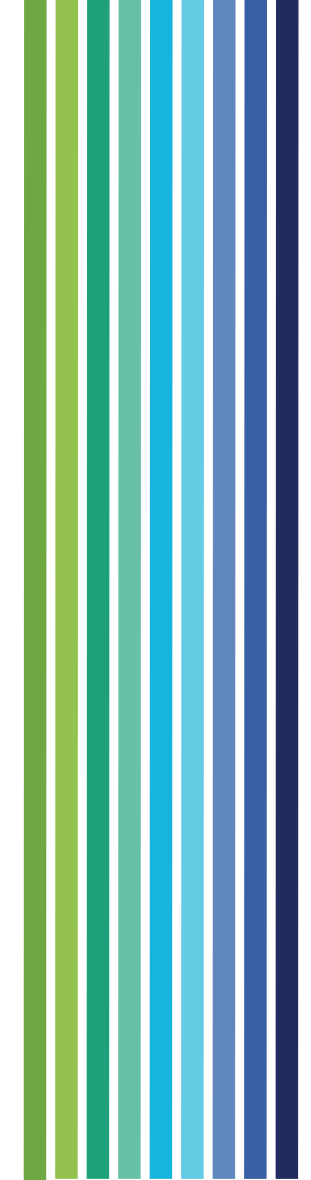


# EVALUATION BUILDING BLOCKS

*A GUIDE*





## INTRODUCTION

Tēnā koutou katoa. Welcome, we created this guide to present our approach to evaluation, developed through ten years of collaborative work. There are many other guides to evaluation and different ways to do it. This approach emphasises the place of evaluative reasoning in evaluation. It draws on the work of many evaluation theorists, as well as our practice-based body of knowledge.

We find that using the building blocks explained in the guide helps us to do credible and useful evaluation. Our clients tell us it gives them meaningful and insightful findings that they can use to take action.

We acknowledge Waikato Regional Council for providing the impetus and support to write this guide.

Noho ora mai, Kate McKegg, Judy Oakden, Nan Wehipeihana, Julian King



**EVALUATION BUILDING BLOCKS**

## SET UP THE EVALUATION FOR SUCCESS



### QUESTIONS TO ASK

- What is evaluation?
- Why bother with evaluation?
- Who are we doing this for?
- What's the purpose of our evaluation?
- Who will use the evaluation?
- What will the consequences of the evaluation be?
- What is it that we are evaluating?
- What is the strategy, project or programme trying to achieve?
- How is it going to achieve this?
- What's going on in the current situation?

### GUIDANCE

- It's important for everyone to have a shared understanding from the outset of an evaluation. For a start, this means making sure that everyone understands what evaluation is. For example, there are some similarities with processes like monitoring, review, audit and research.
- People often think evaluation is about measuring outcomes or seeing whether objectives have been met. Certainly these may be part of evaluation, but on their own they are not enough to serve as an evaluation.
- **Evaluation is the systematic determination of the quality, value or importance of something.**<sup>1</sup> Evaluation always focuses on quality, value and importance. And evaluation matters because it is useful to:
  - check the quality of what we are doing
  - know what difference we are making
  - find out what is and isn't working
  - ensure we channel limited energy and resources to what matters
  - support insightful planning, implementation, delivery, improvement and development
  - be able to demonstrate the quality, value and importance of the investment in the strategy, project or programme to those who have an interest or stake
  - support project and initiative sustainability.
- For an evaluation to be worthwhile it needs to ensure the specific needs of the key stakeholders – particularly the intended primary users – are taken into account as well as to the situation and context of the initiative. Utilisation-focused evaluation is “evaluation done for and with specific, intended primary users for specific, intended uses”.<sup>2</sup>
- It is also important at the start of an evaluation for different stakeholders to understand each other's views, principles and values.
- Thinking about how the findings of the evaluation will be used before starting work on it will help focus the evaluation and maximise the influence it might have. In turn, the way people engage in and use the evaluation is affected by the way the evaluation has been designed and carried out.
- Good evaluation commonly sets out by describing the thing evaluated, what it's trying to achieve and how it will do this. This is sometimes presented in a theory of change, a logic model or results chain. There are many tools and approaches to this. Done well, it helps everyone involved or affected gain a shared understanding of what they are trying to achieve and how this is expected to happen. This shared understanding also comes in handy later in the evaluation process when we are untangling what effects are due to the initiative or what may be due to other factors.<sup>3</sup>
- Clarifying the primary purpose of the evaluation helps guide many decisions as the evaluation progresses. The evaluation purpose may variously be development, improvement, capacity building, learning, accountability, or knowledge and understanding.

