Australian Government

AUSTRALIAN PUBLIC SERVICE ACADEMY

Learning evaluation across the APS

Learning evaluation handbook

Version control

Version	Date	Comment
V1.0	June 2022	Version for apsc.gov.au website.

Contents

Learning evaluation across the APS

Learning across the APS is transforming. Learning practitioners have embraced the vision and targets established by the APS L&D Strategy and APS Workforce Strategy, and learning offerings are evolving to reflect a continuous learning culture and take advantage of a technological revolution.

With so much happening across APS learning, we must capture the stories of positive change and evaluate the effectiveness and impact of this transformation to understand the strengths and ongoing opportunities presented to us.

The APS Learning Evaluation Framework (the Framework) and Learning Evaluation Handbook (the Handbook) assists learning practitioners across the APS to conduct learning evaluation.

Why evaluate?

There are a range of reasons to evaluate learning initiatives. Evaluation is in-depth, values-based and allows an organisation to understand what is happening and why.

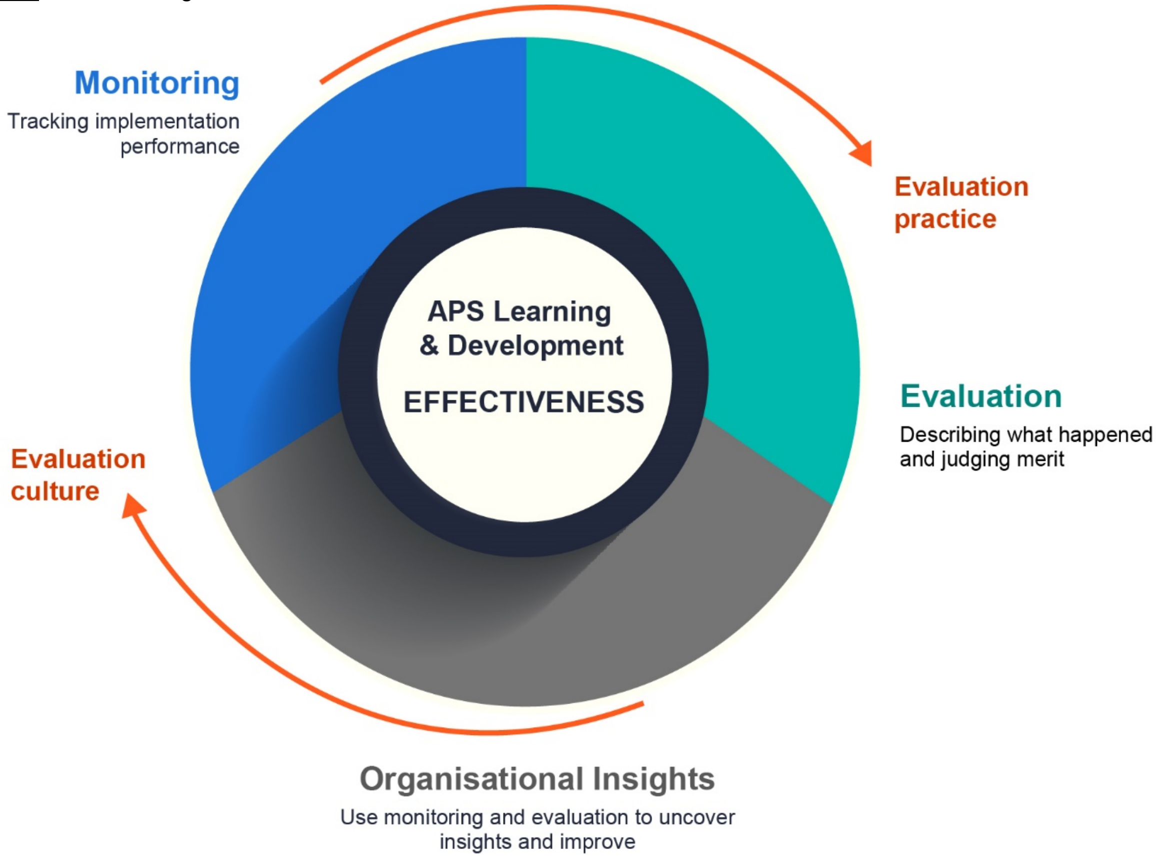
Formative evaluation is the engine that drives improvement and innovation. It is undertaken to assess how well an initiative is, and is not, working and then determines what changes are needed to enhance performance improvement.

Much more than a technical process, formative learning evaluation is deeply influenced by organisational values and the motivation to continually improve learning. When done well, it is aligned with your agency's purpose and strategic goals and leads to improvement.

APS Learning Evaluation Framework

The APS Learning Evaluation Framework (the Framework) emphasises the interdependent relationship between continuous and routine monitoring, evaluation and organisational insights.

Figure 1: APS Learning Evaluation Framework



Evaluating the effectiveness of APS learning is integral to ensuring the quality of L&D initiatives and individual learning. The effectiveness of L&D practice across the APS is determined by evaluating against the following criteria:

- Efficiency
- Impact
- Relevance
- Flexibility
- Accessibility

These criteria are supported by the indicative key evaluation questions and information sources in the table on the next page.

Table 1: Criteria of merit, key evaluation questions and information sources

Evaluation criteria

Effectiveness

Key evaluation questions

- How effective was the initiative in helping the participant perform their work?
- What were the unintended effects (positive and negative) of the initiative?
- To what extent was the initiative successful in achieving its objectives?
- How valuable is the initiative to the participant?

Accessibility

Key evaluation questions

- To what extent did all the targeted audiences benefit equally?
- How well did the initiative contribute to building access to learning?
- How successful was the initiative in reaching hard-to-reach groups?

Flexibility

Key evaluation questions

- How well does the initiative complement other initiatives?
- Are there other initiatives which contradict or inhibit the success of this initiative?
- To what extent can the initiative be adapted and scaled to different environments and contexts?

Relevance

Key evaluation questions

- How relevant is the initiative in responding to the learning needs of the participants, agency and APS?
- How relevant was the initiative for the participants?
- How relevant was the design of the initiative to the environment and context?

Impact

Key evaluation questions

- What difference has the initiative made to participants, agencies, and the APS?
- How has the initiative produced significant change?
- How has the initiative benefited the participants' careers?
- How has the initiative led to other indirect changes?

Efficiency

Key evaluation questions

- How efficient was the implementation of the initiative?
- To what extent were the initiative's objectives met against budget and resourcing?
- How did anticipated costs change? What was the result of this?
- How adequate was the budget in meeting the requirements of the initiative?

Information sources

Primary sources (e.g. participants and supervisors)

- Questionnaires
- Focus groups
- Case studies
- One-on-one interviews
- Pre/post-program participant testing
- User observations

Other sources

- Participation rates
- Participant demographics
- Corporate and business area plans
- Other initiative outlines and work plans
- Initiative proposal and implementation plan
- Benchmarking against comparable initiatives
- Financial and resourcing reporting

Learning evaluation guidance

To accompany the APS Learning Evaluation Framework, the APSC has developed guidance resources for learning practitioners to use when evaluating learning – called the Learning Evaluation Handbook. The Handbook is broken up into nine tasks:

Figure 2: The Nine Evaluation Tasks

Learning Evaluation Handbook

What is the Handbook?

The Learning Evaluation Handbook (the Handbook) has been developed by the APS Academy as part of the APS Learning Evaluation Framework (the Framework) and guidance. The Handbook supports the APS Learning and Development Strategy and the Strategy, Policy & Evaluation craft capability from the Australian Public Service Commission and APS Academy respectively.

The Handbook is underpinned by the Commonwealth Evaluation Policy and provides a short, practical introduction to evaluating learning. It also links to the APS Learning Quality Framework, which emphasises the importance of evaluation being included in learning design for effective and impactful learning experiences.

Intended audience

The Handbook has been created for learning practitioners who want to conduct learning evaluation, but are unsure where, or how, to begin.

Contact the Academy

For further information and support, or to provide feedback on the Handbook, please visit the APS Academy's contact page.

Task 1 – Plan & scope

To ensure evaluation aligns to the guiding principles for evaluation across the Commonwealth and will be useful for your agency, the first task focuses on clarifying the circumstances in which the evaluation has been sought and will be undertaken. Once this is complete, an overarching evaluation plan can be developed. Fact sheet 2 sets out the principles guiding evaluation activity across the Commonwealth.

Understand the operating context

A sound understanding of the reason why the evaluation is required will lead to a fit for purpose and useful evaluation. It will secure the evaluation within your agency's forward work plan.

Supporting operating conditions

Operating conditions which support evaluation are summarised below. These conditions increase the likelihood the evaluation will be fit for purpose, useful and worthwhile.

