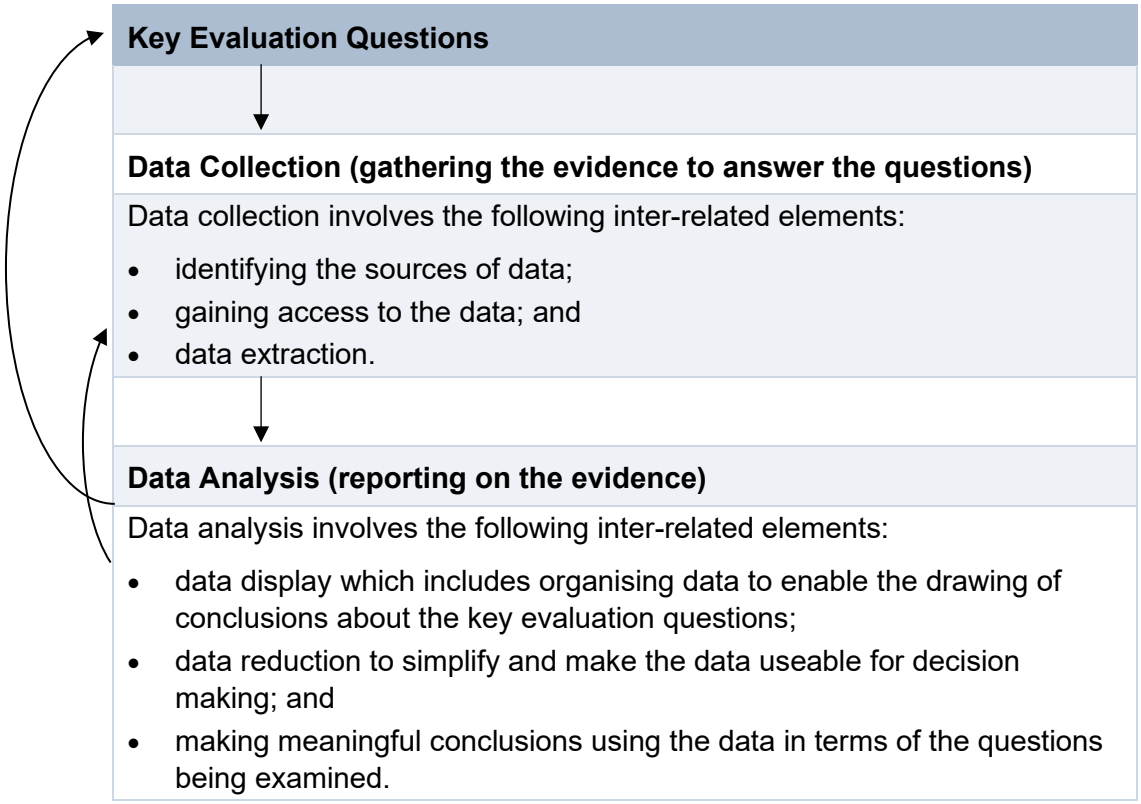
The collection of data (assembly of evidence) relevant to each evaluation question and the analysis of this data (analysis of evidence) form the basis of data management in evaluation.

There needs to be a balance between the rigour in which data is collected and analysed (to answer the key evaluation questions) and the resources that are available to complete the evaluation. It is important to document the limitations of the findings and conclusions.

Figure 7 provides an overview of the link between evaluation questions and data collection and analysis to assess whether results have been achieved.

[24] Owen, J 2006, *Program Evaluation Forms and Approaches*. 3rd edn, Allen and Unwin, Singapore.

Figure 7: Link between Key Evaluation Questions and Data Collection and Analysis in Evaluation[25]



Data Collection

A data collection plan should be embedded into program design and implementation during the program's development.

No single method of data collection is likely to be suitable to answer all evaluation questions. For most evaluations, a combination of both qualitative and quantitative information is used to evaluate a program.

Choosing the right data to collect is key to getting valid information that stakeholders will perceive as useful for decision making.

The strength of evaluation findings is usually found in the bringing together of data from different sources. Data collection methods should be designed and scaled to the evaluation in accordance with the program's size, risk and significance. As there is a large amount of

[25] Adapted from Owen, J 2006, *Program Evaluation Forms and Approaches*. 3rd edn, Allen and Unwin, Singapore, p. 99.

'secondary' data collected through a variety of sources[26], always look to use this type of data where possible.

Table 8 illustrates data sources and collection methods. For each data source, examples of data collection methods are provided.

Table 8: Data Sources and Collection Methods[27]

Sources of data	Data collection method
Directly from individuals identified as sources of information	**Self-reports** • Diaries or anecdotal records. • Checklists or inventories. • Rating scales and semantic differentials. • Written responses (survey). **Personal products** • Tests. • Samples of work. • Interviews.
Compiled by an independent observer	• Written accounts. • Observation forms such as observation schedules, rating scales, checklists and inventories. • Oral responses either singularly or by a group (focus group).
Compiled by use of mechanical or electronic devices	• Audiotape. • Videotape. • Time lapse and still photography. • Internet responses.
Through unobtrusive techniques	• Role playing ('testing').
Existing records	These include published or unpublished documents. • Productivity Commission's Report on Government Services, Agency annual reports. • *Australian Bureau of Statistics*. • *Australian Institute for Health and Welfare*. • *Australian Social Science Data Archive* which houses statistics from the Australian Bureau of Statistics as well as data sets from public sector

[26] Sources may include Report on Government Services (released by the Productivity Commission), Australian Bureau of Statistics or information gathered by the Telethon Institute for Child Health Research.