<sup>1</sup> ANZEA and SuPERU (2015) Evaluation Standards for Aotearoa New Zealand. Scriven, M. (2007). The logic of evaluation. In H.V. Hansen, et. al. (Eds), *Dissensus and the Search for Common Ground*, CD-ROM (pp. 1-16). Windsor, ON: OSSA.

<sup>2</sup> Patton, M. Q. (2008). *Utilization-Focused Evaluation*, Sage, p37.

<sup>3</sup> Rogers, Patricia (2014). *Theory of Change. Methodological Briefs: Impact Evaluation 2*. Florence: UNICEF Office of Research.

## DEVELOP KEY EVALUATION QUESTIONS



### QUESTIONS TO ASK

- What do we want to know about quality, value and importance in relation to this strategy, project or programme?
- What do we want to learn?
- How might we apply the findings to future decisions about the current initiative? Will we need to use or apply the findings to another initiative?

### GUIDANCE

- Key users need credible, well-justified findings for making decisions, developing greater understanding and gaining insights for action.
- **Evaluation questions** provide both a focus and a frame for an evaluation. They are critical for producing robust conclusions. They need to be carefully thought about and prioritised. This will ensure the evaluation ends up meeting the information needs of key users and is able to have any other possible or intended wider influence or consequence.
- **Good evaluation questions are explicitly evaluative**, that is, they seek answers about the quality and value of things.
- For example: a **descriptive question** about implementation might ask, “Was the programme implemented as intended?” In contrast, **an evaluation question** would ask how well it was implemented – for example, “Was it implemented thoroughly, fairly, ethically, culturally appropriately, efficiently, professionally?”<sup>4</sup>
- Evaluation questions may cover the following areas:
  - Need and relevance
  - Quality of content, design, and delivery or implementation
  - The extent to which the strategy, programme or project contributes to change
  - The value of results and outcomes
  - Learnings about specific aspects such as barriers and enablers
  - Overarching value or worth
  - Considerations about value in other settings, ongoing viability, and other opportunities and threats.<sup>5</sup>
- Ideally, it's best if there are only a few high-level evaluation questions with more detailed questions sitting beneath these. There are a set of core high-level evaluation questions we often use that can be applied to any strategy, programme or project. These include the following:
  1. What is this strategy, programme or project? Why was it needed, and who is it for? What resources have been used to create, implement or maintain it?
  2. How well was the strategy, programme or project implemented and/or delivered?
  3. How effectively did the strategy/programme/project realise key results and outcomes?
  4. How valuable are the outcomes for participants, to the organization, the community, the economy?
  5. Overall how worthwhile is or was the strategy, programme or project?
  6. What are the learnings that can be applied going forward?

Often, there is an inbuilt assumption that an intervention is going to cause change, as if in a vacuum. There is also often an assumption the intervention will work the same way for everybody. In our experience, real life is more complex than that. Most strategies, programmes and projects, contribute to change in complex ways, if they work. Also, most don't work the same for everybody.

<sup>4</sup> E. Jane Davidson (2012) *Actionable Evaluation Basics: Getting succinct answers to the most important questions*. A minibook. Real Evaluation Ltd.

<sup>5</sup> Davidson, E.J. (2005) *Evaluation Methodology Basics, The Nuts and Bolts of Sound Evaluation*. Sage, Thousand Oaks, California.



## AGREE ON PERFORMANCE CRITERIA AND DETERMINE IMPORTANCE

### QUESTIONS TO ASK

- What are the needs we are trying to address?
- What's valued in this context?
- What's important to focus on?
- How important is each of the criteria in relation to the other criteria?

