

NSW GOVERNMENT
EVALUATION FRAMEWORK
AUGUST 2013



Disclaimer

While every reasonable effort has been made to ensure that this document is correct at the time of printing, the State of NSW, its agents and employees, disclaim any and all liability to any person in respect of anything or the consequences of anything done or omitted to be done in reliance or upon the whole or any part of this document

Copyright

In keeping with the NSW Government's commitment to encourage the availability of information, you are welcome to reproduce the material that appears in the NSW Government Evaluation Framework for personal, in-house or non-commercial use without formal permission or charge. All other rights are reserved. If you wish to reproduce, alter, store or transmit material appearing in the NSW Government Evaluation Framework for any other purpose, request for formal permission should be directed to Strategic Initiatives Branch, Department of Premier and Cabinet, GPO Box 5341, Sydney NSW 2001. You are required to acknowledge that the material is provided by the NSW Government Evaluation Framework or the owner of the copyright.

ISBN: 978-0-7313-5499-3

DCC1305

CONTENTS

1. Executive summary	2
2. Introduction.....	3
3. Understanding program evaluation	4
Defining programs	4
Defining evaluation.....	5
Evaluation is an integral part of the program lifecycle.....	6
4. A consistent approach to evaluation.....	8
Being strategic – deciding what to evaluate	8
Good practice principles	9
Scaling evaluation.....	10
Using evaluation – Informing decision making and communicating results	11
5. Building evaluation into the public sector.....	12
Evaluation and performance.....	12
New program proposals	12
Program evaluation plans	13
Cluster evaluation schedules	13
Transparency and accountability	13
Roles and responsibilities.....	14
6. Building evaluation capability	15
Community of practice	15
Partnerships with specialist bodies.....	15
7. Transition	16
Implementing the recommendations of the Commission of Audit	16
Program review	17
A toolkit to support evaluation and review.....	17
Review of the Framework.....	17
8. References	18
Appendix A Glossary of terms.....	19
Appendix B Elements of an Evaluation Plan.....	20
Appendix C Capability target groups and their needs	21

1. EXECUTIVE SUMMARY

The government has committed to a series of reforms that will **increase transparency** of expenditure on programs and provide a **better understanding** of the outcomes programs delivered. This evaluation framework is part of those reforms and complements financial management reforms underway.

As highlighted in the recommendations of the Final Report of the NSW Commission of Audit into Government Expenditure¹, evaluation is a key tool to support **evidence based policy and decision making** in government, to help government **learn and adapt** to changing environments and as a tool for **communicating and sharing** valuable information. When planned, designed and conducted in accordance with good practice standards, evaluation can provide the necessary evidence to improve services and guide better resource allocation decisions.

In NSW we recognise evaluation as an integral part of managing government programs at every stage of the program lifecycle. It requires policy and program managers to think about evaluation from the very beginning of program design, and explains how different types of evaluation can be used to ask and answer different sorts of questions at different times.

To guide consistency in evaluation across the sector, it is important the framework:

- emphasises a strategic approach and outlines program features that should be considered when prioritising evaluation efforts
- describes key principles of good evaluation practice
- illustrates how evaluations may be scaled based on the characteristics of different types of programs
- explains how findings generated by evaluation can be used for learning and better decision making.

Evaluation will become a core component of the way we work to improve the quality of evidence upon which decisions are made. The framework outlines the requirements for program evaluation plans and agency evaluation schedules, mandates transparency and provides advice on how to share and publish evaluation findings, and clarifies the role of clusters and the Strategic Centre.

Building evaluation capability across the sector is crucial to the success of the framework. The framework identifies the needs and options to develop the capability of the four major groups engaged in evaluations – senior executives, evaluation experts, other public servant employees and non-government service providers. Understanding their needs will assist clusters to properly target training and development activities.

The integration of evaluation into existing processes will take time. The framework also sets out transitional arrangements as we develop the systems, capability and capacity to regularly evaluate our programs.

¹ NSW Commission of Audit, Final Report Government Expenditure, May 2012

2. INTRODUCTION

This document provides a framework to guide the consistent and transparent evaluation of government programs to inform decision making on policy directions, program design and implementation.

It has been designed to support the Government's commitment to return quality services through evidence based policy and decision making.

Governments face the ongoing challenge of solving complex problems and providing quality services to the community in an environment of constrained resources. A key part of meeting this challenge is testing new ideas and learning what works and what doesn't. More importantly, it means **using that knowledge** – to improve performance, and share learning. Providing agencies with a mandate to evaluate will assist their efforts to learn more about how they can best develop effective policies and programs.

The framework has been developed in response to the Final Report of the NSW Commission of Audit into Government Expenditure², which was critical of the lack of information and transparency on government expenditure and the degree to which programs are meaningfully evaluated to determine their ongoing relevance, effectiveness and value for money. The Commission made wide-ranging recommendations to improve the provision and use of information about expenditure, including:

- improving program information routinely collected and reported on by agencies
- better scrutiny of new proposals and building evaluations into all approvals
- adherence to program review or termination dates
- mandating program evaluation
- ensuring the independence of major and significant program evaluation through third party review.

² Ibid

3. UNDERSTANDING PROGRAM EVALUATION

It is important to have a shared understanding of what we mean by 'program' and 'evaluation'. This section defines both terms for the purposes of the Framework.

Defining programs

For the purposes of the Framework, 'program' is broadly defined as:

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

'Program' is sometimes used interchangeably with project, service, initiative, or policy.

In practice, programs vary in size and structure. They can be large and significant such as Keep Them Safe, a whole of government action plan to reform the delivery of family and community services. These types of programs are typically comprised of many sub-programs and are delivered by multiple agencies and/or in partnership with non-government organisations. 'Program' may also refer to agency specific programs like Rent It, Keep It, a training package developed by Housing NSW to assist people who are seeking private rental accommodation gain the skills to secure and maintain a tenancy.

Figure 1. Programs at different levels



Regardless of program size, when designed and conducted well, evaluation can yield useful evidence about the effectiveness or otherwise of programs at each level. It can be used to assess entire programs or components of programs (i.e. sub-programs or projects). See Figure 1 (below) for an illustrative example of programs at different levels.

Defining evaluation

Well planned and executed evaluation provides timely evidence of the efficiency and/or effectiveness of programs. This will help strengthen community confidence in the investment of public money.