[27] Adapted from: Owen, J 2006, *Program Evaluation Forms and Approaches*. 3rd edn, Allen and Unwin, Singapore, p.100-101.

	agencies, university based researchers and the private sector. • *World Development Indicators* which contain statistical data from the World Bank for over 331 development indicators and time series data from 1960-2008 for over 209 countries and country groups. Data includes social, economic, financial, natural resources and environmental indicators. • The Developmental Pathways Project for *Telethon Institute for Child Health Research* has data over many years which may be accessed on a needs basis with permission.

The establishment and collection of baseline data before program implementation is recommended to enable meaningful comparisons where possible.

If baseline data cannot be obtained, the use of benchmarking against similar programs or best practice research can be used as a baseline guide for the program evaluation.

Table 9 illustrates some data collection principles to ensure that the data collected will be reliable, useable and useful for decision making.

Table 9: Data Collection Principles

Key principles for data collection include:
• Data collection must be reliable and consistent across different time points. • All individual sets of data should be collected in the same way, using the same set of instruments (for example, a questionnaire) at each time point. • Where one on one interviews, focus groups or broader surveys are used, an effort should be made to maintain consistency among respondents (for example, using the same people or target group). • All interviewees and focus group participants should be assured that all contributions will be treated confidentially and will not be attributed to them.

Secondary Data

Secondary data refers to existing data collected for another purpose. Sources of secondary data include other studies previously conducted, administrative records as well as (though not limited to) official statistical sources.

Collection of new, primary data for an evaluation may not always be practical. The use of existing data may provide an alternative means of answering the evaluation questions.

The resources provided by the Public Sector data collection agencies (such as Report on Government Services and *Australian Institute for Health and Welfare)* and public archives (such as *Australian Bureau of Statistics and Australian Social Science Data Archive)* make

secondary data easier to access and use when compared to the costs and practical problems associated with gathering primary data. Data may come from the program itself, from other programs within the agency, or from other agencies (including other levels of government).

When using secondary data, relevant information needs to be extracted from those records and tabulated in order to yield the desired information. In addition, secondary data can form a benchmark against which primary data can be compared.

After identifying a secondary data source, it is important that its relevance and quality is determined by the program evaluation working group before using it as part of an evaluation.

Additional information on data collection:

- Questions to Ask before using Secondary Data, *Appendix F*;
- Hatry, H 2006, Performance Measurement Getting Results, 2nd edn, The Urban Institute Press, Washington DC.
- Northwest Centre for Public Health Practice, Data Collection for Program Evaluation.
- Australian Government, Data Collection Methods, Office for Learning and Teaching.

Data Analysis

Data analysis involves sorting the data in different ways to expose or create new insights. The use of a *combination of quantitative and qualitative* data collection and analysis techniques will improve an evaluation by ensuring that the limitations of one data type is balanced by the strengths of another.

Be mindful of the *Pareto Principle* (the 80/20 rule) when examining and analysing data. That is, 80% of the effects are due to 20% of the causes. The value of the Pareto Principle is that it reminds you to focus on the 20% that matters.

The first step to data analysis is to create a database of the range of quantitative and qualitative data collected. Consider the following questions:

- how will responses/data be organised/tabulated?
- are separate tabulations from different locations or groups required?
- what, if any statistical techniques will be used?
- how will the narrative data be analysed?
- who will organise and analyse the information?

Analysing data to summarise findings and look for trends is an important part of every evaluation. Note: In planning an evaluation, this work is often under budgeted for (time and funds) and therefore needs careful thought.

Mixed method data analysis techniques may be used depending on whether qualitative or quantitative data is collected. Several analytical techniques should be used to ensure accurate and reliable findings. This may be useful in multi-site evaluations where evaluation of data across different sites is essential.

Quantitative Analysis

Quantitative analysis deals in numbers. The most useful way of displaying numerical data is through a Histogram or Polygon.

A Histogram is a bar chart for grouped numerical data in which percentages of each group of numerical data are represented as individual vertical bars. The variable is plotted along the horizontal x axis and the frequency or percentage is plotted along the vertical y axis.

A Polygon is a line representation of a histogram. It allows for multiple group comparisons.

Other quantitative analysis techniques include:

- *correlation*;
- *cross tabulations*;
- *data and text mining*;
- *frequency tables*;
- *measures of central tendency*;
- *measures of dispersion*; and
- *time series analysis*.

Qualitative Analysis

Qualitative analysis deals in words and is guided by fewer rules and standardised procedures than quantitative analysis. Qualitative data may be found embedded in information and in less easily reducible forms than quantitative data. For example, a relevant piece of qualitative data may be found interspersed within portions of an interview script, multiple excerpts from field notes, or in a comment or cluster of comments from a focus group.

It may be useful tallying responses into categories and then presenting the frequency or percentage of each category in tables and charts.

Summary tables and bar or pie charts may be used to facilitate analysis using categorical data. In addition, a *Pareto diagram/chart* can identify situations in which the Pareto Principle occurs[28]. Using a Pareto diagram/chart, categorised responses are plotted in descending

[28] Levine, D. M et al. 2014, *Statistics for Managers – Using Microsoft Excel*, 7th edn, Pearson Hall, New Jersey, (Levine's website is: *http://www.pearsonhighered.com/levine/*)

order, according to their frequencies, and are combined with a cumulative percentage line on the same chart.

Other qualitative analysis techniques include:

- *Content Analysis*;
- *Thematic Coding*; and
- *Most Significant Change*.

The following questions may help to guide the data analysis:

- what patterns and common themes are emerging? How do these patterns (or lack thereof) help to answer the key evaluation questions?
- are there any deviations from these patterns? If yes, are there any factors that might explain these atypical responses?
- what interesting stories emerge from the responses? How can these stories help to illuminate the broader evaluation question(s)?
- do any of these patterns or findings suggest that additional data may need to be collected? Do any of the evaluation questions need to be revised?
- do the patterns that emerge corroborate the findings of any corresponding qualitative analyses that have been conducted? If not, what might explain these discrepancies[29]?