### GUIDANCE

- One of the biggest challenges in evaluation is reaching judgements about quality and value in a way that key stakeholders perceive as transparent, fair and credible.
- To do this, you need to work with them to define what aspects of quality and value matter. This step is about developing a suite of selected **success or performance criteria**. These will be used to look at the strategy, programme or project and assess how good or valuable it is or was. These criteria are not the same as specific objectives, goals or indicators, although there may be some overlap.
- It's preferable that these criteria are informed by a systematic understanding of the needs that a strategy, programme or project is trying to address. The needs may include effectiveness, efficiency, economic impact, equity, fairness, relevance, justice or sustainability. Other criteria may come from legislative, organisational requirements, or from cultural and ethical requirements and needs.
- A complete suite of success or performance criteria should include attributes for the design, content, implementation, delivery, results or outcomes, impact, value for money and potential exportability of a strategy, programme or project.<sup>6</sup> This is a comprehensive list and you are encouraged to identify within this the areas of key priority for the evaluation.
- It's good practice that the criteria to be used in any evaluation are developed in such a way that they have widespread acceptance, credibility and buy-in by key stakeholders. We've found that when key stakeholders can see what value the criteria represent, they are more likely to engage in the evaluation process.<sup>7</sup>
- Every evaluation context is different, so **the process for selecting success and performance criteria for evaluation needs to reflect the unique circumstances and conditions of each situation**. This will emerge out of a mixture of "people, politics, history, context, resources, constraints, values, needs, interests, and chance".<sup>8</sup>
- As indicated, it is critical at this stage to **meaningfully engage and involve key stakeholders in the selection of the evaluation criteria**. This makes sense on very practical as well as ethical grounds. Authentic engagement at this stage ensures the success or performance criteria reflect what is valued and considered important by those involved in or affected by a strategy, programme or project. It anchors the evaluation process and improves the chances of findings being used and having an influence. This has an ethical dimension. It enhances the possibility that the evaluation will actually contribute to decisions for changes and improvements that make a difference in people's lives, and ultimately to our environment, communities and economy.

<sup>6</sup> Davidson, E.J. (2005) *Evaluation Methodology Basics, The Nuts and Bolts of Sound Evaluation*. Sage, Thousand Oaks, California.

<sup>7</sup> King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013) Evaluative Rubrics: a Method for Surfacing Values and Improving the Credibility of Evaluation, *Journal of Multidisciplinary Evaluation*, 9 (21)

<sup>8</sup> Kusters, C., van Vugt, S., Wigboldus, S., & Williams, B. (2011) *Making evaluations matter: A practical guide for evaluators*. Centre for Development Innovation, Wageningen University & Research centre, Wageningen, The Netherlands, p 1:20.

Continues on next page →

# 3

## AGREE ON PERFORMANCE CRITERIA AND DETERMINE IMPORTANCE



### GUIDANCE

- In complex situations, a set of more **high-level success or performance criteria** can be helpful as they provide some coherence to the evaluation process. There are a range of such criteria regularly used in public sector contexts, such as relevance, efficiency, effectiveness, impact, sustainability. It's important, when using these high-level criteria, to make sure they are sufficiently fleshed out so that their meaning is widely understood and shared by key stakeholders for the particular evaluation situation. For example, the meaning of “relevance” or “sustainability” in one context may be quite different from another.
- The next step is to **determine the relative importance of these criteria**. The criteria will not be equally important – and knowing which are more important is essential for being able to make effective, credible judgements about overall results, effectiveness, quality and value. There are a number of ways to determine the relative importance of criteria, for example: key stakeholders could vote, stakeholders with specialist expertise could be asked to give their view, or evidence from the literature could be used. All have their strengths and weaknesses including the levels of skill and expertise required, time and cost, range of perspectives.

# 4



## DEVELOP A FRAMEWORK TO ASSESS PERFORMANCE

### QUESTIONS TO ASK

- What does quality and performance look like in this context?
- What does good quality and performance look like?
- What does poor quality and performance look like?
- What phase are we in and so what performance can we reasonably expect at this stage of development?

### GUIDANCE

- Once success and performance criteria have been agreed to, the next step is to agree on or develop a **performance framework** for assessing the strategy, programme or project, against the criteria. The framework spells out what “how good” or “how well” or “how much” mean in practice.
- There are a few kinds of performance frameworks commonly used in evaluation practice. These are commonly called rubrics.<sup>9</sup> The most common kind is a descriptive rating scale that can be applied to any of the criteria, or to clusters of criteria that logically group together. Below is an example of a generic descriptive rating scale commonly used in evaluation.