Evaluation has a key role in improving services as well as guiding better resource allocation. It is an important tool to help us learn and understand what's working, what's not, and why. Evaluation can also shed light on where we can make changes to programs, big and small, that could lead to better outcomes.

For the purposes of this framework, evaluation is defined as:

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

A key component of evaluation involves establishing bases of comparison to enable the interpretation of data. This goes beyond reporting a statistic, and helps us understand whether a situation is better or worse than before, how it might compare to other regions or jurisdictions, or how it compares to another approach. Fundamentally, evaluation is about asking questions of our programs, such as:

- Do they meet the needs of the community?
- Are they achieving their intended outcomes?
Or producing unintended outcomes?
- Have they been implemented as planned?
- Do they provide value for money?
- Should they be continued, expanded, modified, or discontinued?
- Is there a better way to achieve the same result?
Can resources be allocated more efficiently?

Evaluation is part of a spectrum of activities used to assess what we do. These activities do not constitute evaluation and need to be continued to be undertaken as part of the suite of activities to test the value of outcomes. They are:

- Program reviews - typically quicker, more operational assessments of "how we are going," often to inform continuous improvement. Reviews often take place after implementation has started and may be useful when there is insufficient information to conduct an evaluation. They can also be used to set up an evaluation framework for an existing program that previously did not have one.
- Monitoring - a management process to periodically report against planned targets (or KPIs). For the most part, monitoring is not concerned with questions about the purpose, merit, or relevance of programs. In government, monitoring is frequently based on outputs as opposed to outcomes.
- Research - closely related to evaluation, and when asking questions about the effectiveness of a program, may even be the same thing. However, research can also ask different types of questions that may not be related to merit or worth of a program (e.g. questions about population groups).

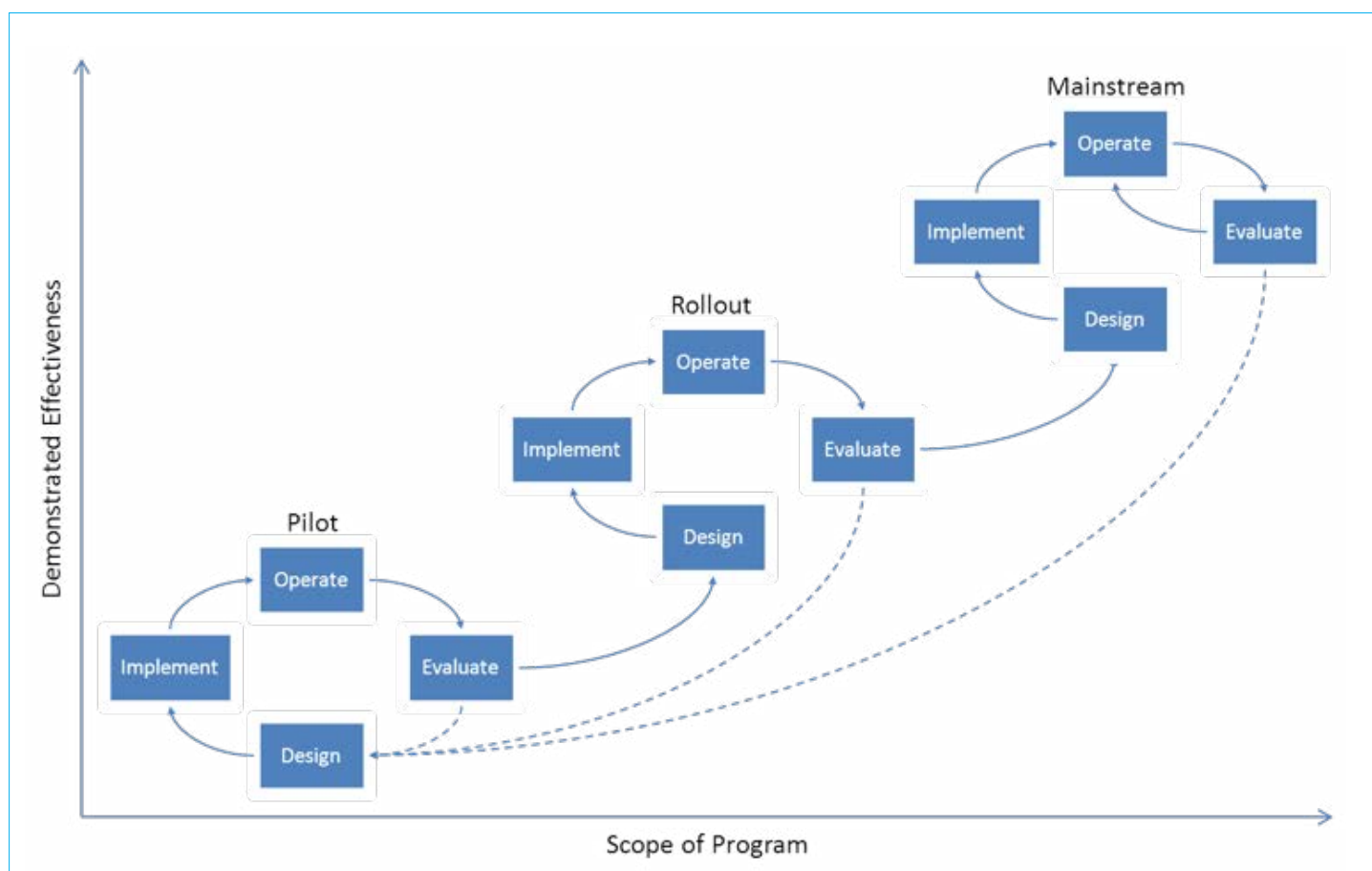
Good Practice Example: Fire & Rescue NSW

Fire & Rescue NSW use a range of activities in addition to evaluation to assess what they do, including internal audit, accreditation, benchmarking, performance reporting, and incident analysis. Reporting on the outcomes of these processes is integrated into their governance framework and findings used to inform operational decisions.

Evaluation is an integral part of the program lifecycle

Evaluation can and should take place across the lifecycle of a program, from design and piloting through to implementation and ongoing mainstream delivery see Figure 2 (below). It has an equally important role to play in testing the impact of new policies and testing whether existing mainstream programs are continuing to deliver outcomes effectively.

Figure 2. Evaluation in the (ideal) program lifecycle



Different types of evaluation provide different information and support different decisions. That's why it's important to plan upfront what questions need to be answered, how they will be answered, and by when.