Table 2. Organisation operating conditions

Vision	- Commitment to evidence-based strategy and policy decision making - Clear responsibilities and accountabilities for embedding a culture of evaluation have been articulated
Enabling environment	- Leadership committed to launching and sustaining an evaluation system over the term - An outcome-oriented, knowledge-sharing, insights and improvement culture and a supportive accountability regime are in place - Permission and space to ask hard questions and curiosity to find out what works and what doesn't - Support for the evaluation principles - Evaluation reporting system and knowledge transfer are in place - Appropriate organisational resources are allocated - Barriers such as resistance and limited buy-in are addressed
Demand-side infrastructure	- Capacity and incentives to deliberately seek out and use evaluation findings - Clear expectations of where, when and how evaluation findings will be used
Supply-side infrastructure	- Skilled personnel with capacity to conduct evaluation - Infrastructure to ensure a credible approach (evaluation policy, standards, framework, etc.)

Fact sheet 3 provides question prompts to better understand your evaluation system.

Why do we need evaluation?

Evaluation is only relevant in certain circumstances and should be reserved for situations where the results of in-depth analysis and values-based appraisal can be used. Each evaluation should be purposeful and make good use of the time and resources allocated to them.

Establishing priorities for a purposeful evaluation requires you to consider who will benefit from the evaluation and who will use the findings. The overarching message is to be clear about how evaluation will help to achieve agency purpose/s or strategic objectives.

Fact sheet 4 outlines when evaluation is warranted in more detail.

Create the overarching evaluation project plan

An overarching evaluation plan outlining how the evaluation will be conducted is essential. The plan should cover:

- **Who will undertake** the evaluation and who else will be involved.
- The **approaches that will be used** for stakeholder engagement to elicit stakeholder interests, expectations for the evaluation and understand why it is important to them. Identify whether there are strong preferences regarding evaluation methods, reporting formats and organisational insight requirements as well as how their engagement can be supported throughout the evaluation.
- **Decision making processes**, including how decisions will be made about the evaluation, who will provide advice, who will make recommendations, and who will be involved in collaborative processes of reflection, insights and deciding action in response to the evaluation findings and recommendations.
- **Resource parameters** such as practical and financial constraints, and how much time is needed to engage participants to undertake data collection, provide feedback and make sense of the data, and to reflect and learn from the findings.
- **How evaluation capacity will be strengthened** such as through accessing internal agency resources, the APS Academy or Evaluation Centre of Excellence, or accessing external resources.
- **The objectives for your evaluation**. Evaluation objectives can be informed by what a good evaluation should do. They are the specific principles-based achievements you want to accomplish.

Fact sheet 5 contains recommendations for evaluation objectives.

Task 2 – Clarify the object being evaluated

Not everything can be, or needs to be, evaluated. This Handbook is for evaluating learning initiatives which comprise a set of related inputs and activities, with a particular long-term aim of helping to achieve agency purpose/s or strategic objectives.

To better understand what is being evaluated and confirm that it can actually be evaluated, this second evaluation task involves clarifying the object to be evaluated.

Logic modelling for evaluation

In the learning and development sector, Kirkpatrick's four-level outcomes hierarchy evaluation model has been a popular approach to learning evaluation for over 40 years. Unfortunately, this approach simplifies the complexity of learning pathways and overlooks individual and organisational factors which may be helping or hindering application of the learning in the workplace.

Learning evaluators are now moving towards more sophisticated forms of logic modelling which capture the complex reality of learning - a practice widely adopted in the broader evaluation field.

To avoid a 'one size fit's all' approach to evaluation, the object of the evaluation needs to be clearly specified before moving onto subsequent tasks. When we clarify a learning initiative through the lens of evaluation, we aim to:

1. Establish the strategic value of the initiative;
2. Ascertain the diversity of learner inputs and learning pathways;
3. Provide a framing for a useful evaluation; and
4. Strengthen evaluative thinking.

Using logic models

To develop and use an effective logic model (also referred to as program logic, program theory, or theory of change), a two-step clarification process should be followed:

Step 1 - Describe the learning initiative
This step should focus on reaching agreement about how the initiative makes a difference and is of strategic value, this will clarify what should be included within the scope of the evaluation.

Step 2 - Develop the logic model
Logic models are a powerful tool for explaining the causal relationships between how an initiative operates and the results it is expected to achieve. They can be used to fully describe the set of tangible inputs and activities, and the intended sequence of results that produce learning transfer in the workplace, and the effects of that learning transfer on agency performance.

Potential negative outcomes that are detrimental and make something worse, rather than better, should also be included. In addition, logic models air assumptions and the reasoning behind why the change process are understood to happens in the way described, why this is of strategic value and what is important about the initiative.

Logic models draw on internal and external sources of expertise, learning practitioner's tacit knowledge, stakeholders experience and what was learned from previous evaluations. Confirmation of the accuracy of the logic model by stakeholders is necessary as logic models guide an evaluation.

Fact sheet 6 describes the main elements to include when developing logic models in evaluation.

Task 3 – Construct key evaluation questions against criteria of merit

This third evaluation task aims to determine exactly what we want to know about the value of a learning initiative. Front-end analysis of the needs of decision makers allows you to establish key evaluation questions and line of enquiry which stakeholders are interested in knowing. This leads to a highly focused, purposeful evaluation which makes good use of limited time and resources and will be taken-up by stakeholders.

Constructing key evaluation questions

Key evaluation questions (KEQs) communicate the scope and focus of an evaluation, guide its conduct and set the structure of the report. KEQs are what make evaluative knowledge more useful and relevant than simply describing what has happened. No more than six KEQs should be agreed upon as evaluation against each KEQ takes time and resources. Too many KEQs can also dilute attention being paid to the main issues of interest.

Using evaluation criteria of merit

Evaluation criteria of merit are used to form evaluative judgements about the quality, value, merit and significance of an initiative. They are the specific dimensions used to distinguish a more successful and worthwhile initiative from one that is less successful and of less strategic value.

Criteria of merit serve as a prompt to ensure different aspects of the initiative are considered and receive the appropriate degree of emphasis. Criteria of merit are not intended to be applied in a routine, standard or fixed way for every initiative or used in a tick-box fashion.

The APS Learning Evaluation Framework's criteria of merit are based on a set from the Organisation of Economic Co-operation and Development's (OECD) Development Assistance Committee. The criteria are:

- Effectiveness
- Accessibility
- Flexibility
- Relevance
- Impact
- Efficiency

These criteria have been designed for use across all APS Learning.

Fact sheet 7 summarises KEQs and criteria of merit in evaluation.

Developing evaluation rubrics

Rubrics are a valuable tool for evaluation. They provide detailed descriptions of the agreed performance dimensions and make explicit the different levels of performance that can occur.

Fact sheet 8 discusses rubrics in evaluation and provides an example.

Task 4 – Develop and implement the monitoring plan

Learning and Development has a strong tradition in performance measurement as part of managing learning delivery and accountability.

While performance measurement can provide immediate data (e.g. participant feedback and attendance/completion rates), we must turn to evaluation-led monitoring and evaluation for a deeper understanding and systematic values-based appraisal of learning.

What to include in a monitoring plan

Along with the evaluation plan, the monitoring plan (link to example below) is built from your logic model and key evaluation questions. Both use a matrix layout to organise their key feature which allows for specification of the focus of the monitoring, relevant measures/indicators and targets, data collection sources, responsibilities and timeframes.

Keep your monitoring relatively straightforward. Monitoring is best placed to capture factual data about ongoing implementation, particularly to understand implementation trends.

Routine monitoring data for evaluation can be similar to the routine performance measurement data collected for management and accountability reporting. The point of difference is the purpose it will serve to address evaluation questions and arrive at evaluative conclusions.

Fact sheet 7 summarises KEQs and criteria of merit for evaluation.

Fact sheet 9 outlines the characteristics of data quality.

An example monitoring plan is provided below, aligning to two criteria of merit.

Key Evaluation Questions	**Focus of monitoring**	**Indicators**	**Targets**	**Data sources**	**Who is responsible and when**
Effectiveness *To what extent is the initiative achieving its objectives?*	Change in knowledge/ skills/ attitudes following participant in learning	Difference; Before – after comparison	X% of learners report increased knowledge	Pre-post survey	Evaluation manager as appropriate
Efficiency *How well are resources being used?*	Cost against budget	Performance against budget	Less than X% variation	Financial records	Program manager on a quarterly basis

Task 5 – Develop and implement the evaluation plan

Evaluation goes beyond measuring performance against indicators. It answers questions about quality, value, merit and learning achievements such as whether a learning initiative is worthwhile and achieving the expected outcomes, what works well (or not) and why.

Without defensible evaluation evidence, you can't really know whether learning initiatives are effective or are achieving your desired outcomes.

What to include in an evaluation plan

Acquiring rich information about learning transfer and its effects requires different inquiry methods than those used to measure attendance and learner perceptions. You should choose appropriate evaluation data collection methods to enhance the quality and potential impact of your organisational insights.

When seeking to answer evaluation questions, chosen inquiry methods must be feasible to implement within the allocated time period and available resources.

An example evaluation plan structure is provided below, replicating a similar structure to the example monitoring plan. The evaluation plan sets out the evaluation data collection methods you will use to develop your evaluation findings.