Rating Level	Description
<b>Excellent (Always)</b>	Clear example of exemplary performance or very good practice in this domain: no weaknesses.
<b>Very good (Almost always)</b>	Very good to excellent performance on virtually all aspects; while strong overall not exemplary; no weaknesses of any real consequence.
<b>Good (Mostly, with some exceptions)</b>	Reasonably good performance overall; might have a few slight weaknesses, but nothing serious.
<b>Emerging (Sometimes, with quite a few exceptions)</b>	Fair performance, some serious, but non-fatal weaknesses on a few aspects.
<b>Not yet emerging (Barely or not at all)</b>	No clear evidence has yet emerged that the aspect of performance has taken effect.
<b>Poor (Never, or occasionally with clear weaknesses evident)</b>	Clear evidence of unsatisfactory functioning; serious weaknesses across the board or on crucial aspects.

Source: adapted from Oakden, J. & McKegg, K. (2011, August). *Waste Minimisation Act implementation: evaluation of stakeholder perspectives*. Wellington: Ministry for the Environment. (p19)

<sup>9</sup> King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013) Evaluative Rubrics: a Method for Surfacing Values and Improving the Credibility of Evaluation, *Journal of Multidisciplinary Evaluation*, 9 (21)

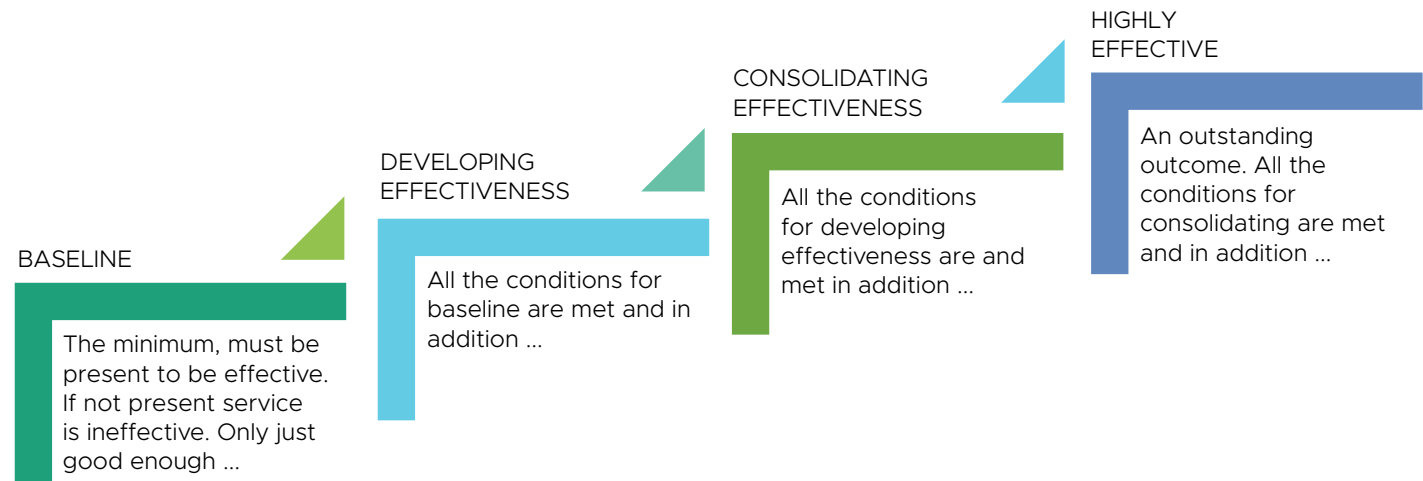
Continues on next page →





## GUIDANCE

An example of a more general rating scale which implies some form of development over time is provided below:



Sourced from [http://www.tokatu.org.nz/index.php?reports\\_publications](http://www.tokatu.org.nz/index.php?reports_publications)

Tokatu. (2014). *How good is good, rubric in action. A Workbook for Evidencing the Quality of Peer Support in Aotearoa New Zealand*. Tokatu Project, Kites Trust. Wellington

Continues on next page →



## GUIDANCE

- A second kind of performance framework has a description of each level of performance for each of the criteria. This is known as an analytic rubric.