In summary, when analysing data:

- compare multiple sources of evidence where possible;
- for multi-site evaluations, analyse information within each case for themes and then across all cases for themes that are either the same or different;
- describe how data collection and analysis is different to original expectations and hypotheses; and
- analyse all information collected to develop a picture of what is happening and why[30]?

Additional information on quantitative and qualitative data analysis: *Australian Bureau of Statistics.*

[29] *National Science Foundation 1997,* Analysing Qualitative Data, *USA.*

[30] Wholey, J et al. 2010, *Handbook of Practical Program Evaluation*, 3rd edn, Jossey-Bass, San Francisco, USA, p. 176-177.

Stage 4: Report and Recommendations

Use Findings to Inform Decisions

The value of an evaluation lies in how it is used to inform decisions. Effective evaluation supports action. Useful evaluation reporting clarifies options, identifies program strengths and weaknesses, and provides information on program improvements and key contextual factors affecting the program. It is important to ensure that the findings are used to achieve the purpose for which the evaluation is intended, and to drive change.

For programs subject to a Sunset Clause, for example, the evaluation should inform Government whether the program:

- is achieving the intended results;
- should continue and/or be improved;
- may be better delivered by an alternative service provider; or
- should cease.

Tips to promote the use of findings to inform decisions include:

- communicate the findings, particularly to key stakeholders and decision-making bodies throughout the evaluation process, not just at the end when the evaluation report is produced;
- align the reporting and dissemination of findings with decision making cycles (that is, prior to Budget, Mid-Year Review, or a Sunset Clause cessation date);
- link the findings to the agency's and Government's strategic outcomes/goals;
- present findings in an understandable format to stakeholders. For example, use a variety of mediums to present the findings such as graphs and tables etc;
- use the results to present an argument – including models or alternative proposals; and
- develop an action plan based on the evaluation's results to implement any required changes.

Evaluation Reporting

How results of an evaluation will be reported depends on the purpose of the report. Different communication formats may be required for different stakeholders[31]. For example, is the evaluation report to be used as a basis for repeating the program somewhere else? Is it to justify ongoing funding? Is it to demonstrate program performance and the achievement of results?

For programs subject to a Sunset Clause, the evaluation report will be used to inform the Government's decision whether to cease, continue or to redesign the program past its set cessation date.

Evaluation findings may also be integrated into other reporting formats such as interim progress reports or extracts included in agency Annual reports.

Regardless of the audience for the evaluation report, there are general principles common to all reports. Table 10 describes the characteristics of good evaluation reporting.

Table 10: Characteristics of a Good Evaluation Report
Results are reported in an accurate and unbiased manner
• Assumptions and value judgements are made explicit. • Data is presented in a comprehensive, rather than selective way. • Avoid over generalising the results. Ensure the results specify to whom the results apply and the likely timeframe for which the results hold true. • Avoid mistaking correlation of data for causality[32] when there is not enough evidence to draw that conclusion.
Report is user friendly
• Be concise and use plain English with little jargon. • Present quantitative results with appropriate contextual statements to aid interpretation. • Break up graphs and tables of numerical data with qualitative feedback that illustrate the points that the data is indicating.
Report is produced in a timely manner
• Provide information in a timely manner, useful for decision makers.
The report considers the ethical and political sensitivities and risks attached to the evaluation
• Write reports that are true and accurate but be mindful of the sensitivities of both the community and the key stakeholders involved.

[31] International Centre for Alcohol Policies, Evaluation Toolkit, Reporting and Dissemination.

[32] **Correlation** is a statistical measure that indicates the extent to which two or more variables (data sets) increase or decrease. A positive or high correlation indicates the extent to which data sets increase or decrease. A negative or low correlation indicates the extent to which one data set increases as the other decreases. **Causality** on the other hand is the capacity of one variable (data set) to influence another. Available from: http://whatis.techtarget.com/definition/correlation and causation.

- Report is to provide clear guidance on the reliability and scope of results and how they should be interpreted.

Additional information for creating an evaluation report *Appendix G.*

Communicate the Findings

Communication is critical to a successful evaluation[33]. While it is important to describe and explain results of activities that are completed, or close to completion, the real impact is in providing recommendations and *justifying the conclusions* so they can be used to drive change and improve the performance of the program over time.

Recommendations are based on the findings of the report about changes in program activities that are likely to bring about improved program effectiveness.

The manner in which recommendations are created and presented to key stakeholders will affect their usefulness for informing decisions. Table 11 provides guidelines for developing useful recommendations.

Table 11: Guide to Developing Recommendations

Guidelines for developing useful recommendations
• Summarise the three to five main points arising from the evaluation that are critical for key stakeholders. Provide recommendations that follow from these findings and develop a plan to ensure they are implemented. • Key stakeholders should be involved in the development of recommendations as much as possible. • Recommendations should be understandable, directed to appropriate persons or groups, appropriate to the context, feasible and practical. • There should be differences in recommendations. Some recommendations will require greater effort and encouragement to adopt than others. • Link recommendations to the evidence where possible.

To promote action, it is essential to seek endorsement for the recommendations from key stakeholders (in particular the Steering Committee where relevant) to ensure they align with Government priorities and the outcomes and goals of the agency(s).

[33] Treasury Board of Canada Secretariat 2009, *Policy on Evaluation*. Available from: http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=15024.

Stage 5: Implementation Plan

Monitor the Program

In the case of programs subject to a Sunset Clause, extension of the program past the initial program cessation date will be dependent on evidence of the program's results and a new implementation and evaluation plan to ensure program implementation continues to achieve the desired results.