Key Evaluation Questions	Summary of monitoring	Focus of evaluation	Method (i.e. sample size, data collection method, etc.)	Who is responsible and when

Selecting the right methods for data collection

Acquiring rich information about learning transfer and its effects requires different inquiry methods than those used to measure attendance and learner perceptions. Learning practitioners should choose appropriate evaluation data collection methods to enhance the quality and potential impact of their work on organisational insights. When seeking to answer evaluation questions, chosen inquiry methods must be realistic to implement within your allocated time period and available resources.

Fact sheet 10 contains information about data collection methods for evaluation.

Telling a story

Narrative methods are a form of qualitative, storytelling data collection that can provide a rich source of powerful feedback about learning transfer and its effects.

Although narrative methods don't need to be included in every evaluation and shouldn't be used on their own, incorporating narrative methods into the evaluation plan provides a range of benefits.

Fact sheet 11 describes the benefits of narrative methods in learning evaluation and outlines two narrative methods you can use.

Task 6 – Assemble evidence

Data is converted into knowledge through the process of assembling evidence. Evaluators use a three step process to assemble evidence against evaluation questions: **1 - Data analysis**, **2 - Data synthesis**, and **3 - Development of recommendations**.

Evidence assembly is informed by in-depth subject knowledge and understanding of context. It is supported by common sense, an analytical and curious mind, and a capacity to build a line of argument.

Data analysis

Data analysis involves organising and interpreting your data sets and considering their interrelationships. Data analysis begins when the identified data from your monitoring and evaluation plans begins coming through. You can then begin to form insights and hypotheses from this analysis and continue to explore.

Findings against evaluation questions can be displayed and summarised in ways that others can engage with and find useful. Only after data analysis has been completed can you begin to draw out meaning and form conclusions from the data that has been collected.

Fact sheet 10 contains information about data collection methods for evaluation.

Data synthesis

Data synthesis sits at the centre of any evaluation. It is the process of combining evidence to form an evaluative judgement, and deals with values in a transparent way.

Performance targets and evaluation rubrics offer transparency as judgements of actual performance and are assessed against the performance of the initiative. Your earlier effort spent on constructing key evaluation questions and criteria of merit rubrics starts to pay off here.

Subsequent value-laden conversations about performance and your conclusions the initiative's merit demonstrate how well values have been understood and shared.

Fact sheet 7 summarises KEQs and criteria of merit for evaluation.

Fact sheet 12 provides an example of how evaluative judgements can be made transparent.

Form recommendations

The final step of the evidence assembly task involves translating evaluation conclusions and judgements into a set of preliminary recommendations and lessons for adaptation and improvement.

Recommendations suggest a course of action for how to proceed in light of your evidence and conclusions. Your recommendations for further consideration need to be useful, feasible and carefully worded as they will attract scrutiny. Before they are finalised, recommendations can be incorporated in your draft report and discussed during facilitated conversations for organisational insights.

Fact sheet 13 outlines good practice for developing recommendations.

Task 7 – Report

Good evaluation reports convey evaluation findings, conclusions and recommendations in ways that are useful and inform the audience about the changes required.

Evaluation report writers should consider audience information needs and timelines, and follow a standard evaluation report structure.

Established information and decisions

When considering audience information needs and timelines, evaluation report writers can return to decisions made at the outset of the evaluation and captured in the Terms of Reference. Specific choices relating to who requested and who will use the evaluation, and what needs to be learned from this evaluation to improve learning and development in the agency have been pertinent throughout the evaluation process. They continue to be important for the reporting phase. In effect, the report responds to these initial decisions.

Consider you audience

Evaluation reports can have three types of audiences:

1. Your **primary audience** are the people who requested the evaluation and are the major decision makers of the initiative, with responsibility for leading the initiative's design and implementation.
2. The **secondary audience** are the people involved in the program in some way, although not responsible for leading the initiative's design and implementation.
3. Your **tertiary audience** are the people positioned outside of the agency. They may need more background information and descriptive information about the initiative and its evaluation approach to be able to consider the initiative's relevance for improving similar initiatives in their own context.

Report structure

Learning evaluation reports follow a standard reporting structure, such as:

1. Title page
2. Acknowledgements
3. Table of contents
4. List of tables and figures
5. List of acronyms or a glossary of terms
6. Executive summary
7. Introduction
8. Initiative description including the logic model
9. Evaluation approach
10. Data collection methods
11. Results (i.e. findings and conclusions) organised by key evaluation questions
12. Recommendations
13. References and appendices

Adhering to this standard reporting structure assists the audience to navigate the detailed analytical content of an evaluation report.

Fact Sheet 14 provides tips to expedite the reporting process and a detailed checklist for evaluation report content.

Task 8 – Facilitate insights, adaptation and improvement

Formative evaluation drives improvement and innovation. When learning practitioners share evaluation knowledge through reporting, they're helping others and themselves to learn.

This knowledge sharing taps into subject matter expertise and helps everyone learn from success and mistakes to improve learning initiatives.

Acting on insights

The insights generated through formative evaluation can foster a knowledge-sharing culture, guiding the allocation of learning resources, improving the design and implementation of initiatives, and achieving better outcomes.

Ultimately what matters is that the new evaluative knowledge that has been created is shared. Only then can learning initiatives be fine-tuned to advance agency outcomes in the future. This fine-tuning may include making adjustments to the logic model and what is needed to achieve the initiatives intended outcomes.

Facilitation

For this task, your role as evaluator transforms into that of a facilitator. The facilitator's job is to make it easier for the group to learn how to proceed together. To be constructive and improvement-focused, this participatory process must overcome biases that can thwart organisational insights, including:

- Preoccupation with success;
- Favouring action over reflection;
- Codes of engagement that limit innovation; and
- An over-reliance on external experts.

Embed evaluative thinking

As the evaluation facilitator, you must embed evaluative thinking and evaluation findings into the process of reflection, dialogue and action planning. You must also enable group learning, tend to group dynamics and enable the flow of discourse. When what should be done is far from obvious, the facilitator focuses on collaborative sense-making and prioritises action.

- Collaborative sense-making occurs though reflection and dialogue when people engage in the process of making sense of information about complex and dynamic situations. Often, this requires being able to scrutinise evaluation findings with an open mind.
- Prioritise action-taking by considering possible options which make a positive difference and review the recommendations to operationalise the options. Good decision making relies on knowing what action it will take to bring about improvement, being clear about what needs to change, and as a result prioritising resources to achieve better outcomes.

Fact sheet 15 provides a set of questions to prompt collaborative sense-making and action-taking, and describes deeply ingrained biases that can interfere with organisational insights.

Task 9 – Wrap up and review

A useful evaluation does not end with reporting, organisational insights or even adaptation and initiative improvement. A final task remains to support both ongoing evaluation and to improve evaluation practice across the APS learning.

Wrapping up the evaluation includes publishing your report, identifying any follow up actions to implement the evaluation recommendations, and reviewing the overall project to consider how well the evaluation achieved its objectives.

Publish and disseminate

Evaluation findings should be made transparent by default, unless there are good public policy reasons not to. Publication supports accountability, continuous improvement and helps to embed a culture of evaluation across the Commonwealth.

Each agency will have their own formal authorisation process for disseminating evaluation reports and related material which should be followed. There are many opportunities to contribute to the system-wide culture of learning by connecting with other people across APS learning such as the APS Academy and APS L&D Community of Practice, as well as evaluation networks and other forums.

Identify follow-up actions

A program management response should be articulated to convey whether each recommendation in the evaluation report is accepted, partially accepted or rejected. If the recommendation is rejected, a reason why should be provided. Where a recommendation is accepted, the appropriate learning team(s) should proceed to implementation.

The status of follow-up actions should be periodically reviewed and updated in a tracking system and reported to senior managers with responsibility for their implementation.

Review

Improving the quality of your evaluation and critically reviewing the evaluation process and its outcomes against the evaluation's initial objectives completes the evaluation process.

Evaluation review seeks to improve future evaluation practice for enhancing the value, influence and use of evaluation.

The objectives that you initially set for your evaluation were informed by what a good evaluation should do. The initial intention was to comply with the principles-based approach of the Commonwealth Evaluation Policy and the APS Learning Evaluation Framework. The review of the evaluation processes and its outcomes can be made against those original objectives. This desktop exercise should seek to:

- Describe how conducting the learning evaluation contributed to agency outcomes to enhance the value, influence and use of evaluation; and
- Generate lessons for future evaluations of learning initiatives through assessing what parts of the evaluation went well, what didn't and the overall success of the evaluation.

Fact sheet 16 provides a checklist to review the evaluation process and its outcomes.

Page left intentionally blank.

Fact sheets

This page left intentionally blank.

Fact sheet 1: Formative Evaluation

The Learning Evaluation Framework takes a formative evaluation approach that it is future-orientated, constructive and improvement focussed. Evaluation is considered *formative* when it is conducted with the intention to identify aspects of an initiative that can be improved to achieve better results.

We turn to 'formative evaluation' to know whether the initiative is unfolding as intended, identify obstacles and limitations, or opportunities that have arisen unexpectedly. It focuses on the concerns of Senior Executives and others interested in enhancing agency performance. Formative evaluation is undertaken to help manage uncertainties, gain an understanding of what makes a difference, and identify aspects of an initiative that can be improved to achieve better results. Accordingly, formative evaluation is a valuable part of the ongoing management process to fine-tune learning initiative design and delivery.