For the purposes of this framework, we discuss three main types of evaluation, formative evaluation, process evaluation and summative evaluation³:

Table 1. Types of evaluation

TYPE	DESCRIPTION	EXAMPLE QUESTIONS TO BE ANSWERED
<p>Formative</p>	<p>Formative evaluation can provide information on how the program might be developed (in the case of new programs) or improved (both new and existing programs). Examples include:</p> <ul style="list-style-type: none"> • Needs assessment to determine who needs the program, how great the need is, and what might work to meet the need. • Development of a program logic to ensure there is a clear picture of how and why the program will produce the expected outcomes. • Business case to define the program, delivery mechanisms, the target population, and the possible outcomes. • 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. 	<p>What is the problem we're trying to solve?</p> <p>What are the characteristics and needs of the target population?</p> <p>Where is the problem and how big or serious is it?</p> <p>What is the most appropriate thing to do?</p> <p>Is government intervention appropriate?</p>
<p>Process</p>	<p>Process evaluation investigates how the program is delivered, and may consider alternative delivery procedures. It can help to differentiate ineffective programs from failures of implementation.</p> <p>As an ongoing evaluative strategy, it can be used to continually improve programs by informing adjustments to delivery.</p>	<p>How is the program being implemented?</p> <p>Are the activities being delivered as intended?</p> <p>Are participants being reached as intended?</p>
<p>Summative</p>	<p>Summative evaluation generally reports when the program has been running long enough to produce results, although it should be initiated during the program design phase. It may also be known as:</p> <ul style="list-style-type: none"> • Outcome evaluation: determine whether the program caused demonstrable effects on specifically defined target outcomes. • Cost-effectiveness and cost-benefit analysis: address questions of efficiency by standardising outcomes in terms of their dollar value to answer questions of value for money. These types of analyses can also be used in formative stages to compare different options. 	<p>What are the net effects?</p> <p>To what extent can changes be attributed to the program?</p> <p>Is the program the best use of the resources it costs?</p>

³ Adapted from Research Methods Knowledge Base by William M.K. Trochim (2006)

4. A CONSISTENT APPROACH TO EVALUATION

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

This requires common understanding of the principles that underpin good evaluation, the ability to scale evaluation relative to different programs, and an understanding of how findings can best be communicated for learning and better decision making.

Being strategic – deciding what to evaluate

Clusters and their agencies are responsible for planning and managing evaluations of their programs. It is not feasible, cost effective or appropriate to fully evaluate everything, so agencies will need to decide what should be evaluated, at times with advice and support from central agencies. As evaluation is only one of the activities we undertake to assess what we do, agencies should use discretion to determine whether it is best to conduct an evaluation or other review of effectiveness and efficiency.

For some programs, evaluation will be mandatory and will need to meet certain requirements (see Scaling evaluation, below). In these instances, it is likely that central agencies will have a role in advising on, planning or managing the evaluation in partnership with agencies. They may also sit on the steering committee for the evaluation. These programs are likely to be major reforms that are large, complex, and expensive (e.g. Stronger Together 2 or Keep Them Safe). The involvement of central agencies in program evaluations will be negotiated with line agencies.

Clusters are expected to evaluate a significant proportion of their programs. The selection of programs is a matter of judgement and may be influenced by the capacity of each agency or cluster. To encourage consistency and guide decision making around prioritisation, agencies should consider the following:

- **Size of investment** – large scale or resource intensive programs should be evaluated to ensure they deliver intended outcomes and provide the best value for money.
- **Terminating programs** – programs that have been funded on a time-limited basis should be prioritised for evaluation (prior to the termination date) to inform and support future decision making on their continuation.
- **Strategic importance to government** – programs that feature in agency strategic plans, election commitments, *NSW 2021*, or National Partnership Agreements (NPs) should be evaluated, particularly to determine the extent to which they are contributing to government priorities. These evaluations are likely to be of interest to a wider range of stakeholders or may be a contractual obligation (e.g. for programs funded under NPs).
- **Risk** – programs that may pose risks to government, an agency, its clients/stakeholders, or the community should be carefully monitored and evaluated to assess impacts.
- **Innovation** – where there is a level of uncertainty about the design, delivery and results of a program, or where a program with an established evidence base is being applied in a new environment (e.g. different location or target group), evaluation should be designed to inform decisions about future investment and improvement.
- **Cross-sector involvement** – programs that are partly or wholly funded by other jurisdictions, such as the Australian Government, should also be considered, particularly if there are performance payments tied to funding and to determine if the program is worthy of ongoing state investment. Delivery partners should be involved in the design of the program and the evaluation (e.g. other agencies within or across clusters, between jurisdictions, or with non-government organisations).
- **Complexity** – whether the program or its delivery is relatively simple, complicated, or complex should inform decisions about the type and extent of the evaluation. Other considerations may be whether the need for or design of the program is contested by others, and the degree to which any outcomes might be attributed to a single program or combination of programs.

Good practice principles

Key principles that underpin the conduct of good evaluations include:

- **Evaluations should be built into program design**
– planning an evaluation before a program is implemented can strengthen the rationale for the program, improve the design of the program, and increase the power of the evaluation. This includes articulating evaluation questions, developing a program logic and clear outcome measures, and identifying data needs and key stakeholders. Ideally, evaluators should be engaged during the formative stage to ensure that programs are designed in such a way that they can be evaluated.
- **Evaluations should be methodologically rigorous, with appropriate scale and design** – where possible, evaluations should employ best practice methodologies and relevant data to provide clear and accurate information that can be relied on. Since resources and time are required to conduct evaluations, they should be designed and scaled to each program in accordance with the program's size, risk, and significance. An evaluation toolkit will be developed to provide guidance on different methodologies and the quality of evidence produced by those methodologies (see Transition, p. 18).
- **Evaluations should be conducted with the right mix of expertise and independence** – the person or agency conducting the evaluation should be independent from program managers. However, program managers should be involved in the planning of evaluations and input from subject matter experts (including program managers) should be included. Draft evaluation reports should be discussed with program managers prior to finalisation.
- **Evaluations should be timely to support and influence decision making** – planning of evaluations should commence before implementation with the selection of methodologies and collection of baseline data. When planning evaluations, the primary audience and decision making body should be explicitly nominated. Evaluations should conclude before decisions are made, with consideration given to the realistic amount of time needed to conduct them to answer the questions being asked. Summative evaluations should not be undertaken too early, in recognition of the time it can take to accrue sufficient evidence and produce measurable outcomes.
- **Evaluation processes should be transparent and open to scrutiny** – comprehensive information on all aspects of the evaluation should be systematically recorded, including choice of methods, analyses, and conclusions. The extent to which evaluation findings, methods and data are shared within and external to government should be determined during planning. An evaluation report on any program delivering services to the public should be publicly released in a timely manner, except where there is an overriding public interest against disclosure, to foster accountability, knowledge building, and transparency.