The example below illustrates this kind:

Breakfast in Bed: Analytic Rubric					
	Beginning 1	Developing 2	Accomplished 3	Exemplary 4	Score
<b>Food</b>	Most food is colder or warmer than it should be, is under- or over-seasoned, or is under- or overcooked.	Some food is colder or warmer than it should be, is under- or over-seasoned, or is under- or overcooked.	All food is at the correct temperature, adequately seasoned, and cooked to the eater's preference.	All food is perfectly cooked and seasoned to the eater's preference. Additional condiments are offered.	
<b>Presentation</b>	More than one item (tray, napkin, or silverware) are dirty or missing.	Tray, napkin or silverware may be dirty or missing.	Food is served on a clean tray with napkin and silverware. Some decorative additions may be present.	Food is served on a clean tray with napkin and silverware. Several decorative touches are added.	
<b>Comfort</b>	Wake-up is abrupt, little to no help with seating, and the recipient is rushed and crowded during the meal.	Wake-up is somewhat abrupt, recipient may struggle with seat adjustment, or there may be some rushing or crowding during eating.	Recipient is woken gently, assisted in seat adjustment, and given reasonable time and space to eat.	Recipient is woken gently and lovingly, assisted until seating is just right, and given abundant time and space to eat.	

Sourced on 29 August 2016 from <http://www.cultofpedagogy.com/holistic-analytic-single-point-rubrics/>. Posted on May 1, 2014 by Jennifer Gonzalez.

Continues on next page →



## GUIDANCE

Another example of this kind of rubric (applied to evaluative leadership in an organisation):

### EXAMPLE: Evaluation leadership dimension

#### EMERGING:

There is little to no encouragement or support available for evaluation. Any monitoring, evaluation or performance development is generated by individuals.

#### DEVELOPING:

Leadership recognises the need to comply and ensures that this is achieved.

#### CONSOLIDATING:

Leadership sets a direction for evaluation and performance development and strongly encourages stakeholder to participate.

#### HIGHLY DEVELOPED:

Leadership shares a clear vision for performance, results, improvement and development; demonstrates a commitment to learning and models an insatiable curiosity to improve.

A third type of performance framework uses all the evaluation criteria with thresholds set for each level of performance. This is called a holistic rubric.

*Continues on next page →*



## GUIDANCE

The examples below are from two very different evaluations:

### Māori and Pacific Education Initiative – evaluation rubric<sup>10</sup>

Highly effective	<p>All of the conditions for developing effectiveness are met and, in addition:</p> <ul style="list-style-type: none"> <li>• The vast majority of projects show educational outcomes at least as positive as those achieved by pre-existing exemplar Māori and Pasifika education programmes.</li> <li>• Virtually all projects achieve outcomes that their families and communities widely value. Māori and Pasifika communities endorse and celebrate the success of these models for the contribution they make to them realising their educational aspirations.</li> <li>• There is clear evidence about why and how the models work for their target populations, including validation of the role of culture and the specific Māori and Pasifika cultural elements that matter in this context.</li> <li>• The Foundation is a recognised and respected leader, innovator and influencer of education policy and/or philanthropy.</li> <li>• Government/communities (for example, schools, iwi, and others) implement successful models pioneered through the Māori and Pacific Education Initiative.</li> <li>• Other philanthropic organisations and/or Government(s) recognise the value of the high-engagement approach, and seek to learn from Foundation North.</li> <li>• The Initiative influences the focus of education in New Zealand, for example, how value and success in education are defined.</li> </ul>
Consolidating effectiveness	<p>All of the conditions for developing effectiveness and any of the conditions for highly effective are met.</p>
Developing effectiveness	<p>All of the conditions for minimally effective are met and, in addition:</p> <ul style="list-style-type: none"> <li>• Government (for example, Ministers or departments) or communities (including, schools and iwi) show an interest in the models, or the Initiative enables the Foundation to engage in other significant policy dialogue that otherwise would not have been possible.</li> <li>• Learnings from the high-engagement investment are identified and acted upon. There is evidence of ongoing refinement and improvement of the funding approach.</li> </ul>
Minimally effective (basic requirements; “only just good enough”).	<p>All of the following conditions are met:</p> <ul style="list-style-type: none"> <li>• The majority of projects show better educational outcomes than previously achieved with Māori and Pasifika children of equivalent year groups within the communities served by the projects.</li> <li>• The majority of projects achieve outcomes that their families and communities value.</li> <li>• The overall outcomes achieved through the Initiative investment (such as, educational outcomes, associated social and economic benefits or other benefits of value to the Foundation) are commensurate with the overall level of investment.</li> <li>• There is a clear rationale to support why and how each of the models are intended to work – including the specific Māori and Pasifika cultural elements that matter in this context.</li> <li>• At least one project was able to secure external sustainable funding.</li> </ul>
Ineffective	<p>Any of the conditions for minimally effective are not met.</p>

<sup>10</sup> Kinnect Group & Foundation North. (2016). *Kua Ea Te Whakangao Māori & Pacific Education Initiative: Value for investment evaluation report*. Auckland: Foundation North.