Because learning and development is never complete, formative evaluation should be the primary mode of evaluation to provide critical feedback for learning, improvement and innovation. Formative evaluation offers a systematic and structured way to harness feedback across the system. It can be conducted during the development or delivery of an initiative with the intention of providing feedback to improve the current initiative, or between six weeks to six months after the initiative ended in anticipation of improving future iterations of the initiative.

Formatives evaluation is appropriate for both:

- 'Process evaluation' which refers to the activities and events that occur as an initiative is delivered such as the administration, procedural/operational matters, and implementation. To increase the likelihood of success through rectifying problems early, formative evaluation may be conducted to review whether delivery is unfolding as planned, attendance rates match those expected, and identify obstacles or unexpected opportunities that have arisen. Otherwise implementation is evaluated after the initiative has ended to enhance processes in the future.
- Outcome/ impact assessment which determines the merits to the changes, results and impacts that flow from the initiative that may be positive or negative, intended or unintended. This includes:
  - Immediate capability gains and participant views on the initiative's relevance and quality.
  - Post-implementation to ascertain whether the initiative met its objectives. It focuses on application in the workplace of the capability, including enablers and barriers to application, and participant views on the benefits of the capability.

Formative evaluation is different from summative evaluation which is often undertaken to meet accountability obligations and to secure further departmental funding. Formative and summative evaluation can be seen as two sides of the same coin in that they are closely related but emphasise different aspects of the same situation. The similarities and differences between formative and summative evaluation are summarised in the box below.

	Learning and improvement focused **Formative evaluation**	**Accountability focussed** **Summative evaluation**
Process evaluation	Process focused on internal dynamics and intended to inform decisions about improvement, primarily regarding implementation	Process focused on internal dynamics and intended to inform decisions about whether or not to refund the initiative
Outcome/ Impact evaluation	Focussed on interim or long-term results and intended to inform improvement	Focussed on interim or long-term results and intended to inform decisions about whether or not to refund the initiative

Fact sheet 2: Principles guiding evaluation

The Commonwealth Evaluation Policy 2021 describes the principles guiding evaluation activity across the Commonwealth[1]: These principles are:

- **Fit for purpose** – the choice to evaluate and the scale of effort and resources allocated to an evaluation should be proportional to the value, impact, strategic importance and risk profile of the program or activity. Methods should differentiate between evaluations to inform program administration and evaluations to inform policy decisions.
- **Useful** – evaluations to inform program delivery should be designed for the purposes of continuous improvement and accountability against objectives, while evaluation for decision-making should be designed for the purpose of defining achievable outcomes, taking account of any pilots, prototyping or experience from other jurisdictions. A strong understanding of Government policy intent is required, both when evaluation is used as a monitoring tool and when it is an input to new program design.
- **Robust, ethical and culturally appropriate** – evaluations should be well-designed, identify potential evaluator bias and take account of the impact of programs and evaluations on stakeholders. Robust data and evidence should provide performance insights and drive continuous improvement for programs in the delivery stage. Where vulnerable cohorts interact with a program, ethically and culturally appropriate approaches should be considered for collection, assessment and use of information.
- **Credible** – evaluations should be conducted by people who are technically and culturally capable. The collection and analysis of evidence should be undertaken in an impartial and systematic way, having regard to the perspectives of all relevant stakeholders. Evaluations should adhere to appropriate standards of integrity and independence.
- **Transparent where appropriate** – to be useful, evaluation findings should be transparent by default unless there are appropriate reasons for not releasing information publicly. To support continuous improvement, accountability and decision-making, evaluation findings should be provided to appropriate stakeholders.

[1] Commonwealth Evaluation Policy. Evaluation Centre of Excellence, Department of Finance (December 2021). https://www.finance.gov.au/government/managing-commonwealth-resources/planning-and-reporting/commonwealth-performance-framework/commonwealth-evaluation-policy

Fact sheet 3: Question prompts to better understand your evaluation system

1. Key drivers	• What is driving the demand for evaluation?
2. Leadership commitment and accountability	• How does leadership support for evaluation manifest? • Who is ‘championing’ evaluation? • Is there an ongoing commitment to sustain evaluation? • Who is responsible for ensuring that evaluation is a functioning part of the Learning team?
3. Resourcing	• Where do resources to develop evaluation infrastructure and build evaluation capacity come from?
4. Technical Capacity	• Is there data system and infrastructure capability to collect reliable data? • Is there adequate analytical capacity? • Is there an agency evaluation unit that could serve as credible ‘partners’?
5. Infrastructure to use evaluation findings	• Do agency guidelines set how evaluation information gets ‘used’? • Are ‘incentives’ for using evaluation findings adequate? • What mechanisms are in place for reporting or tabling evaluation findings?
6. Infrastructure to conduct evaluation	• Is there agreement to adhere to Commonwealth Evaluation Policy? • Are agency structures in place that describe roles, responsibilities and expectations of the operation and use of evaluation? • Is adequate resourcing in place?

Fact sheet 4: Circumstances where evaluation is warranted

Evaluation is warranted for learning initiatives intended to make a significant contribution to an agency's strategic objectives. The initiative must also meet at least one of the following conditions:

1. The initiative has not been evaluated recently where previous evaluation recommendations are still being implemented.
2. There is room to improve the quality of regular and ongoing performance information for accountability reporting.
3. There has been substantive change to the initiative's environment and/or design.
4. Anecdotal information or performance information suggests that initiative performance can, and should, be improved.
5. An understanding and clear definition of what learning success is poorly understood. In particular, little is known about high order learning level outcomes such as:
    a. 'Learning transfer' – i.e. when a person uses what was learned to perform their work tasks successfully; and
    b. The effects of this learning transfer on the person, their co-workers and agency.

Fact sheet 5: Checklist for setting evaluation objectives

Purposeful and fit for purpose	
☐	The evaluation is tailored to the specific circumstances and improvement needs of learning and development.
☐	The rationale for the evaluation, its scope, evaluation questions and findings articulation, are agreed early.
☐	An intentional and purposeful planning process culminates in the Terms of Reference.
☐	Attention is devoted to the full range of individuals and groups invested in the object being evaluated.
☐	The evaluation is completed in a timely manner so as to ensure the usefulness of the findings and recommendations.
Useful	
☐	The evaluation, and its report, takes full account of the diverse attributes of APS people and their agency's settings.
☐	The evaluation will acknowledge the diversity of learning initiatives objectives.
☐	The evaluation will identify the assumptions of the learning initiative's operation.
☐	The evaluation will build evidence about what has made a difference.
☐	The evaluation's reporting format and dissemination strategy are appropriate and accessible.
☐	The evaluation conveys the process of how the initiative causes change and impacts on agency outcomes.
☐	The evaluation will provide useful information about what makes a positive difference on decision making.
☐	The lessons learned about conducting evaluation and how the process can be improved are documented.
Ethical and culturally appropriate	
☐	Actionable recommendations to enhance adaptability, process and use will result from this evaluation.
☐	The evaluation will manifest professional integrity, respect the rights of individuals and be sensitive to the setting.
☐	Pressure will not be exerted to suppress, distort or alter findings or recommendations.
☐	Transparent arrangements for publishing evaluation reports are followed.
Credible	
☐	Appropriate data can be retrieved and/or collected to answer the key evaluative questions.
☐	The evaluation will build understanding about how well the initiative contributes to agency objectives.
☐	Evaluation questions are decided early and guide the structure of the report.
☐	Evaluation questions inform data collection and analysis.
☐	The diverse attributes of APS people, the agency settings and the adequacy of resources are taken into account.
☐	Findings describe positive outcomes where initiatives are making a difference and situations where they didn't.
Transparent	
☐	The program logic is present in the evaluation and an assessment on if it needs to be revised is conducted.
☐	The Criteria of Merit are defined in unambiguous terms and applied in a transparent manner.
☐	Stakeholder participation is meaningful.
☐	Findings will be transparent unless there is a legitimate public policy reason they cannot be released publicly.

Fact sheet 6: Simple hints for developing logic models in evaluation

Logic models are usually developed in the planning phases to explain expected results. Subsequently, they can be revised when more information comes to light, including information about additional positive results (which are beneficial) or negative results (which are detrimental). It can also show other factors which contribute to producing results, such as context and other projects and initiatives.

The main elements of traditional pipeline logic models are

- inputs - financial, human resourcing and other inputs that are required to deliver learning initiatives
- activities involved in delivery such as design, coordination
- learning products and services (referred to as ‘outputs’)
- learner’s outcome pathways leading to learning transfer that takes place in the workplace and the effects of that transfer for the agency (i.e. long-term impacts); and
- assumptions that underpin how expected changes will happen.

Traditionally, logic models are drawn as box and arrow diagrams (inputs->activities->outputs->outcomes->long-term impacts). Sometimes they are shown in a table. The visual representation of a logic model can incorporate creative forms, shapes, labels and colour coding.