Scaling evaluation

The scale of an evaluation should be proportionate to the size or significance of a program. Table 2 (below) outlines the extent to which different types of programs are expected to be evaluated. Decisions will be primarily based on the judgment of agencies. There may also be times where Cabinet or central agencies will request an evaluation.

Table 2. Program characteristics and relative scale of evaluation

TIER	CHARACTERISTICS OF PROGRAM	CHARACTERISTICS OF EVALUATION
1	<ul style="list-style-type: none"> Limited investment of resources Low strategic priority Low risk 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 	<p>Evaluate at Agency CEO discretion, but may be a less formal review process with limited data needs and few resources allocated. Regardless of intent to evaluate, the following program information should be routinely collected:</p> <ul style="list-style-type: none"> Program rationale, objectives Budget Risks Milestones Performance measures (such as KPIs) Responsible Executive and/or Program Manager.
2	<ul style="list-style-type: none"> Moderate investment (relative to agency) Named in agency strategic plan Low to moderate risk Responsibility of single Minister or agency Not recently reviewed (in the case of existing programs) 	<p>Agency CEO discretion to evaluate, and decide:</p> <ul style="list-style-type: none"> At what points to evaluate (may be summative only) To evaluate internally or contract out If contracted out, whether to manage contract through program area or cluster evaluation unit Budget source and amount Need for a Steering Committee and its composition Publication of evaluation report.
3	<ul style="list-style-type: none"> Significant investment (relative to cluster/agency) Named in <i>NSW 2021</i>/strategic plan Moderate to high risk Joint responsibility of two or more Ministers Involvement of external stakeholders or delivery partners Complicated design, with multiple elements (or sub-programs) Not recently reviewed 	<p>Evaluation expected:</p> <ul style="list-style-type: none"> Evaluation plan agreed with partners with clear KPIs and responsibility for data collection Quarantined evaluation budget (for new proposals) Consider independent evaluation by Treasury, consultants or by cluster specialists Steering Committee with membership representing all responsible agencies, NGOs and possibly central agencies Peer review, perhaps by evaluation expert in another cluster Report to responsible Ministers and DGs/CEOs Support and advice from DPC/Treasury if sought Publication of evaluation report.
4	<ul style="list-style-type: none"> Significant investment (relative to government) Resource intensive High priority (at cluster or whole of government level) High risk Complex Controversial Innovative or a pilot, trial or proof of concept Involves multiple delivery partners External reporting / evaluation requirement 	<ul style="list-style-type: none"> Formal evaluation mandatory Detailed evaluation plan covering the program lifecycle with questions, methods, data collection, analyses, KPIs agreed prior to implementation Quarantined evaluation budget Independent evaluation, in consultation with Treasury, with governance led outside the affected agency or cluster Evaluation steering committee chaired or co-chaired by agency Executive; DPC/Treasury Executive representation Peer review by content and evaluation experts and/or Treasury (if not carrying out evaluation) Report results and next steps to relevant Ministers & Cabinet Publication of evaluation report.

Using evaluation – Informing decision making and communicating results

Evaluation is an investment, and to maximise the return on investment evaluation results should be acted on and communicated widely to inform decision makers, stakeholders and the community. How the findings of an evaluation will be used and disseminated should be considered at the planning stage of the evaluation⁴.

Communication of evaluation helps to:

- disseminate knowledge, experiences and key lessons
- promote transparency and accountability
- improve evaluation quality
- contribute to learning and the development of stronger evidence bases
- reduce duplication of effort.

The overall aim of communicating evaluation is to inform decision making at all levels – programs, agencies/ clusters, and government. Therefore, it is important to identify the primary audience for the evaluation, that is, the specific people who will use the evaluation findings and who have the capacity to effect change⁵. Evaluation evidence can be used to inform a range of different types of decisions, such as:

- immediate decisions about the program, including whether to roll out a pilot as a mainstream program
- longer term decisions about the program, including informing budget reviews and the future scale of investment
- how the program should be improved, for example, if the evaluation identifies issues or challenges to delivery
- how future policies and programs should be designed and implemented⁶.

The timing of evaluations should be aligned, where possible, with decision making cycles, with consideration given to budget cycles, corporate planning, Cabinet meetings and other external decision making processes⁷. Evaluation findings and recommendations of major programs should be reported to the relevant Cabinet Committee, where requested.

Evaluation findings can also be communicated to others for different reasons, such as building confidence in the effectiveness of a program, sharing lessons learned from the evaluation with other evaluators in the sector, or promoting the work of an agency.

Good Practice Example: Motor Accidents Authority

The Motor Accidents Authority plans how findings will be disseminated from the outset of program design. The Authority regularly publishes evaluation reports on significant programs on its website, involves other jurisdictions and academic experts in peer reviews, and facilitates dissemination within and external to government through information workshops.

4 HM Treasury – The Magenta Book: Guidance for evaluation, 2011

5 Better Evaluation – Report and support use, 2013, <http://betterevaluation.org/plan/reportandsupportuse>

6 Ibid. 3

7 ACT Government – Evaluation Policy and Guidelines, 2010

5. BUILDING EVALUATION INTO THE PUBLIC SECTOR

The first part of the framework provided an introduction to program evaluation and its importance in designing and delivery programs, as well as some guidelines to promote consistency and transparency in evaluation across the sector.

The experience of other jurisdictions indicates the importance of strong links between evaluation and existing government processes to embed evaluation into the way we do business⁸. The following sections outline how evaluation will be built into the NSW public sector with clear links to performance management, the budget process, and access to government information.

Evaluation and performance

NSW 2021 is the Government's 10-year plan to guide policy and budget decision making and, in conjunction with the NSW Budget, to deliver on community priorities. It sets ambitious long term goals and challenging targets, providing direction and focus for public sector activity. If we are to meet these goals and targets, we need to experiment and innovate, and we need to know what works and what doesn't. We have to be prepared to stop doing those things that have not proven to be effective.