To ensure changes to a program (as a consequence of the evaluation) are having the desired effect, the program should be monitored (if still active) and an implementation plan developed to progress the decisions that have been made following the evaluation.

An Implementation Plan is a guide for further developing the program. It should include a timeline of actions or activities as a result of the evaluation.

An Implementation Plan:

- provides a process for thinking through critical components of the program;
- allows for anticipation of program challenges and details the critical steps in advance;
- provides a common understanding among stakeholders, particularly the staff implementing the program;
- clearly articulates the goal of the program; and
- encourages forward and proactive thinking.

Another evaluation may be necessary (at a later date) to ensure the program continues to be effective, efficient and is appropriate to meet the need for which it was intended to address.

In addition to ongoing monitoring of the program, an assessment of the evaluation is beneficial. By critically analysing the effectiveness and appropriateness of completed evaluations, evaluations may be improved which will, in turn, contribute to better programs.

Questions to consider when evaluating the evaluation technique include:

- was it the right type of evaluation?
- were the objectives and scope of the evaluation reasonable?
- was the stated purpose of the evaluation too broad?
- were stakeholders adequately engaged?

- was the right evaluator chosen?
- were resources adequately and efficiently used?
- did the level of evaluation reflect the risk involved with the policy or program? Were risks avoided?
- was the data reliable and interpreted fairly?

References and further information

9. Argyrous, G (editor) 2009, *A practical guide: Evidence for policy and decision –making*, UNSW Press Book, Australia.

10. Australian Capital Territory Government 2010, *Evaluation Policy and Guidelines*, Chief Minister's Department, Canberra. Available from: *http://www.cmd.act.gov.au/__data/assets/pdf_file/0004/175432/ACT-Evaluation-Policy-Guidelines.pdf*.

11. Australian Government, *Evaluation Resources*, Office for Learning and Teaching. Available from: http://www.olt.gov.au/evaluation/key-evaluation-questions.

12. Bouvard, F, Carsouw, R and Labaye, E et al. 2011, *Better for Less: Improving public sector performance on a tight budget*, McKinsey and Company Discussion Paper.

13. Cummings, R 2006, 'What if': The Counterfactual in Program Evaluation, *Evaluation Journal of Australasia*, vol. 6 (new series), no. 2.

14. Victorian Government, *Evaluation Toolbox*, Community Sustainability Engagement, Victoria. Available from: http://evaluationtoolbox.net.au/index.php?option=com_content&view=article&id=1&Itemid=162.

15. International Centre for Alcohol Policies 1995, *Evaluation Toolkit, Undertaking Evaluation, Developing a Monitoring and Evaluation Plan* Policy Tools. Available from: http://www.icap.org/PolicyTools/Toolkits/EvaluationToolkit/tabid/437/Default.aspx.

16. Khan, B and Kael Consulting 2008, *Excerpts from Review of Evaluation Frameworks*, Saskatchewan Ministry of Education, Canada.

17. MEERA, My Environmental Education Evaluation Resource Assistant 2004, *Evaluation: What is it and why do it?* University of Michigan, USA. Available from: http://meera.snre.umich.edu/plan-an-evaluation/step-6-analyze-data.

18. New South Wales Government 2013, *Evaluation Framework*, Department of Premier and Cabinet, NSW.

19. New South Wales Government 2013, *Evaluation Toolkit*, Department of Premier and Cabinet. Available from: http://www.dpc.nsw.gov.au/programs_and_services/policy_makers_toolkit/evaluation_toolkit

20. New South Wales Government 2011, *Stakeholder needs template*, Office of Environment and Heritage. Available from: *http://www.environment.nsw.gov.au/resources/4cmas/0971stakeholderneeds.doc*.

21. Western Australian Government 2004, *Outcome Based Management: Guidelines for use in the Western Australian Public Sector*, Department of Treasury. Available from: http://www.treasury.wa.gov.au/cms/uploadedFiles/_Treasury/Publications/Outcome_Based_Management.pdf.

22. Owen, JM 2007, *Program Evaluation: Forms and Approaches*, 3rd edn, The Guildford Press, New York.

23. Queensland Government 2011, *Developing an Evaluation Framework*, Queensland. Available from: *http://www.qld.gov.au/web/community-engagement/guides-factsheets/evaluating/evaluation-framework.html*.

24. Patton, M 2002, *Utilization-Focused Evaluation (U-FE) Checklist*, University of Michigan. Available from: http://www.wmich.edu/evalctr/archive_checklists/ufe.pdf.

25. Victorian Government 2011, Guide to Evaluation, *How to plan and conduct effective evaluation for policy and programs*, Department of Treasury and Finance.

26. Sanderson, I 2002, *Evaluation, Policy Learning and Evidence-Based Policy Making*, Blackwell Publishers.

27. South Australian Government 2014, Guidelines for the Evaluation of Public Sector Initiatives, Part A and B. Department of Treasury and Finance.
Part A available from:
http://www.treasury.sa.gov.au/__data/assets/pdf_file/0004/1768/ti17-guidelines-part-a.pdf.
Part B available from:
http://www.treasury.sa.gov.au/__data/assets/pdf_file/0009/2898/ti17-guidelines-part-b.pdf

28. Staley, L 2008, *Evidence-based Policy and Public Sector Innovation*, Institute of Public Affairs.

29. Svendsen, A 1998, *The Stakeholder Strategy, Profiting from Collaborative Business Relationships*, Berrett-Koehler Publishers, USA

30. Taylor-Powell, E, Steele, S and Douglah, M et al. 1996, *Planning a Program Evaluation*, University of Wisconsin-Extension. Available from: http://learningstore.uwex.edu/Assets/pdfs/G3658-01.pdf.