Approaches to drawing logic models should not be limited to conventional pipeline logic models that emphasise the initiatives implementation over the learning transfer that happens in the workplace. However, this design can be very effective for initiatives that have a well-defined and predictable relationship between the application of routine, non-expert capability and the simple organisational problems they address.

Program theory or theory of change approaches are better suited to most learning initiatives. While they retain the main elements of pipeline logic models, more attention is given to the results chain leading to agency outcomes such as incorporating other factors which contribute to producing impacts such as external contextual and other related projects or initiatives, more causal pathways and differentiated outcome patterns. This approach to visualising logic models tends to be more concerned with whether the set of underpinning ideas are sound.

Fact sheet 7: Key Evaluation Questions and Criteria of Merit in evaluation

The summary table below describes seven Criteria of Merit (CoM) based from the Development Assistance Committee (DAC) of the Organisation of Economic Co-operation and Development (OECD). Definitions, example KEQs to illustrate how the specific dimensions can be applied, and the inherent challenges faced in addressing these questions are described against each CoM.

These criteria can be used to shape Criteria of Merit for your own Evaluation Framework.

Evaluation Criteria of Merit (CoM) and Evaluation Questions

Relevance

The extent to which the learning initiative is suited to the priority learning needs of individual and the agency, and can be adapted when circumstances change. [Learn more about Relevance.]

Key evaluation questions:

- Is the initiative doing the right things?
- Is the problem being addressed of critical importance?
- Is the initiative aligned to organisational priorities/ goals?

Sub questions:

- How relevant is the initiative in responding to the learning need(s) of the agency/APS and participants?
- How relevant was the initiative to the targeted individuals?
- How relevant was the initiative design to the environment - political, economic and cultural?

Dimensions of relevance

- Sensitivity and responsiveness to context
- Quality of design
- Adapting over time

Challenges

- Where the context changes over time but it is unclear if, and how, the initiative was adapted

Coherence

Consistency and compatibility of the initiative with other Learning and Development initiatives, agency policies, and priorities. [Learn more about Coherence.]

Key evaluation questions:

- How well does the initiative fit?

Sub questions:

- How well does the initiative complement other agency initiatives?
- Are there other initiatives which contradict or inhibit the success of this initiative?
- To what extent can the initiative be adapted and scaled to different contexts?

Dimensions of relevance

- The initiative supports or undermines other initiatives
- Other initiatives support or undermine this initiative

Challenges

- Knowledge of policy priorities and access to relevant documents, data and informants
- Too broad a scope

Effectiveness

The extent to which the initiative achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups. [Learn more about Effectiveness.]

Key evaluation questions:

- Is the initiative achieving its objectives?

Sub questions:

- How effective was the initiative in helping participants perform their work?
- What were the unintended effects (positive and negative) of the initiative?
- To what extent was the initiative successful in achieving its objectives?
- How valuable is the initiative to the participant?

Dimensions of relevance

- Achievement of objectives
- Varying importance of objectives and results
- Differential results across groups and understanding the factors that influence outcomes
- Unintended effects

Challenges

- Lack of appropriate information or disaggregated data
- Poorly phrased, vague or difficult-to-measure objectives/purpose statements
- Attribution of results

Efficiency

The extent to which the initiative delivers, or is likely to deliver, results in an economic and timely way.

Learn more about Efficiency.

Key evaluation questions:

- How well are resources being used?

Sub questions:

- Was the initiative designed to optimise technology to enable access and support for everyone, anywhere, anytime?
- How efficient was the implementation of the initiative?
- To what extent were objectives met against the outlay of budget and resourcing?
- To what extent are stakeholders satisfied with initiative outcomes against their outlay?
- How were anticipated costs adjusted throughout implementation? What was the result of this?
- How adequate was the budget in meeting the requirements of the initiative?

Dimensions of relevance

- Economic efficiency - conversion of inputs (funds, expertise, natural resources, time, etc.) into outputs, outcomes and impacts, in the most cost-effective way possible
- Operational efficiency

Challenges

- Finding suitable comparisons
- Navigating different concepts of efficiency and the associated methods and tools

Impact

The extent to which the initiative has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects. Learn more about Impact.

Key evaluation questions:

- What difference does the initiative make to individuals and the agency?

Sub questions:

- What difference has the initiative made to participants, agencies, and the APS?
- How has the initiative produced significant change in how the participants work?
- How has the initiative benefited the participants' careers?
- How has the initiative led to other indirect changes?

Dimensions of relevance

- Higher-level effects
- Significance
- Differential impacts
- Unintended effects
- Transformational change

Challenges

- Securing clarity on what impact means as a concept for the relevant stakeholders, what impact was intended for this initiative and how it would be achieve
- Availability of baseline and indicator data
- Identifying the degree to which the initiative causes the impact and where there are significant unexpected or unintended effects

Sustainability	The extent to which the net benefits of the initiative continue or are likely to continue. Learn more about Sustainability.

Key evaluation questions:

- Will the benefits last?
- How does the initiative contribute to achieving a better and more sustainable future?

Dimensions of relevance

- Understanding the enabling environment
- The continuation of positive effects
- Risks and trade-offs

Challenges

- The initiative has not achieved its intended results (assessed via effectiveness), positive unintended benefits or made contributions to impact
- Timing of the evaluation
- Evaluating factors likely to influence sustainability
- Securing clarity on what impact means as a concept for the relevant stakeholders, what impact was intended and how it would be achieved
- Availability of baseline and indicator data
- Identifying the degree to which the initiative causes the impact and where there are significant unexpected or unintended effects

Fact sheet 8: Rubrics in evaluation

Rubrics help make the process of evaluative judgment-making accessible and boost the commitment within an organisation to deliberately seek out, act on and value evaluation. They consist of two of dimensions of performance – the evaluation focus (i.e. Criteria of Merit); and (ii) merit determination - the definition of what performance looks like at each level (Oakden, 2013).

A strength of rubrics is that they have been found to facilitate clear thinking and collective sense-making, and strengthen engagement throughout all stages of an evaluation. They keep everyone involved in the evaluation focused on the things that matter. What really counts are the conversations that occur around the rubric, as they promote transparency and help everyone involved to unpack the aspects of the evidence that are their key focus (King et al, 2013).

When using a rubric to evaluate a learning initiative, it is important to remember that you are trying to find the best description of performance based on the evidence that you have. Using consistent, agreed upon, and objective criteria of rubrics to make evaluative judgements will provide a range of benefits, including[2]:

- Support the development of a culture of evaluation within an agency.
- Assessment against established Criteria of Merit is transparent, reduces bias and stops ‘moving the goalposts’ in order to gain an advantage and hope that no-one notices.
- Keep evaluation focused on the things that matter.
- Forming an assessment against established Criteria of Merit that resonate within a learning and development team will enhance the credibility of new evaluative knowledge and the set of findings against key evaluation questions, and increases the likelihood that they will be used.
- Descriptions of the most salient criteria of exceptional and below expectations levels of performance can help practitioner’s judge quality and improve their practice.
- Reporting can be more explicitly evaluative and is often more focused and concise.
- An assessment's reliability can be enhanced when rubrics are used consistently and objectively across cycles of an initiative being implemented.

An example rubric for implementation is provided below. It shows an example rating scale to answer the ‘effectiveness’ key evaluation question ‘*is the initiative achieving its objective.* As an example, the rubric categorises its ratings against the Continuous Learning Model’s learning methods, outlined in the APS Learning and Development Strategy.

[2] King, McKegg, Oakden and Wehipeihana (2013)

Rating scale example for effectiveness criterion

RATING	WORK	PEOPLE	COURSES	RESOURCES
Clearly effective - a consistent story showing that the learning initiative method is effective in delivering the learning and development needed	Evidence that most APS people deliberately developed relevant knowledge and skills through on the job experience, taking on an increased span of control, and undertaking reflection	Evidence that most APS people deliberately developed relevant knowledge and skills through connecting with other people	Evidence that most APS people deliberately developed relevant knowledge and skills through participating in online courses, MOOCs and webinars, and/ or attending classroom training	Evidence that most APS people deliberately developed relevant knowledge and skills through searching the internet and intranet for resources and information; reading books, scholarly articles and magazines; and accessed online podcasts, videos, articles and blogposts
Mixed effectiveness - in some respects the learning initiative is effective in delivering the learning and development needed, but in other significant respects, it was not effective	Evidence that the learning initiative is effective, but in other significant effects or for a significant proportion of APS people, the work method was not effective in delivering the learning and development needed.	Evidence that the learning initiative is effective, but in other significant effects or for a significant proportion of APS people, the people method was not effective in delivering the learning and development needed.	Evidence that the learning initiative is effective, but in other significant effects or for a significant proportion of APS people, the courses methods was not effective in delivering the learning and development needed.	Evidence that the learning initiative is effective, but in other significant effects or for a significant proportion of APS people, the resources method was not effective in delivering the learning and development needed.
Clearly not effective - the learning initiative was not effective in delivering the learning and development needed	Evidence that the work method was not effective in supporting the learning and development needs of most APS people	Evidence that the people method was not effective in supporting the learning and development needs of most APS people	Evidence that the courses method was not effective in supporting the learning and development needs of most APS people	Evidence that the resources method was not effective in supporting the learning and development needs of most APS people

Fact sheet 9: Characteristics of data quality

- **Validity** – related to truth. The degree to which the data represent what they purport to measure or describe.
- **Reliability** – data are collected consistently; definitions and methodologies are the same when doing repeatedly over time.
- **Completeness** – data are complete and there is no missing data or data elements.
- **Precision** – data have sufficient detail.
- **Integrity** – data are protected from deliberate bias or manipulation for political or personal reasons.
- **Availability** – data are accessible so they can be validated and used for other purposes.
- **Timeliness** – data are up-to-date current and available on time.
- **Data backup where appropriate** – onsite and offsite, automatic and manual processes to guard against the risk of data being lost or corrupted.
- **Archive data for future use** – systems to store non-identified/ de-identified data so that they can be accessed for verification purposes or for further analysis and research in the future.