By undertaking evaluations we will be able to measure the performance of programs, agencies, and the sector as a whole. Evaluation will contribute to the evidence base to support agencies to learn how they can best achieve the results they set out to. It will be a mechanism to drive and inform action where performance is not at the level it should be.

Evaluation will strengthen the existing performance management processes and structures in place to track progress in delivering the Government's commitments in *NSW 2021*. Already, in instances where performance is off track against a particular target, agencies are required to produce a remediation plan to demonstrate how performance will be improved. To do this, agencies will need to identify the programs contributing to the target and demonstrate evidence of their effectiveness.

Where evidence is not strong or lacking, agencies should review or evaluate the contributing program (or programs) to help identify issues and performance improvement initiatives.

Good Practice Example: Promoting evidence-based policy development

The Government has established the NSW Statistical Council to promote the use of evidence in policy development and service delivery and assist with improving data quality. The Council includes statistical and data experts from the Australian Bureau of Statistics, the NSW Chief Scientist and Engineer, the NSW Bureau of Crime Statistics and Research, the Bureau of Transport Statistics and the Bureau of Health Information, as well as policy and service delivery specialists from each Cluster.

New program proposals

Program proposals brought forward to the Expenditure Review Committee (ERC) will be expected to include an evaluation or review component to inform decision making. All new program proposals to ERC must:

- include an evaluation plan attached to the business case (see below, Program evaluation plans)
- specify an explicit date for program termination, by which time an evaluation or review to inform future decisions on expenditure is to be completed. For major programs or as requested by ERC, evaluation findings and recommendations must be reported to ERC. Should the program be approved to continue in any form, agencies may also be required to report back to ERC on progress to implement the endorsed recommendations within an agreed timeframe
- specify and quarantine an evaluation budget based on the type of evaluation to be undertaken and estimates of the costs associated (i.e. funding cannot be used for other purposes, unless the program is terminated before the evaluation occurs).

Proposals for additional funding to extend, significantly modify or expand programs should be accompanied by the findings of an evaluation or review including discussion, where appropriate, that recommendations that have been agreed to will be implemented. For example, proposals to expand programs from pilot stages to further rollout or mainstream, or proposals for significant policy changes.

⁸ See the Treasury Board of Canada Secretariat's Policy on Evaluation, the New Zealand Treasury's Circular Using evaluative activity to improve decision making, and South Africa's National Evaluation Policy Framework.

Program evaluation plans

To maximise the benefit of an evaluation, it should be well planned during the program design phase. Agencies are responsible for developing evaluation plans for programs they intend to evaluate. These plans should be negotiated and agreed with all delivery partners, including non-government organisations. Consideration should also be given to the involvement of other key stakeholders and service users. Evaluation plans submitted to ERC will be reviewed by DPC and Treasury as part of normal Cabinet processes.

The details and scope of the evaluation plan will differ from evaluation to evaluation, however, all plans should consider:

- the purpose of the evaluation
- key questions the evaluation will seek to answer
- primary audience of the evaluation
- resources to be allocated to the evaluation
- who will conduct the evaluation
- baseline data and methodology
- budget and timeline
- plans to disseminate and/or publish findings.

More detail on the elements that should be included in an evaluation plan is at Attachment B.

Cluster evaluation schedules

Existing programs should also be periodically evaluated or reviewed to assess their continued relevance, relationship to other programs and government priorities, efficiency and effectiveness in delivering intended outcomes.

To help manage the evaluation of new and existing programs over the medium term, clusters will be required to prepare multi-year rolling evaluation schedules that are reviewed annually.

In the interim, clusters will be required to develop a rolling annual schedule from the 2013/14 financial year (see Transition, below, p. 18). The schedules will be expected to include:

- a list of programs planned for evaluation or review, their expected completion date, and an indication of whether they are suitable for a Treasury-led evaluation
- who will evaluate or review the listed programs
- the governance processes for the schedule, including internal monitoring and reporting
- when the schedule will be reviewed and updated.

In selecting programs for the schedule, agencies should consider the factors outlined previously in Being Strategic – Deciding what to evaluate (p. 9).

Agencies should give priority to programs:

- identified as contributing to *NSW 2021* targets that are off track
- scheduled to terminate during the current financial year
- for which evaluation is required by Cabinet (or a committee) or the Australian Government.

These schedules should be aligned to agency corporate planning cycles and internal decision making processes. They should also take account of reporting requirements to Cabinet, the Australian Government and other external bodies. Schedules should be developed in consultation with DPC and Treasury and submitted to ERC for approval once finalised.

Transparency and accountability

Government is committed to the principles of transparency and accountability, which should be evident at all stages of the process to gather, analyse, interpret and present findings generated by evaluation.

It is considered good practice to publish the findings of evaluations. Publication enhances accountability and transparency by:

- communicating expenditure choices and outcomes to the community
- providing Cabinet with the necessary information to base decisions on
- contributing to quality assurance processes, as methods and findings are open to scrutiny
- contributing to publically accessible evidence bases, which may be used in other research and evaluation activities and to inform the design of future services.

In line with GIPA, agencies are mandated to proactively and publicly release the findings of program evaluations, unless there is an overriding public interest against disclosure of the information. Agencies are encouraged to make evaluation reports available, however large and complex evaluation reports should be provided to Cabinet for approval prior to publication. Publication should be accompanied by a plain English executive summary on the commissioning agency's website.

Beyond written reports, agencies may also consider holding briefing sessions with key stakeholders, participating in seminars and conferences, publishing in peer reviewed journals, and presenting information and data more interactively.