31. The World Bank 2004, *Monitoring and Evaluation. Some tools, methods and approaches*, Washington DC, USA. Available from: http://lnweb90.worldbank.org/oed/oeddoclib.nsf/24cc3bb1f94ae11c85256808006a0046/a5efbb5d776b67d285256b1e0079c9a3/$FILE/MandE_tools_methods_approaches.pdf.

Appendix A: Glossary of Evaluation Terms

Evaluation Term	Definition
Activities	The tasks that are required to be done in order to achieve program outputs.
Appropriateness	Extent to which a program is, or remains, in accordance with the original government decision.
Benchmark	Standards of excellence or achievement against which measurement can be made.
Bias	In statistics, bias describes the extent to which a measurement or sample underestimates or overestimates the true value.
Effectiveness	The extent to which a program's intended results were achieved.
Efficiency	A measure of how economically resources/inputs (funds, expertise and time etc.) are converted into results.
Equity	The quality of being fair and impartial.
Evaluation	The systematic collection and analysis of information to enable judgements about a program's effectiveness, appropriateness and efficiency.
Impact	Looks beyond the immediate results of an initiative and identifies longer-term effects including unintended or unanticipated consequences.
Inputs	The resources (number of employees and/or funds) expended on the policy/program.
Objective	A specific and measurable result that can be reached to accomplish a particular goal.
Outcomes	The results, impacts or accomplishments of the program. It is important to capture both intended and unintended results. Planned outcomes are also known as 'SMART' results for the purposes of programs with Sunset Clauses.
Outputs	The product or service delivered.
Policy	A statement of principle that articulates, and aligns with legislative, regulatory or organisational requirements.
Process	How inputs are translated into outputs during service delivery.
Program	A group of related activities (may be called a program, project, policy, intervention, initiative, strategy or service) undertaken by or for

	Government that intends to have a specific impact (that is, Government is choosing to do something to achieve a result).
Relevance	The extent to which the program's objectives meet the target group's needs or priorities.
Result	An observable measure of achievement, performance or change. It provides evidence of activities, success or otherwise.
'SMART' result	A describable and measurable change that is derived from a cause-and-effect relationship. In this guide, 'SMART' results are the same as outcomes and are defined as Specific, Measurable, Attainable, Relevant and Time-bound.
Sunset Clause	From 1 January 2014, all new programs (or the extension of existing programs) that impact the State's net operating balance by $5 million or more in any one year will be subject to a Sunset Clause. Under a Sunset Clause, the initial submission to Government will include an evaluation plan. This will specify the intended program results and the resources allocated for a finite period. The continuation of a program with a Sunset Clause is subject to Government consideration informed by a detailed evaluation.

Appendix B: Program intent and 'SMART' results

The overarching program goal should describe the program's intended purpose or expected results. Results are the achieved outcomes, observed characteristics or consequences of the program. Results should align with the overarching intent (or objective) of the program and measure program effectiveness, cost effectiveness, appropriateness and efficiency.

A 'SMART' result should meet the following criteria:

'SMART' result criteria	
Criteria	**Description**
Specific	Clear and well defined.
Measurable	The need for concrete criteria for measuring progress and to know when it has been achieved.
Attainable	Is there a realistic path to achievement? Neither out of reach nor below standard performance.
Relevant	Choosing results that matter within the availability of resources, knowledge and time. The results that will drive the program forward.
Time-bound	Reasonable timeframe to achieve the goal. A time-bound result is intended to establish a sense of urgency. For example: Can data be collected with the frequency that aligns with the required reporting timelines?

Additional factors to consider when designing a 'SMART' result:

A 'SMART' result should be:	
Comparable	Does the measure allow for comparisons over time and across similar programs? A baseline set of data for future comparison will need to exist or be collected.
Administratively simple and cost effective	Do the benefits from collecting the data outweigh the administrative burden and cost of the data collection?
Accurate	What data will be used? Will it be of a sufficient quality to confidently draw conclusions?
Attributable	Is the result being measured clearly attributable to the program?

Hypothetical examples of 'SMART' results
Interim Result (After one year) By the end of the first year of the primary school Breakfast program, student absenteeism will fall for those who participate in the program (at least three days a week) by 20% compared to the year before the program was introduced. There will be a 10% improvement in school discipline incidences requiring the Principal's attention.
Interim Result (After two years) After two years school results of Western Australian students who regularly attend the after-school tutoring program (at least 3 days a week) will indicate a 20% improvement in academic achievement in the areas of literacy and numeracy compared to those who have not participated in the program.
Final Result (After three years) At the end of the third year of the Whooping Cough Vaccination Program for all Year 7 students, Western Australian hospital statistics will show a decrease of presentations for confirmed whooping cough by 80% in the 12-15 year age group.

Appendix C: Common Evaluation and Data Collection Types[34]

Evaluation and data collection types	Example uses	Limitations	Benefits
Literature search **Best practice models**	• Identify what's happening in other jurisdictions, new and best practice ideas. • Enhance understanding of causes of problems and identify strategies and options.	• Causal relationships can be difficult to test. • Often needs supplementary information.	• Useful for all evaluations. • Inexpensive way to collect information. • Useful for program planning as it may identify service delivery options, community needs and improve and inform policy.
Comparative analysis including benchmarking and performance measurements	• Measures program performance relative to comparable performance standards. • Identify best practice standards and how to adopt these practices to improve the program. Possible Sources:	• May be some degree of bias in the information due to data integrity issues. • Requires base data information. • Requires comparable performance measures. • May need supplementary information and analysis. • Requires expert knowledge and time to identify relevant measures and to analyse results.	• Useful for all evaluations. • Used to investigate service delivery. • Supports budget priorities. • Improves accountability and transparency. • Drives organisational learning and development.