Fact sheet 10: Data collection methods in evaluation

The field of evaluation now accepts a wide variety of methods because they each enable different insights to emerge to better understand the complexity of learning and development. The trend towards mixed methods evaluation involves the planned use of two or more different kinds of quantitative and qualitative methods in the same evaluation. Learning practitioners can be more thoughtful about choosing appropriate methods to enhance the quality and potential impact of their work on organisational insights.

The options for organising the assembly of data and summarising can be grouped into two categories and correct protocols for each should be respected: quantitative data and qualitative data collection and analysis.

Quantitative data generally refers to observations that are represented in numerical form such as cost in dollars, ages and number of participants, events, hours and days. Quantitative data can be analysed with statistics to allow for numerical descriptions of populations and samples, such as frequency of occurrence, percentages, ratios and other trends. Typically, tables, charts and graphs are used to present the emerging themes and trends of quantitative data.

Qualitative data are text based in representation. Text refers to spoken or written word that result from interviews, observation notes, audio or video recordings, documents, publications, speeches and discussions, and multi-media forms of communication such as an agency intranet and email communications, websites, podcasts and social media.

Contextualised understanding is the object of collecting quantitative data. Qualitative approaches are invested in the diversity of contextualised, value-laden meaning constructed by various participants in a given setting. The analytical process of qualitative data typically involves sorting and coding before analysis begins which is done entirely manually or with the assistance of qualitative data analysis software tools. Either way, a person is needed to organize, interpret and make meaning from the data.

Using a combination of qualitative and quantitative data can improve an evaluation by ensuring that the limitations of one type of data are balanced by the strengths of another. This will ensure that understanding is improved by integrating different ways of knowing. Most evaluations will collect both quantitative data and qualitative data, however it is important to plan in advance how these will be used.

Fact sheet 11: Narrative methods in learning initiative evaluation

Narrative methods provide a range of benefits in improvement orientated evaluation of initiatives with complicated outcome patterns such as:

- Narrative data can be analysed using existing theoretical or conceptual frameworks such as program logic or assessed for emergent themes.
- They rely on people making sense of their own experiences and environments.
- Stories can focus on particular initiatives and contextual factors that influence outcomes.
- Stories can be systematically gathered and claims can be verified from independent sources or methods.

The Success Case Method, and Most Significant Change are two examples of case based story-telling approaches with an improvement orientation. Although they have slightly different purposes and procedure, they are both contextually responsive, structured around notions of success and provide useful insights about learning transfer and the effects of that transfer.

The Success Case Approach

The Success Case Method, developed by Robert O. Brinkerhoff (2003), combines in-depth analysis of extreme groups with a case study and storytelling for the purpose of finding out how well an initiative is working and explaining the role of contextual factors that differentiate successful from unsuccessful adoption of new learning. The method is based on the assumption that any learning initiative, no matter how successful or unsuccessful, will always include some success and some failure. The method seeks to uncover the most impactful successes and failures of an initiative and then tell the stories behind them, backed by evidence.

The Success Case method is not designed to help judge the overall value and merit of an initiative, but rather to learn from the most and least successful cases. The method has been most frequently used to evaluate staff training and related human resource programs, although conceptually it could be applied to other types of programs as well.

The Success Case method has two parts. First the evaluator reviews self-reported survey data to find the outliers in order to be able to study them in more detail. These are the cases where the learner has been particularly successful or unsuccessful at applying the new techniques and tools that the learning initiative intended to give them and what, if anything, they are accomplishing. Second, extreme cases are randomly selected from the two ends of the success continuum to dig deep and explore (i) exactly what the participant used, when and how they used it; (ii) what was accomplished; (iii) the merit, significance or value of what was accomplished; and (iv) the organisational contextual factors that enabled or constrained their application and results achieved. Selected participants are invited to tell descriptive stories about both the features of the training that were or were not helpful as well as other organizational factors that facilitated or impeded success.

The most compelling and descriptive success cases are communicated in story form within an evaluation report. A success story is not a testimonial or critical review but a factual and verifiable account that describes how valuable a participant found a new method, tool or capability. When necessary, independent evidence is sought that would corroborate the success claims.

The significance of comparing stories of successful learning with stories of unsuccessful learning from an initiative is that it allows program managers to identify systemic factors that enable some participants to successfully leverage the initiative while others have nowhere near that success.

The Most Significant Change

The ‘Most Significant Change’ (MSC), developed by Jess Dart and Rick Davies, is a participatory, values inquiry story-based technique. Essentially the method involves participants in an initiative reporting what they perceive to be a ‘significant change’ as a result of their involvement in the initiative. The collection of stories is presented to designated stakeholders such as learning program designers, decision makers and implementation staff who deliberate on the value of participant’s outcomes in a systematic and transparent manner. This deliberation process involves discussion to select the stories that are considered to be most illustrative or significant.

What distinguishes MSC from other narrative approaches is the explicit focus on valued change trajectories such as learning pathways. Initiative design and implementation practices and processes become more open to adaptation and improvement when certain change trajectories are endorsed as the most desirable outcomes, and evaluative judgements are made more visible. As a consequence, more of these explicitly valued outcomes can be actively pursued[3].

Other possible benefits of MSC for learning initiative evaluations is that decision makers are provided with a more realistic understanding of the initiative and the difference that it makes. Obtaining a more realistic understanding in the form of learner success stories promotes recognition of the diverse values among stakeholders, and identifies unintended outcomes.

The MSC approach has three parts. The first is to establish domains of change that should be evaluated – expected outcomes may come from strategic planning document, program logics or evaluation rubrics. The second part of MSC involves collecting and reviewing stories of change – stories of significant change are collected from people involved in the learning initiative using a simple question. For example, ‘*During the past month, in your opinion, what is the most significant change that has taken place as a result of the program*?’ Participants are also encouraged to report why they consider a particular change to be the most significant one. Next the stories are analysed and passed on to designated stakeholders (i.e. learning and development program designers, decision makers and implementation staff) who review the stories and determine which ones represent the most significant change against the domains of change. Making the results of the review available to others then provides a further source of feedback to subsequent review. An elimination process occurs when a selection of the stories that represent the most significant stories of change are passed up through organisational levels of authority. All of the stories selected in the final elimination round can be collated into a document, together with the reasons they were selected.

If required, verification of the accuracy of the stories can be undertaken. Evaluators can gather additional information about the changes reported or gather evidence of ongoing change. The process of the MSC itself can be at least as important as the evaluation data contained in the stories as it increases involvement and interest in evaluation, and improves understanding of program outcomes and the dynamics that influence them.

---

[3] Dart. J. (2005) Evaluation Encyclopaedia. p. 261-262.

Fact sheet 12: Transparency of judgements in evaluation

The example rubric introduced in Fact Sheet 8 is presented again as an example of how evaluative judgements of performance can be displayed. This display makes transparent that the evidence supports rating of a combination of clearly not effective, mixed effectiveness and clearly effective.

Following a discussion of these evaluative assessments, recommendations can be made on how to address the short-comings that have been uncovered.