To help build deeper evidence bases and provide a safeguard against errors in design and analysis, agencies are also encouraged to:

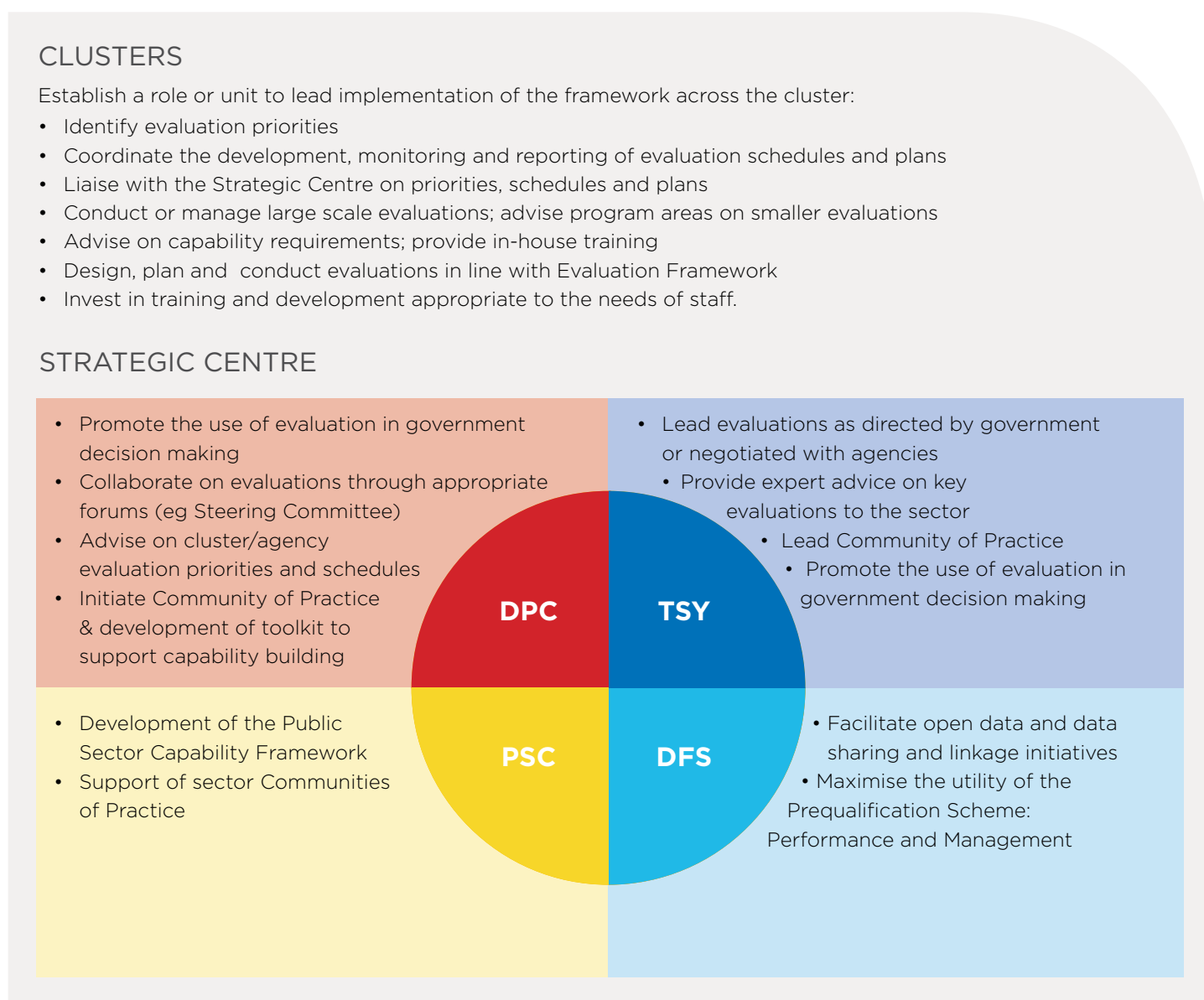
- make de-identified raw data used in evaluations available, preferably in reusable formats
- make analytical assumptions, methods and their testing explicit
- make theoretical perspectives explicit
- make the relationship with past research explicit
- declare financial and other interests⁹.

⁹ Adapted from Evidence Based Policy: Principles of Transparency and Accountability by George Argyrous (2012)

Roles and responsibilities

Roles and responsibilities of clusters and the Strategic Centre in relation to the Framework are outlined in Figure 3 (below).

Figure 3. Roles and responsibilities



6. BUILDING EVALUATION CAPABILITY

Evaluation needs to be integrated into the way we work, rather than seen as discrete activity that occurs after programs are delivered, to satisfy some external requirements, or as a threat to funding.

Building the evaluation capacity and skills of our people is paramount to facilitating the development of this culture and improving the use and quality of evaluation across the sector.

There are already pockets of considerable evaluation expertise in some clusters, but there is much more to be done to develop these skills more broadly across the sector. Clusters will need to invest in training and development to improve evaluation capability and quality. As much as possible, these experts should be involved in the up-skilling of other staff in their cluster.

There are different groups in the sector which each have differing levels of expertise and play different roles during evaluations. To help clusters target appropriate training, four groups have been identified, alongside their needs and options for developing capability at Appendix C.

Community of practice

A community of practice will be established to foster a culture of evaluation across the sector, but also with a broader focus on evidence based policy development. The community will encourage:

- promotion of the role of evaluation in delivering public value for the people of NSW
- understanding of the strengths and limitations of evaluation and its role in evidence based decision making
- closer alignment between evaluation and program/policy design
- sharing of knowledge, expertise and experiences across agency boundaries
- improving skills and sharing resources.

The community will be open to all staff, but will primarily provide an opportunity for evaluation specialists and program staff, including non-government partners, to meet periodically to reflect on their practice, share their experience with others, and explore new approaches to evaluation. DPC and Treasury will have a coordinating role to organise meetings, seminars and disseminate information. The community will be supported by an interactive website to encourage discussion and share information and resources.

Partnerships with specialist bodies

Clusters and their agencies may also wish to pursue partnerships with evaluation bodies relevant to their portfolios to build capability and capacity.

Good Practice Example: Legal Aid

Legal Aid has established a research alliance with the Law and Justice Foundation. As part of this alliance, the Foundation has been engaged to undertake a number of large scale evaluations and to act in a capacity building role through the provision of research and evaluation workshops delivered to internal staff. The workshops include building evaluation into the project design stage and identifying data needs and collection avenues. The alliance also provides ad-hoc consultation processes for current projects and their evaluation.

7. TRANSITION

It will take time to develop the capability and capacity of the sector to periodically and rigorously evaluate programs.

The move towards regular evaluation as an integral part of the program lifecycle will need to be supported by sustained and inspired leadership, sufficient resourcing, and a shift in culture. This section outlines what we aim to achieve over the next three years to support this transition.

Implementing the recommendations of the Commission of Audit

Clusters will have to take steps to implement the recommendations of the Commission of Audit into Government Expenditure agreed to by Government¹⁰ including:

- putting in place adequate financial and data systems to capture and report on program information (such as that listed for Tier 1 programs in Scaling evaluation, p. 11)
- ensuring appropriate information management exists within agencies to ensure program knowledge and history is maintained¹¹.