## GUIDANCE

### Sustainable Farming Fund – value for money rubric<sup>11</sup>

Level of performance	Sustainable Farming Fund Value for Money Rubric
Excellent value for money	<ul style="list-style-type: none"> <li>• Sufficient results from successful projects to provide clear evidence of positive return on investment from SFF investment overall AND</li> <li>• Credible contribution to export opportunities and improved sector productivity; and increased environmentally sustainable practice AND</li> <li>• Evidence of exemplary contributions to enhanced environmental, social and cultural outcomes including significant outcomes at Level 6 of the Bennett's Hierarchy and emergent outcomes at Level 7.</li> </ul>
Very good value for money	<ul style="list-style-type: none"> <li>• Sufficient results from successful projects to demonstrate we have already broken even on the SFF investment overall AND</li> <li>• Emerging contribution to export opportunities, improved sector productivity; and increased environmentally sustainable practice AND</li> <li>• Evidence of significant contribution to enhanced environmental, social, or cultural outcomes including significant outcomes at Level 6 of the Bennett's Hierarchy.</li> </ul>
Good value for money	<ul style="list-style-type: none"> <li>• Sufficient results from successful projects to credibly forecast break-even on the SFF investment overall AND</li> <li>• Credible contribution in encouraging primary sectors partnering, encouraging and co-investing in industry innovation and adoption, partnering innovative approaches to environmental challenges, and engaging with Māori AND</li> <li>• Evidence of emerging contribution to enhanced environmental, social, OR cultural outcomes including significant outcomes at Level 5 of the Bennett's Hierarchy and emergent outcomes at Level 6.</li> </ul>
Minimally acceptable value for money	<ul style="list-style-type: none"> <li>• The SFF is sufficiently well-utilised on a range of sufficiently promising projects to have a credible prospect of breaking even overall AND</li> <li>• Funds are being allocated and used in accordance with the intended purpose and strategic priorities of the SFF AND</li> <li>• Emerging contribution in encouraging primary sectors partnering, encouraging and co-investing in industry innovation and adoption, partnering innovative approaches to environmental challenges, and engaging with Māori AND</li> <li>• Evidence of emerging contribution to enhanced environmental, social or cultural outcomes – meets Levels 1–4 (Resourcing, activities, participation and reactions) on Bennetts Hierarchy and there are emerging examples from Level 5 (KASA – Knowledge, Attitudes, Skills and Actions).</li> </ul>
Poor value for money	<ul style="list-style-type: none"> <li>• Fund is not sufficiently well-utilised on a range of sufficiently promising projects and has no credible prospect of breaking even OR</li> <li>• No evidence of contribution to enhanced environmental, social, or cultural outcomes at Bennetts Hierarchy Level 5 (KASA – Knowledge, Attitudes, Skills and Actions) or higher.</li> </ul>

<sup>11</sup> Oakden, J., & King J. (2018). Evaluation, Chapter 15 in Tolich, M., Davidson, C. (eds), *Social Science Research in New Zealand*. Auckland University Press. Auckland; Oakden, J., King, J. & Allen, W. (2014). *Evaluation of the Sustainable Farming Fund: Main Report, prepared for Ministry for Primary Industries*. Wellington, New Zealand: Kinnect Group. (p.57)



## AGREE ON CREDIBLE EVIDENCE

### QUESTIONS TO ASK

- What evidence is credible to whom, and in what circumstances?
- What kind of evidence is feasible in context?
- What different sources of data are there for triangulation, particularly on important criteria?
- Do we have multiple sources of data for each evaluation criterion?
- Where there is a shortage of data, what information gives approximate answers to important questions?
- What perspectives are included and missing?
- Can we use existing data or someone else's, or do we need to collect new data?