Rating scale example for displaying evaluative judgement – Effectiveness criterion

RATING		WORK		PEOPLE		COURSES		RESOURCES
Clearly effective - a consistent story showing that the learning initiative method is effective in delivering the learning and development needed		Evidence that most APS people deliberately developed relevant knowledge and skills through on the job experience, taking on an increased span of control, and undertaking reflection	✓	Evidence that most APS people deliberately developed relevant knowledge and skills through connecting with other people		Evidence that most APS people deliberately developed relevant knowledge and skills through participating in online courses, MOOCs and webinars, and/or attending classroom training		Evidence that most APS people deliberately developed relevant knowledge and skills through searching the internet and intranet for resources and information; reading books, scholarly articles and magazines; and accessed online podcasts, videos, articles and blogposts
Mixed effectiveness - in some respects the learning initiative is effective in delivering the learning and development needed, but in other significant respects, it was not effective		Evidence that the learning initiative is effective, but in other significant effects or for a significant proportion of APS people, the work method was not effective in delivering the learning and development needed.		Evidence that the learning initiative is effective, but in other significant effects or for a significant proportion of APS people, the people method was not effective in delivering the learning and development needed.	✓	Evidence that the learning initiative is effective, but in other significant effects or for a significant proportion of APS people, the courses method was not effective in delivering the learning and development needed.	✓	Evidence that the learning initiative is effective, but in other significant effects or for a significant proportion of APS people, the resources method was not effective in delivering the learning and development needed.
Clearly not effective - the learning initiative was not effective in delivering the learning and development needed	✓	Evidence that the work method was not effective in supporting the learning and development needs of most APS people		Evidence that the people method was not effective in supporting the learning and development needs of most APS people		Evidence that the courses method was not effective in supporting the learning and development needs of most APS people		Evidence that the work method was not effective in supporting the learning and development needs of most APS people

Fact sheet 13: Good practice in developing preliminary recommendations

Approaches for formulating recommendations can involve individual critical reflection or group critical reflection with the practitioners who will implement the recommendations. The formative evaluation approach involves the evaluator first developing preliminary recommendations from evaluation findings before considering them further and they are finalised with workplace colleagues. When developing conclusions into recommendations for more prescriptive steps forward:

1. Take care and make it clear how recommendations were arrived at. They should be based on evidence and logical reasoning that is useful, feasible and carefully worded because they may attract considerable scrutiny.
2. Respond to performance issues such as whether the initiative should continue or be repeated in the future, how the risk of failure can be reduced, how to correct negative outcomes and short-comings/ less than expected outcomes, and how the initiative can be improved.
3. Be explicit about the type of recommendations being made. Recommendations can either:
    a. Enhance or extend current activities such as making the initiative available to more people, calling for more or better resources, and suggest that something should be done to remediate a deficiency that has been uncovered.
    b. Specify strategic or tactical changes to the theory of change such as eliminating, strengthening, augmenting or replacing current initiative activities to increase the likelihood that in the future, the initiative will operate with fewer problems and outcomes will be brought into line with the original intentions. These type of recommendations engage with redesign issues.
    c. Specify changes in objectives to better reflect changes to agency objectives.

Fact sheet 14: Report writing

Tips for speedy report writing

Experienced Learning and Development practitioners expedite the evaluation reporting process by following good report writing practice:

- Don't delay - begin report writing as soon as possible in the evaluation process.
- Maintain a detailed audit trail explaining the evaluation processes and the changes made along the way.
- Undertake simultaneous data collection and analyses as much as possible in the lead up to the reporting phase. The different strands of data collection and analyses can be undertaken in parallel prior to data synthesis with preliminary findings being used to guide decisions about additional data collection. An approach referred to as 'Rapid evaluation'[4] involves starting data analysis as the data is coming in and while other data is still being collected. Data collection and analyses happens simultaneously with the evaluation team meeting almost daily to debrief about the emerging themes and issues. This cycle is repeated until new data cease to provide new information of significance to the evaluation questions.
- Draw up the report structure before starting to write the report. A standard reporting structure can be followed. The following checklist suggests a reliable reporting structure.

[4] Rapid Evaluation and Assessment Method (PDF) Methods of Rapid Evaluation, Assessment, and Appraisal (researchgate.net)

Checklist for Evaluation Report Structure[5]

Title page	
☐	Title: Provide a succinct, informative name for the report. Include the initiative name, the word 'evaluation'; and report timing, such as annual, midterm, or final report.
☐	Recipient(s): Identify the name, title, agency, and contact information of the individual(s) to whom the report is being submitted.
☐	Author(s): Identify the name, title, organization, and contact information of the individual(s) who wrote the report. (If the person submitting the report is different from the author, identify that person separately.)
☐	Date: Identify the month and year when the report was complete.
☐	Preferred citation: Provide complete reference information so that others may cite the report. Include the author, year, title, and web address, if available.
☐	Acknowledgements: The acknowledgements section identifies and thanks individuals who directly or indirectly assisted or facilitated the evaluation process.
☐	Contributors: Identify each contributor by name. If desired, identify their specific contributions.
Acknowledgements - The acknowledgements section identifies and thanks individuals who directly or indirectly assisted or facilitated the evaluation process.	
☐	Contributors: Identify each contributor by name. If desired, identify their specific contributions.
Table of contents	
☐	Headings: List all first - and second-level headings, including the titles of all documents in the appendices.
☐	Page numbers: Identify the page numbers on which each of these components belong.
List of tables and figures	
☐	Titles: List the exact titles of all tables and figures.
☐	Page numbers: Identify the page numbers on which each table and figure begins.
List of acronyms	
☐	Definitions: List acronyms alphabetically and identify the terms they represent.
Executive Summary - The executive summary is a synopsis of key information from the main report. It usually includes important findings, conclusions, and recommendations. The executive summary tends to be the most widely read part of a report, so should make sense when read separately from the main report.	
☐	Most important (need to know) content: Highlight key content from the report, based on the needs of the report's main audiences.
Introduction - The Introduction orients the reader to the type of information included in the report.	
☐	Overview: Identify the program that was evaluated and what the report is about.
☐	Structure: Describe how the report's content is organized.
☐	Intended audience: Identify the groups or individuals for whom the report was developed.
☐	Purpose and intended use: Briefly note why the evaluation was conducted and how the results are intended to be used.

[5] This checklist has been drawn from Kelly N. Robertson and Lori A. Wingate. The Evaluation Centre, Western Michigan University. https://wmich.edu/evaluation/checklists

Initiative Description - The initiative description section includes details about the program that was evaluated to help readers understand the context of the evaluation's implementation and results.	
☐	Goals and/or objectives: Identify the specific achievements the program is designed to bring about.
☐	Funding arrangements: Identify the entities that sponsor the initiative and the total program budget.
☐	Organizations involved: Identify organisations involved in the program and their roles.
☐	Identify the groups or types of people the initiative is designed to serve.
☐	Program design: Describe the program's activities and how they are supposed to bring about desired changes. If the program has a logic model or theory of change, include it here.
☐	Context: Describe relevant economic, political, environmental, cultural, social, or other important factors that influence the conditions in which the program operates.
☐	History: Identify the program's stage of maturity, such as whether it is a new initiative, has been operating for a long time, or is winding down for closure. Describe how the program has changed over time.
Evaluation background - Key factors that influenced the evaluation's planning and implementation. This section helps readers understand the general orientation of the evaluation and the opportunities.	
☐	Purpose and intended use: Identify why the evaluation was conducted. Describe how the results are intended to be used, such as to inform program improvement.
☐	Scope: Identify the boundaries of the evaluation in terms of time period, location, and the specific program components that were evaluated.
☐	Stakeholder engagement: Describe how stakeholders were involved in and influenced the evaluation's planning and implementation—beyond serving as data sources.
☐	Budget: Identify the total funding for the evaluation and the percentage of the overall program budget it constituted.
☐	Evaluation team: Briefly describe the composition of the evaluation team and each member's role. Disclose any real or perceived conflicts of interest—relationships or factors that could affect the credibility of the evaluation—and describe how they were managed.
☐	Prior evaluation: If the program has been evaluated before, summarize key takeaways and implications for the current evaluation.

Methods - How the evaluation was implemented and how the evaluation results were obtained. Organise this section so that it is clear which indicators, data sources, and methods were used to address each evaluation question.	
☐	Approach: Briefly describe how the Handbook and other evaluation resources informed the evaluation's focus, design, or implementation.
☐	Evaluation questions: Identify the questions that framed the evaluation and criteria chosen to judge the programs performance.
☐	Indicators: Identify what was measured for each evaluation question or criterion.
☐	Data sources: For each indicator, identify the type and source of information collected— such as individuals, documents, or institutional databases.
☐	Sample size and description: If sampling was employed, describe how many individual data sources were selected for inclusion in the sample and the actual number from which data were gathered. For each data source, describe how individual cases were chosen—such as through a census or specific sampling techniques.
☐	Data collection methods: Describe how the information was gathered from each data source—such as through interviews, surveys, focus groups, observations, or document review. If mixed methods were used, describe the extent to which and how qualitative and quantitative approaches were integrated.
☐	Data collection procedures: Include pertinent procedural information, such as how respondents were invited or encouraged to participate in data collection. Instruments: Identify the tools used to implement each data collection method, such as questionnaires and protocols for interviews, document reviews, focus groups, or observations. Include copies of instruments in appendices if possible. Otherwise, provide a brief description of each instrument.
☐	Instruments: Identify the tools used to implement each data collection method, such as questionnaires and protocols for interviews, document reviews, focus groups, or observations. Include copies of instruments in appendices if possible. Otherwise, provide a brief description of each instrument.
☐	Timeline: Identify when each method was implemented and when major evaluation tasks were completed.
☐	Data analysis: Describe the specific procedures used to organize and transform raw data into findings. Include enough detail so that others could reproduce the analysis for both qualitative and quantitative data. Indicate whether and how multiple data sources or methods were used to measure the same thing.
☐	Interpretation: Describe how findings were used to answer the evaluation questions and reach conclusions about the program's quality, value, or importance. Identify who was involved in that process. Include enough detail so that others could reproduce the process and arrive at similar conclusions.
☐	Limitations: Describe factors that may have adversely affected the accuracy or credibility of the evaluation results. This should include significant limitations that were within or outside of the evaluation team's control.
Evaluation Results - What was learned? This section will likely be the longest part of the report, as it includes the most important and substantive information. Organize results by the evaluation questions.	
☐	Findings: Present the analysed data and other evidence used to formulate the conclusions. Provide relevant information about the representativeness of the data, such as response rates or data source characteristics.
☐	Conclusions: Answers to the evaluation questions. To enhance transparency, remind the reader of the relevant findings and interpretation procedures used to reach conclusions.
Recommendations - Suggestions for actions that align with intended evaluation uses.	
☐	Development process: Explain how the recommendations were generated.
☐	Recommendations for the initiative: Identify suggested actions for stakeholders to consider. Refer to the specific evaluation results to support each recommendation.
☐	Recommendations for future evaluations, if relevant.
References - Information about literature cited in the report, enabling readers to locate sources if desired.	
☐	Sources: Use a consistent reference style. Provide website addresses for publicly accessible documents.