The Strategic Centre will support this work through the development of the new financial management framework, led by Treasury, which will also:

- clarify the definition of 'program' in consultation with line and central agencies ensuring consistency with the definition in this Framework
- set minimum requirements for agency financial management systems to report data at the program level
- work with central agencies to ensure alignment on other program-level reporting needs, such as workforce and performance data¹².

Treasury has also established a program evaluation unit to work with agencies to evaluate large and significant programs, as recommended by the Commission of Audit¹³. The unit comprises evaluation experts with experience in the government, private and academic sectors. The unit aims to contribute to the development of evaluation capability and an evidence base of program effectiveness by:

- leading selected evaluations in close collaboration with agencies
- providing a source of advice to agencies and other parts of Treasury on evaluation design
- contributing to the development of an evaluation community of practice.

The evaluation unit will start with one or two pilot evaluations before moving into a rolling series of evaluations through this transition process.

¹⁰ NSW Government response to the Final Report of the Commission of Audit, 2012
¹¹ Commission of Audit into Government Expenditure, 2012, Recommendation 127
¹² NSW Treasury, Financial Management Framework: Discussion Paper, December 2012

¹³ Ibid

Figure 4. High level timeline to 2015

NOW

- Limited evaluable data
- Ad hoc evaluations & reviews
- Inadequate information & systems
- Capability & capacity deficits

2015

- Reviews of existing programs
- Develop information systems
- All new program proposals have:
 - Review date
 - Business case
 - Evaluation plan
- Community of Practice in place
- Toolkit launched and maintained

Program review

It is expected that the majority of existing programs have not been designed to be readily evaluated. While more formal evaluations may not be appropriate or feasible for a majority of existing programs, there are other approaches that can be used to test efficiency and effectiveness. Agencies will be required to review their existing stock of programs and put in place systems to enable evaluation at a later date. This might include developing key performance indicators, reviewing data needs, tightening up monitoring and reporting processes for individual programs.

Program reviews, as discussed earlier, are generally focused on program implementation and operation. They seek to analyse and provide answers to variances to planned activities. These reviews will enable agencies to adjust and remediate programs and will help inform future funding decisions. Where a full-scale evaluation is not possible or appropriate, a program review may provide valuable input into decision-making. Programs planned for review should be included in clusters' evaluation schedules. DPC will provide tools and guidance for agencies to use to help them review their programs.

The Figure 4 (above) outlines the high level timeline of what we aim to achieve over the next three years.

A toolkit to support evaluation and review

For evaluations and reviews to be most effective it is necessary for government to have a common understanding of what these activities involve, how they differ in their purpose, as well as how and when they are to be applied in practice. To enable this common understanding, an evaluation toolkit will be developed by DPC and Treasury to provide a set of evaluation and review tools that agencies and their evaluation experts can use.

There is a considerable body of knowledge and a range of high quality tools already being applied by agencies. By opening up access to these resources, and drawing on best practice, we will develop tools that embody the good practice principles and planning processes outlined in this document for use by all parts of government. These tools will be developed progressively and will play an integral part in improving evaluation and review capability and capacity across the sector.

Review of the Framework

The evaluation framework will be reviewed towards the end of 2014 to ensure its alignment with the Financial Management Transformation and the processes surrounding the budget and *NSW 2021*.

8. REFERENCES

1. ACT Government (2010). *Evaluation Policy and Guidelines*, ACT Chief Minister's Department, Canberra.
2. Argyrous, G (2012). *Evidence Based Policy: Principles of Transparency and Accountability*, Australian Journal of Public Administration, vol. 71, no. 4, p. 457-468.
3. Department of Performance Monitoring and Evaluation (2011). *National Evaluation Policy Framework*, The Presidency Republic of South Africa, Pretoria.
4. HM Treasury (2011). *The Magenta Book: Guidance for Evaluation*, UK Government, London.
5. New Zealand Treasury (2003). *Circular 2003/13 Using evaluative activity to improve decision making*, The Treasury, Wellington.
6. NSW Commission of Audit (2012). *Final Report: Government Expenditure*. NSW Commission of Audit, Sydney.
7. NSW Government (2012). *NSW Government response to the Final Report of the Commission of Audit*, NSW Government, Sydney.
8. Owen, JM (2007). *Program Evaluation: Forms and Approaches*, 3rd edition, The Guildford Press, New York.
9. Treasury Board of Canada Secretariat (2009). *Policy on Evaluation*, Treasury Board of Canada Secretariat, Ontario.
10. Trochim, WM (2006). *Research Methods Knowledge Base*, 2nd edition, Accessed www.socialresearchmethods.net/kb/ (version current as at October 20, 2006).

APPENDIX A

GLOSSARY OF TERMS

TERM	DEFINITION
Baseline	Information collected before or at the start of a program that provides a basis for planning and/or assessing subsequent program progress and impact.
Cost Benefit Analysis / Cost Efficiency Analysis	Evaluation of the relationship between program costs and outcomes. Can be used to compare different programs with the same outcomes to determine the most efficient intervention.
Evaluation	A systematic and objective process to make judgments about the merit or worth of one or more programs, usually in relation to their effectiveness, efficiency and/or appropriateness.
Findings	Factual statements about a program which are based on empirical evidence. Findings include statements and representations of the data, but not interpretations, judgments or conclusions about what the findings mean or imply.
Formative evaluation	Evaluation which is generally undertaken while a program is forming (prior to implementation), directed at optimising a program. Typically used to identify aspects of a program that can be improved to achieve better results.
Key Performance Indicator (KPI)	Quantitative or qualitative variable that provides a reliable way to measure intended changes. Performance indicators are used to observe progress and to measure actual results as compared to expected results.
Monitoring	A process to periodically report against planned targets (Key Performance Indicators). Monitoring is typically focused outputs rather than outcomes and is used to inform managers about the progress of a program and to detect problems that may be able to be addressed through corrective actions.
Outcome	A result or effect that is caused by or attributable to the program.
Output	The products, goods, and services which are produced by the program.
Process evaluation	Evaluation which investigates how the program is delivered, and can be used to continually improve programs by informing adjustments to delivery.
Program	A set of activities, managed together over a sustained period of time, that aim to deliver an outcome for a client or client group.
Program Logic	A management tool that presents the logic of a program in a diagram (with related descriptions). It links longer term objectives to a program's intermediate and shorter term objectives. The program logic is used to ensure the overall program considers all the inputs, activities and processes needed to achieve the intended program outcomes.
Program review	Typically quicker, more operational assessments of delivery to get a sense of "how we are going" and to inform continuous improvement. Reviews often take place well after implementation has started and may be a useful alternative when there is insufficient information to conduct an evaluation.
Summative evaluation	Evaluation generally reports when the program has been running long enough to produce results, although it should be initiated during the program design phase. It assesses the positive and negative results of a program, as well as the intended and unintended outcomes delivered. This form of evaluation is used to determine whether the program caused demonstrable effects on specifically defined target outcomes.