### GUIDANCE

- This step entails considering what data is needed in order to present credible evidence. There are a range of different perspectives about what counts as evidence and what is considered credible<sup>12</sup> in a particular context. Credibility of evidence will depend on: what the evaluation is being done for (purpose), who it is being done for and with (stakeholders), and the types of evaluation questions being addressed. It's important that key stakeholders discuss this and agree on what they consider credible sources of evidence to be, in the given context.
- In any evaluation, a mix of forms of evidence should be presented that will credibly demonstrate the quality of the strategy, programme or project and to what extent it delivered desired outcomes and value for the overall investment.
- Also, it's generally considered good practice to have **more than one source of evidence** to assess performance on any criterion. All data has flaws, there is no single source of data that tells a complete story. So, it's important to try and draw on more than one source of data or evidence.
- Evidence doesn't have to be new; it can come from existing sources. In many organisational contexts there are already rich sources of credible data available and using this can save lots of time and resource. Traditional forms of data include published reports, internal reporting and longitudinal monitoring. Other existing forms of data include administrative data, social media feedback and website analytics. However, existing sources of data and evidence usually exist for other reasons and may not suit evaluation purposes. It is important to weigh up if the data can be used easily or requires effort make it fit for use in the evaluation. It might be necessary to design and collect new types and forms of data and evidence if there are significant gaps in what is available.

<sup>12</sup> Credibility implies a level of confidence someone has about the authenticity and veracity of the sources of evidence as well as the sufficiency of evidence for the conclusions reached – see Superu (2016) *In Focus: Standards of evidence for understanding what works: International experiences and prospects for Aotearoa New Zealand*, Social Policy Evaluation and Research Unit, Families Commission, Wellington.



### QUESTIONS TO ASK

- What methods of data collection are appropriate and relevant in this context?
- What's the most efficient way to collect and gather the data needed?
- What are the trade-offs that have to be made?
- When do we have enough data?
- How credible is the data, and to whom?
- How will we address causality or contribution?

### GUIDANCE

- **Evaluation design** means the combination of methods chosen for collecting evidence. There are many examples of evaluation design but nearly all of them combine a range of methods. Some examples are experimental and quasi-experimental designs, case study, theory based and qualitative designs.
- There is always a trade-off between what might be ideal and what is practical, given the current circumstances. Factors to consider are: time, budget, accessibility and the expertise available in the team to design, collect and analyse the data.
- It is useful to have people on the evaluation team with expertise in data collection, both of a qualitative and quantitative nature. This ensures the strengths and limitations of different data collection methods are well managed. Otherwise, additional external expertise might be needed to help with data collection.
- For a mixed-methods evaluation design it's important to consider how each of the methods will be implemented.
  - Will they be independent of each other and then connected during analysis?
  - How will they be used in interaction with each other throughout the evaluation process?
  - Will each method have the same relative weight or is one considered to be more important than the other?
- Combining qualitative and quantitative data collection methods ensures that the limitations of one kind are balanced by the strengths of the other. There's now little doubt in the evaluation field that a mix of methods results in a stronger, more complete evaluation than if only one method is used.<sup>13</sup>
- An important evaluation design issue is how to address the causal question: how can we tell if changes we see are a result of the policy, program or intervention we are evaluating, or of something else?
- There are a range of strategies for addressing this question, and the approach taken depends on what is being evaluated. It depends whether results can feasibly be isolated to a single cause or (as is more often the case) there are multiple factors contributing. It also depends on what sorts of decisions will be made from the evidence, and how certain and precise the evidence of causality needs to be.
- It is important to consider the counterfactual when determining the effects of an intervention: what would have happened without it? In some cases this can be measured, e.g., by using a control group. In many social programs, however, we use theory-based approaches. This involves some detective work to look at the multiple strands of evidence, align them with the theory of change, consider alternative explanations, and make a reasoned judgement about the extent to which the intervention might have contributed to the results seen.
- There are many examples of such approaches. One is contribution analysis (Mayne, J. 2008. *Contribution Analysis: An approach to exploring cause and effect*. ILAC Methodological Brief).

<sup>13</sup> Greene, J. (2008) Is Mixed Methods Social Inquiry a Distinctive Methodology? *Journal of Mixed Methods Research* 2008; 2; 7.