[34] Adapted from *Guide to Evaluation* 2011, Department of Treasury and Finance, Victoria, p. 9-11.

	• Report on Government Services (released by the Productivity Commission); • *Commonwealth Grants Commission*; • *Australian Bureau of Statistics*; and • *Australian Institute of Health and Welfare*.		
Pilot and Case studies	• Identify what works and what doesn't to guide future program development and implementation.	• Difficult to apply learning from specific pilot or case studies. • Pilot studies are costly and time consuming. Case studies are generally easier to do as they rely on historical information.	• Useful for all evaluations. • Drives organisational learning and improvement, make adjustments and improvements, share and transfer learning.
Performance reports such as: • **surveys;** • **internal and external reports;** • **compliance reports; and** • **management reports.**	• Measures progress achieved against initial performance targets.	• Causal relationships can be difficult to test. • May need supplementary information and analysis. • Can require some expertise to interpret.	• Useful to provide a starting point and background for all evaluations. • Input to cost-benefit and net impact analysis. • Improves organisational accountability. and transparency as well as organisational learning and development.

Overarching Evaluation Questions (Formative, Process and Summative)	• Set of evaluation questions to enable the evaluator to make judgements. For example: • how relevant is the program in meeting the Government's objectives? • is there an ongoing need? • how efficient are the program activities (value for money)?	• Can be subjective. • Requires base data information. • May require benchmark data.	• Useful for all evaluations. • Helps to set baseline data requirements.
Statistical data collections	• Identify quantitative, qualitative, economic, social and environmental impacts of a policy or program. • Inform the generation and testing of alternative models and options.	• May require supplementary information and analysis. • Not all results can be quantified. • Requires expert data analysis. • Availability and quality of data may vary.	• Useful as an input to evaluations that measure cost-benefit or the net impact of policies or programs. • Improves and informs policy decisions. • Supports and informs budget priorities.
Cost-Benefit Analysis (Formative, Summative)	• Identifies and quantifies program costs and benefits. • Quantifies inputs and outputs. • Objective way to compare and rank alternative program options.	• Focus is on costs of a program – some costs and benefits may not be able to be measured in dollar terms. • Relies on predictive data generated by other methods. • Requires expert statistical modelling and use of cash flow techniques such as net present value and discounted cash flow.	• Support budget priorities. • Identify cost effectiveness and efficiencies. • Guide resource allocation decisions. • Forces disciplined consideration of choices, including status quo option). • Makes hidden costs and benefits explicit.

Net impact evaluation or outcome based evaluation (Summative)	• Identify what would happen if the program had not occurred. • Identify the extent to which the program or policy has achieved its desired results once it has been implemented. • Review unintended consequences and impacts of programs.	• Time and resource intensive. • Can be difficult to obtain baseline data for comparisons. • Requires specialist expertise, knowledge or input such as economic and social research skills.	• To improve and inform policy by providing evidence to develop alternatives and future policy options. • Improves accountability and budget priorities by identifying potential improvements.
Cost Effectiveness Analysis (Summative)	• Measures/compares results to costs. • Measures outputs and results in both quantitative and qualitative terms. • Compares and ranks programs in terms of their costs for reaching given results, to reduce costs of a program.	• Time and resource intensive. • Requires specialist economic or social research expertise and program knowledge to objectively assess effectiveness. • Need clear measures for results.	• Useful to compare the costs and results. • Usually through ratio – denominator being a gain in a variable (for example, improvement in life expectancy) and the numerator being the cost associated with the gain.

Appendix D: Framework for developing an Evaluation Plan

1. Program Intent and Rationale
What program or sub-program is being evaluated? This should include a brief overview of the program objectives, program logic, and history of program's development (that is, who was involved in its development and how long has it been in existence?).
2. Purpose of the Evaluation
Why is the evaluation being conducted? The evaluation may be to develop a program, improve the delivery of a program, to test whether a program has been effective, appropriate and has been delivered efficiently.
3. Key Stakeholders including the Primary Audience
Identifying and understanding the key program stakeholders is critical to the success of every program evaluation. Who will be interested in the information that will be obtained from the evaluation? The stakeholders may be classed as external or internal stakeholders. External stakeholders for an evaluation include the program's target audience (that is, the users or beneficiaries of the program), partners involved in developing and delivering the program, ethics committees and for large programs the general public. Internal stakeholders include the Government, Agency, Board of Directors or Governance bodies and potential data custodians. Stakeholder analysis is a tool used to identify the key people associated with a program and program evaluation. Who will receive and use the information? The primary audience is the person or group that is most likely to use the information produced by the evaluation, whether it be evidence, conclusions, judgements or recommendations. It is important to distinguish the primary audience from clients and interested stakeholders, and may not be the same as the group that initiates and provides the resources for the evaluation.

4. Key Evaluation Questions

Key evaluation questions are questions the users of the evaluation want to know. For example:

- to what extent does the vaccination program reduce hospital admissions?
- is there a correlation between a targeted Year 2 literacy program and improved overall academic results in Year 3?
- will the juvenile justice program reduce recidivism in the five-year period after program completion?

Asking the right questions requires the selection of the most important aspects of the program to be examined. Carefully consider what you really want to know from the evaluation to keep the scope manageable. Avoid asking too many questions (particularly with time or resource constraints) or questions which are not amenable to evaluation.

5. 'SMART' Results

S = Specific

Use specific rather than generalised language.

M = Measurable

Be clear about what will be changed and by how much. Setting this clearly at the start makes it easier to evaluate. An example of a measurable result: Within six weeks of running a Year 2 early intervention literacy program in Western Australia's primary schools, students' spelling errors will have decreased by 20% compared to the start of the program.