APPENDIX B

ELEMENTS OF AN EVALUATION PLAN

1. Specifying the subject of the evaluation

What program or sub-program is being evaluated?

This should include a brief overview of the program objectives, existing evidence base (or program logic if untested), history of development (i.e. who was involved in its development and how long has it been in existence?), and responsible officer/s.

2. Purpose of the evaluation

Why is the evaluation being done?

It could be to help develop a program, to improve the delivery of a program, or to test whether a program is effective.

3. Key evaluation questions

What are the key questions the evaluation will seek to answer, for example:

- To what extent does... ?
- Is there... ?
- In what ways does... ?

This involves selecting the most important aspects of the program to be examined. Carefully consider what you really want 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.

4. Primary audience

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, judgments or recommendations. It is important to distinguish the primary audience from clients and interested stakeholders, and may not be the same as the commissioners of the evaluation (the group that initiates and provides the resources for the evaluation).

5. Evaluation resources

Who commissioned and will manage the evaluation?

Who will conduct the evaluation?

What human resources and materials are available?

Evaluations are always done with resource and time constraints. Someone or a group of people will need time to plan, set up appropriate data management systems, and generally be responsible for all aspects of the evaluation. 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. 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.

6. Baseline data and methodology

What is the baseline to compare program outcomes to?

What are the key performance indicators? Can data be collected from existing data sets?

What methodology will be used during the evaluation (e.g. survey, case study, randomised control trial)?

There are a range of methodologies, data collection and analysis techniques so it might be useful to seek advice from specialist data, research or evaluation areas in your agency/cluster. Ideally, the desired outcomes, associated questions used to verify these outcomes and identified data sources will be considered during program design.

7. Dissemination

How will information about and findings of the evaluation be communicated to decision makers, stakeholders and the community?

What kinds of information will be included (e.g. findings, conclusions, judgments, recommendations)?

Effective dissemination of findings increases the likelihood that the evaluation will impact on decision making. The findings should be presented in a way that the primary audience can understand – a lengthy and complicated report might not be the best way to communicate to your primary audience. Be explicit about intentions to publish the final report.

8. 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?

There are 60 Human Research Ethics Committees in NSW, including the NSW Population and Health Services Research Ethics Committee and the Aboriginal Health and Medical Research Council Ethics Committee. You should consult the specialist data, research or evaluation areas in your agency/cluster regarding the need for ethics approval.

9. Budget and timeline

What funding has been allocated to the evaluation (including source)?

What are the key milestones during the evaluation?

Specify the budget for evaluation. Include a high level plan outlining key milestones, including when the findings are needed to report to Cabinet, a Minister, or Director General. The budget and other available resources will impact on what you can reasonably expect to achieve in the timeframe.

Adapted from Program Evaluation: Forms and approaches, John M Owen, 2007

APPENDIX C

CAPABILITY TARGET GROUPS AND THEIR NEEDS

GROUP	EVALUATION SKILLS	PRACTICE SKILLS	OPTIONS TO BUILD CAPABILITY
1. Senior Executives	<ul style="list-style-type: none"> Understand the role of evaluation as a core component of evidence based decision making Communicate evaluation findings to Ministers Ensure evaluation findings are acted upon 	<ul style="list-style-type: none"> Lead cultural change in the sector Support open and accountable processes of evaluation Participate in the training program for Group 2 	<ul style="list-style-type: none"> Build evaluation framework into existing leadership programs
2. Evaluation experts	<ul style="list-style-type: none"> Formulate program logics for large scale or complex projects Conduct complex evaluations Choose and develop core data for monitoring and measuring progress that may be used across evaluations Present and communicate findings 'upwards' Implement ethical standards for evaluation research 	<ul style="list-style-type: none"> Mentoring and coaching skills Coordination and steering functions for evaluation activity Cross agency/cluster knowledge sharing Coordinate repository of completed evaluations Provide expert advice to Group 3 Draw on high-level expert advice where appropriate Participate in the training program for Groups 3/4 	<ul style="list-style-type: none"> Require high-level evaluation expertise at appointment Work-based learning projects to develop and hone evaluation skills Regular sector wide symposiums with all Group members focused around agreed learning needs (part of the community of practice)
3. Other PS employees	<ul style="list-style-type: none"> Build evaluation thinking into all aspects of program and policy design Select appropriate evaluation methodology Formulate clear evaluation questions Formulate program logics for medium to small scale projects Formulate tender briefs for externally provided evaluations Contract management and monitoring of externally provided evaluations Understand when to use economic evaluation tools such as cost-effectiveness/CBA Identify indicators for measuring performance and outcomes Critique the use of quantitative and qualitative data Present and communicate findings 'across' the agency Ensure ethics guidelines are met 	<ul style="list-style-type: none"> Within agency/cluster knowledge sharing Document learning from completed evaluation to inform future evaluation Provide peer review for completed evaluations Act as 'helpful neighbour' to other tier members conducting evaluations Refer to past experience in developing evaluation plans 	<ul style="list-style-type: none"> System of certification involving regular small 'skills' workshops designed around the framework. External providers may be engaged in provision of workshops but with involvement of Group 2 members Agency or cluster based workshops as needed by agency/ clusters to develop specific needs Present findings/learnings from evaluation projects to whole group through seminars/ workshops (internally and as part of community of practice)
4. Non-government service providers	<ul style="list-style-type: none"> Contribute to program logics provided by other Groups Deliver services in a way that will allow rigorous evaluation, including data needs Contribute to the selection of rigorous and appropriate measures of performance, outputs, and outcomes Facilitate the selection and conduct of appropriate quantitative and qualitative data analysis 	<ul style="list-style-type: none"> Document and communicate learning from individual evaluations to Group 2 	<ul style="list-style-type: none"> Specific training requirements to be developed in consultation with the Group 2 Must involve Groups 2/3 in the training program Require participation in the training program for tender bids to deliver services

www.nsw.gov.au/strategic-centre