Appendices - Supplementary information that is pertinent to the evaluation, but not critical to readers' understanding of the report, can be included as appendices. Each document included as an appendix should be referenced in the body of the report.	
☐	Data collection materials: Include data collection instruments and protocols, qualitative coding guides, and blank consent forms.
☐	List of reviewed documents or artefacts: List all reviewed artefacts, databases, documents, or other materials, if they were not already mentioned in the methods section. If possible, format the list using the same style used for references. If the information is publicly available, include website addresses or indicate how others can access the materials.
☐	Supplementary data or findings: If applicable, include additional data tables that may be of interest to some readers but are not required for understanding the evaluation conclusions.

Fact sheet 15: Evaluation facilitation

When working with internal stakeholders, the evaluation facilitator must focus on collaborative sense-making and prioritisation, and may need to overcome deeply ingrained biases.

Collaborative sense-making involves scrutinising evaluative findings with an open mind and raises the following questions:	**Prioritise action-taking** involves giving consideration to possible options in light of the evidence of what makes a positive difference for people in this agency and formulating recommendations on the path forward. The following questions can be useful:
• What has evaluation inquiry revealed? • What does evaluative reasoning suggest about how close we are to what 'good' looks like? • Do we have different interpretations of the data? If so, why? • How do we feel about what has been found?	• What do we need to do and why? • What options do we have? • What can we do to ensure better progress and outcomes for more people? • How big are the changes being considered? • Can the outcomes we want be achieved within the specified timeframes? • Is there sufficient capability to forge ahead with the necessary changes? • What other strengths can be drawn on? • What support/resources might be needed?

Deeply ingrained biases that interfere with organisational insights[6]

Four biases that are inherent in organisations **must be overcome** by the evaluation facilitator are discussed.

1) Preoccupation with success that limits learning

While we might acknowledge that learning can come from disappointing results and failure, in reality there is often a preoccupation with success at the expense of learning. This tendency may not be surprising but becomes unreasonable when avoiding questioning assumptions starts to impede learning, adaptation and improvement. A bias towards success raises several challenges the facilitator must manage to create the potential for improvement:

- Counter a fear of failure through a disciplined approach to evaluative inquiry and reflective practice that is premised on an openness to not always getting it right and to looking for opportunities to adapt and improve.
- Embrace a 'growth mindset' where no matter how good something is, it can always be better and challenges are seen as learning opportunities.

[6] Draws on Francesca Gino and Bradley Staats (2015) *'Why Organizations Don't Learn, Our traditional obsessions—success, taking action, fitting in, and relying on experts—undermine continuous improvement.* Harvard Business Review. https://hbr.org/2015/11/why-organizations-dont-learn

- Emphasise that key evaluation question’s focus on things that matter most.
- Overcome ‘attribution bias’ whereby bad luck, unforeseen changes or contextual conditions are blamed for getting in the way of success but credit is given to competence, skill and diligence for effectiveness and success. Evaluative thinking prompts us to dig deeper to better understand causality and learn from work experience.

2) Bias toward action that curtails reflective practice

Experience alone does not necessarily lead to learning. Setting aside time for deliberate reflection on experience is essential. However, an aversion to ‘inaction’, especially when under chronic time pressure, is detrimental to adaptation and improvement for two reasons. First, exhausted people can be too tired to learn new things or to apply what they already know. Second, being always busy doesn’t allow the time needed for reflecting on what worked well and what didn’t. Reflective practice involves taking a critical stance towards one’s own practice and that of the groups as part of a process of continuous learning, adaptation and improvement. Reflective practice provides an opportunity to recognise assumptions, frameworks and patterns of thought and behaviours that shape thinking and action.

An advantage of reflective practice within the context of formative evaluation is that the reflection is built upon accurate data and engagement with multiple perspectives. Taking time to reflect on changes that can be attributed to the focus of the evaluation and the drivers of success, or barriers that limit it, is a critical part of the formative evaluation process.

3) Bias towards fitting in that limits innovative evaluative thinking

Generally, people are inclined to want to fit in with business-as-usual ways of working together and maintaining the status quo. Despite the comfort that comes with fitting in, some codes of behaviour stifle opportunities for learning through work and innovation. Formative evaluation provides a rare, formal opportunity and a systematic liberating structure to review issues that may otherwise remain out-of-sight and allow new ideas to flourish.

4) Tendency to rely on external experts

An unexpected consequence of some prevailing management approaches to examining how organisations operate is the perception that external industry experts are often the best source of ideas for improvement. Internally conducted formative evaluation help to counteract an over-reliance on external experts that can challenge continuous learning.

Fact sheet 16: Checklist to review the evaluation against its objectives

The objectives that you initially set for your evaluation were informed by what a good evaluation should do (Refer to Fact Sheet 5). The intention was to comply with the principles-based approach of the Commonwealth Evaluation Policy, and the APS Learning Evaluation Framework. Now the review of the evaluation processes and its outcomes can be made against these original objectives.

Matching the review to the initial objective, this desktop exercise should seek to:

- Describe how the evaluation contributes to agency outcomes to enhance the value, influence and use of conducting evaluation in APS learning and development; and
- Generate lessons for future evaluations of learning initiatives through assessing what parts of the evaluation went well, what didn't and the overall success of the evaluation.

Purposeful and fit for purpose	
☐	The evaluation was tailored to the specific agency circumstances and improvement needs of learning and development.
☐	The rationale for the evaluation, its scope, evaluation questions and how the findings were made clear, decided upon early and feasible.
☐	An intentional and purposeful planning process culminated in articulation of the Terms of Reference (TORs).
☐	Attention was devoted to the full range of individuals and groups invested in the initiative being evaluated.
☐	The evaluation was completed in a timely manner so as to ensure the usefulness of the findings and recommendations.
Useful	
☐	The evaluation process and report took full account of the diverse attributes of APS people, their agency settings and the adequacy of workplace resources
☐	The evaluation acknowledged the diversity of learning initiatives objectives.
☐	The evaluation revealed the assumptions the learning initiative operated under and promoted evidence-informed, or at least plausible, assumptions of causality and how change happens.
☐	The evaluation built evidence about what has made a difference, and its reporting format and dissemination strategy are appropriate and accessible.
☐	The evaluation conveys the process of how the initiative causes change and impacts on agency outcomes.
☐	The evaluation provided useful information provided about what makes a positive difference for organisational insights and decision making.
☐	The lessons learned about conducting evaluation and how the process can be improved have been documented.
☐	The evaluation process and report took full account of the diverse attributes of APS people, their agency settings and the adequacy of workplace resources
Ethical and culturally appropriate	
☐	The evaluation has not been ignored – actionable recommendations (i.e. instrumental use) to enhance adaptability, process and/or conceptual use (i.e. changes in understanding) emerged from this evaluation.
☐	The evaluation manifested professional integrity, respected the rights of individuals and was sensitive to the setting.
☐	Demonstrated integrity – Pressure was not exerted to suppress, distort or alter findings or recommendations.
☐	Transparent arrangements for publishing of evaluation reports were followed.

Credible	
☐	Appropriate data was retrieved and/or collected to answer the key evaluative questions and enhance understanding about how well the initiative objectives contributes to meeting agency purpose/s and what needs to happen to improve effectiveness.
☐	Evaluation questions were decided upon early on, and guided the structure of the report.
☐	Evaluation questions informed data collection and analysis methodologies, and appropriate integration of monitoring and evaluation.
☐	Full account was taken of the agency settings and the adequacy of workplace resources.
☐	Performance stories described both positive outcomes where initiatives are making a difference and situations where they didn’t thereby providing opportunities for improvement.
Transparent	
☐	The original underlying program logic was made clear and an assessment on whether it needs to be revised was made, the Criteria of Merit were defined in unambiguous terms and applied in a transparent manner.
☐	Stakeholder participation was meaningful.
☐	Evaluation findings were transparent (unless there are good public policy reasons why they should not be released publicly).