A = Achievable

Be realistic about what the program can achieve in terms of the scale/scope of what is being done, the time and resources available.

R = Relevant

The outcomes need to relate to and be relevant to the results.

T = Time-bound

Be specific about the timeframe in which the program's activities will have an impact on the expected changes/results. For example: At the end of the first year of the Year 11 and 12 teacher assisted after school study program, exam results for students who attended the program 3 or more days a week will indicate a 10% increase in academic achievement compared to those students who did not participate in the program.

6. Timeline, Budget and Evaluation Resources

Timeline for Evaluation

A *timeline* for an evaluation demonstrates the different evaluation activities which will take place during the five stages of a program evaluation, as well as indicates the key reporting deadlines.

A timeline is an essential element of an evaluation plan and should be negotiated with the key stakeholder groups. It enables the program and evaluation staff to schedule the major activities required to complete the evaluation on time and within budget and to track progress to ensure there is a smooth flow of activities. A timeline should include a high-level plan outlining key milestones. This includes when the findings are needed to report to Cabinet, a Minister, or Director General. Note: The budget and other available resources will impact on what you can reasonably expect to achieve in the timeframe.

Budget and Evaluation Resources

Specify the budget for evaluation (generally, the rule of thumb is 1-5% of the total program cost).

In practice, most evaluations are completed with resource and time constraints. The availability of evaluation resources will have implications for the design and scale of the evaluation, as well as what can be realistically achieved by the evaluation.

The evaluator will need time to plan, set up appropriate data management systems, and generally be responsible for all aspects of the evaluation. For programs other than those subject to a Sunset Clause, consider the need for an oversight body, such as a working group or steering committee, and the membership of that body. It is always a good idea to use an evaluator who is independent from the management of the program.

7. Privacy and Ethics

How will client/commercial privacy be safeguarded (if relevant)?

What ethical issues need to be considered and addressed? Does the project need ethics clearance from a relevant body?

8. Baseline Data and Methodology

There are a range of methodologies, data collection and analysis techniques. Seek advice from specialist data, research or evaluation areas in your agency or from the Program Evaluation Unit at Treasury. Ideally, the desired results, associated questions used to verify these results and identified data sources will be considered during program design.

Questions to consider: What is the baseline for which program results will be compared? Describe the ‘SMART’ results? How will I know they have been achieved? Can data be collected from existing data sets? What methodology will be used during the evaluation (for example, survey, case study, randomised control trial)?

9. Communication of Findings

How will information about the evaluation findings be communicated to decision makers, stakeholders and the community? What kinds of information will be included (for example, findings, conclusions, judgments, recommendations)?

Effective dissemination of findings increases the likelihood that the evaluation will inform decision making. The findings should be presented in a way that is understandable to the primary audience. Be explicit about intentions to publish the final report.

Appendix E: Program Evaluation Plan

Program Information
1. **Name of Program**
2. **Program Location(s)**
3. **Lead Agency**
4. **Other Agencies** (delivering the program)
5. **Program Intent and Rationale**
6. **Purpose of Evaluation**
7. **Key Stakeholders** (for the program evaluation)
8. **Program Logic Map.** (attach to the evaluation plan)
9. **Key Evaluation Questions.** (to measure Effectiveness, Appropriateness and Efficiency)
10. **Program Results** ('SMART' results are developed to answer the evaluation questions.)

Interim Results (Year 1)	**Interim Results (Year 2)**	**Final Results (Year 3)**
11. Risk Identification and Analysis		
12. Evaluation Timeline		
13. Who will Evaluate the Program (external or internal service provider)? (rationale for choice of provider)		
14. Baseline Data and Methodology (What is the baseline data? What data collection methods will be used to measure effectiveness, appropriateness and efficiency of 'SMART' results? For example, survey/questionnaire, charts, maps, case study, pre and post program statistics, existing data from ROGS or annual reports.)		
15. Communication Plan (to communicate findings)		

Program Funding				
Is the Program Subject to a Sunset Clause? If so, what is the funding cessation date? Sunset Clause Yes ☐ No ☐ Funding cessation date:				
Is Program Delivery Contracted to an External Provider? Yes ☐ No ☐ If yes, provider name:				
Period of Funding: e.g. 1 July 2014 – 30 June 2017				
Contractual Arrangements				
Service Agreement ☐	Funding Contract ☐	Grant Agreement ☐	MOU ☐	Other ☐
If other, provide a description.				
How many Full Time Equivalent (FTE) people are employed to deliver the program?				
Total Program Funding: $'000		**Period of Funding:** e.g. 1 July 2014 – 30 June 2017		
Annual Program Funding Breakdown				
Year	**State Govt. Funding Totals $'000**	**State Govt. Funding Sources**	**Other Funding Totals $'000**	**Other Funding Sources**
Example. 2012-13	$5,000 $2,000	Education Dept. Budget Health Dept. Budget	$2,000 $500	Commonwealth Private Investment
Evaluation Resource Requirements (FTEs and $) **Note: Include cost of evaluation in total funding costs for program.**				
Evaluation Costs Consider quarantining an evaluation budget.	**Total FTE**	**FTE total $'000**	**Other costs $'000**	**Total cost of Evaluation $'000**
		$	$	$

Appendix F: Questions to ask before using Secondary Data

1. Who collected the data?
2. What was the original purpose for the data collected?
3. Are there any geographic or demographic limitations to the data?
4. When was the data collected?
5. How was the data collected?
6. How were the variables defined?
7. For longitudinal data, have the methods of collection and variable definitions changed over time?
8. In what form is the data available?
9. What is the size of the sample?
10. Are the data collection instruments available?
11. What restrictions, including ethical concerns, are there in the use of the data?
12. Is the data in raw form or summarized into higher aggregates?
13. Was the data collection process itself based on previous research?
14. What are the costs of purchasing the data[35]?

[35] Argyrous, G 2009, *Evidence for Policy and Decision - Making*, UNSW Press, Australia, p.173.

Appendix G: Evaluation Report Checklist

This evaluation report checklist is intended as a guide. Sometimes different versions of the evaluation report will be required for different stakeholders. Before writing a report it is important to understand the depth of report required and who will read it[36].

Checklist for Evaluation Reports		
Program:		
Agency:		
Evaluator:		
1. The Report Structure		
1.0	The Report is well structured, logical, clear and complete.	☐
1.1	Report is logically structured with clarity and coherence (e.g. background and objectives are presented before findings, and findings are presented before conclusions and recommendations).	☐
1.2	The title page and opening pages provide key basic information:	
	• name of the evaluation.	☐
	• timeframe of the evaluation and date of the report.	☐
	• name of evaluator(s).	☐
	• name of the agency commissioning the evaluation.	☐
	• table of contents which also lists tables, graphs, figures and appendices.	☐
	• list of terminology including acronyms.	☐
1.3	The Executive Summary is a stand-alone section of 2-3 pages that includes:	
	• overview of the evaluation.	☐
	• evaluation objectives and intended audience.	☐
	• evaluation methodology.	☐
	• most important findings and conclusions.	☐
	• main recommendations.	☐
1.4	Appendices increase the credibility of the evaluation report. Appendices may include:	
	• evaluation terms of reference.	☐

[36] Checklist adapted from: United Nations Evaluation Group 2010, Quality Checklist for Evaluation Reports.

	• list of persons interviewed and sites visited.	☐
	• list of documents.	☐
	• further information on the methodology, such as data collection instruments, including details of their reliability and validity.	☐
	• evaluator's justification of team composition.	☐
2. Full description of Program		
2.0	The report presents a clear and full description of the evaluated program.	☐
2.1	The inputs, outputs and results of the program are clearly described based on the Program Logic Map.	☐
2.2	The context of key social, political, economic, demographic, and institutional factors that have a direct bearing on the evaluated program is described.	☐
2.3	The scale and complexity of the evaluated program are clearly described. For example: • the number of components, if more than one, and the size of the population each component is intended to serve, either directly or indirectly. • the geographic context and boundaries (such as the region, and/or landscape and challenges where relevant). • the purpose and goal, and organisation/management of the program and its parts. • total resources from all sources, including human resources and budget(s) including Agency, State and Commonwealth funding.	☐
2.4	The key stakeholders involved in the program, including the implementing agency(s) and partners, other key stakeholders and their roles.	☐
2.5	The report identifies the implementation status of the program, including its phase of implementation and any significant changes (e.g. plans, strategies) that have occurred over time. In addition, explain the implications of those changes for the evaluation.	☐
3. Evaluation Purpose and Scope		
3.0	The evaluation's purpose, objectives and scope are fully explained.	☐
3.1	The purpose of the evaluation is clearly defined, including why the evaluation was needed at that point in time, who needed the information, what information is needed and how the information will be used.	☐
3.2	The report should provide a clear explanation of the evaluation's 'SMART' results and scope including key evaluation questions. It should describe and justify what the evaluation did and did not cover.	☐

4. Evaluation Process		
4.0	The report describes the evaluation process and clearly explains how the evaluation was designed to address the 'SMART' result criteria and answer the evaluation questions.	☐
4.1	The report describes the data collection methods and analysis, the rationale for selecting them, and their limitations. Baseline data and benchmarks are included where relevant.	☐
4.2	The report describes the data sources, the rationale for their selection, and their limitations. It includes a discussion of how a mix of data sources was used to obtain a diversity of perspectives, ensure data accuracy, validity and overcome data limitations.	☐
4.3	The report gives a complete description of the stakeholder consultation process during the evaluation, including the rationale for selecting the particular level of consultation.	☐
5. Findings		
5.0	Findings relate directly to the 'SMART' result criteria.	☐
5.1	Reported findings reflect systematic and appropriate analysis and interpretation of the data.	☐
5.2	Reported findings address the 'SMART' result criteria (such as efficiency, effectiveness, sustainability, impact and relevance) and key questions defined in the evaluation scope.	☐
5.3	Findings are objectively reported based on the evidence.	☐
5.4	Gaps and limitations in the data and/or unanticipated findings are reported and discussed.	☐
5.5	Overall findings are presented with clarity, logic, and coherence.	☐
6. Conclusions		
6.0	Conclusions present reasonable judgments based on findings and substantiated by evidence.	☐
6.1	The conclusions reflect reasonable evaluative judgments relating to key evaluation questions.	☐
6.2	The conclusions provide insights into the identification and/or solutions to important problems or issues.	☐
6.3	Conclusions present strengths and weaknesses of the program being evaluated, based on the evidence presented and taking due account of the views of a variety of stakeholders.	☐

7. Recommendations		
7.0	Recommendations are relevant to the program and the evaluation's purpose, are supported by evidence and conclusions, and were developed with the involvement of relevant stakeholders.	☐
7.1	The report describes the process involved in developing the recommendations including consultation with stakeholders.	☐
7.2	Recommendations are based on evidence and conclusions.	☐
7.3	Recommendations are actionable and reflect an understanding of the agency.	☐
7.4	An implementation plan for the recommendations is included within the report.	☐
7.5	If a program is subject to a Sunset Clause, the recommendations need to incorporate the evidence as to why the program should continue, cease or whether alternative service delivery options be sought.	☐