Continues on next page →



## GUIDANCE

Some methods that might be considered are listed below<sup>14</sup>:

Method	Overall Purpose	Advantages	Challenges
Questionnaires, surveys, checklists (these can be qualitative or quantitative)	When need to quickly and/or easily get lots of information from people in a non-threatening way	<ul style="list-style-type: none"> <li>• Can complete anonymously</li> <li>• Inexpensive to administer</li> <li>• Easy to compare and analyse</li> <li>• Can administer to many people</li> <li>• Can get lots of data</li> <li>• Many sample questionnaires already exist</li> </ul>	<ul style="list-style-type: none"> <li>• Might not get careful feedback</li> <li>• Wording can bias responses</li> <li>• Are impersonal</li> <li>• May need sampling expert</li> <li>• Doesn't get full story</li> </ul>
Interviews	When want to fully understand someone's impressions or experiences, or learn more about their answers to questionnaires	<ul style="list-style-type: none"> <li>• Get full range and depth of information</li> <li>• Develops relationship with client/user</li> <li>• Can be flexible with client</li> </ul>	<ul style="list-style-type: none"> <li>• Can take much time</li> <li>• Can be hard to analyse and compare</li> <li>• Can be costly</li> <li>• Interviewer can bias client's responses</li> </ul>
Documentation review	When want impression of how programme operates without interrupting the programme; from review or applications, finances, memos, minutes, project plan etc.	<ul style="list-style-type: none"> <li>• Get comprehensive and historical information</li> <li>• Doesn't interrupt routine</li> <li>• Information already exists</li> </ul>	<ul style="list-style-type: none"> <li>• Can take much time</li> <li>• Information may be incomplete</li> <li>• Need to be quite clear about what looking for</li> </ul>
Observation/site visit	To gather accurate information about how a programme actually operates	<ul style="list-style-type: none"> <li>• View operations as they are actually occurring</li> <li>• Can adapt to events as they occur</li> </ul>	<ul style="list-style-type: none"> <li>• Can be difficult to interpret seen behaviours</li> <li>• Can be complex to categories observations</li> <li>• Can influence behaviours of programme participants</li> <li>• Can be expensive</li> </ul>
Focus groups	Explore a topic in depth through group discussion	<ul style="list-style-type: none"> <li>• Quickly and reliably get a common impressions</li> <li>• Can be an efficient way to get much range and depth of information in short time</li> <li>• Can convey key information about programmes</li> </ul>	<ul style="list-style-type: none"> <li>• Can be hard to analyse responses</li> <li>• Need good facilitator for safety</li> </ul>
Case	To fully understand or depict experience in a programme, and conduct comprehensive examination through cross comparison of cases	<ul style="list-style-type: none"> <li>• Fully depicts client's experience in programme input, process and results</li> <li>• Powerful means to portray programme to outsiders</li> </ul>	<ul style="list-style-type: none"> <li>• Usually quite time consuming to collect, organise and describe</li> <li>• Represents depth of information, rather than breadth</li> </ul>

<sup>14</sup> Adapted from Trotman, R. (2008) *Promoting Good(ness), A guide to evaluating programmes and projects*, developed for Auckland City Council.

Continues on next page →





## GUIDANCE

- Another important aspect of this phase of evaluation is developing an **evaluation plan**. Typically, an evaluation plan will set out the following things:
  - **The rationale and evaluation purpose.** This includes the key evaluation questions, context and background. It also covers evaluation needs, a stakeholder analysis and identification of intended use.
  - **Evaluation approach.** This section includes important principles and theoretical assumptions of the strategy, programme or project. It documents the theory of change (if there is one). It also covers the values or evaluative criteria and the performance framework to be used in the evaluation. The approach to synthesis is also documented.
  - **Evaluation process or design.** This section includes the key data sources used as they relate to the evaluation criteria. It also includes methods of data collection, limitations and mitigation strategies.
  - **Evaluation ethics and standards.** This section discusses the ethics that pertain to the evaluation and how they will be included in the evaluation process. It also identifies the standards that will be applied. The anticipated consequences of undertaking the evaluation could also be included here.
  - **Evaluation management.** This section outlines the evaluation roles and responsibilities. It details the activities, schedule, budget, and human, material and other resources.
  - **Evaluation reporting and use.** This part of the plan sets out agreed reporting outputs. It also signals other more adaptive and emergent reporting to meet key stakeholder needs. At different phases of the evaluation, evaluators may try to plan for and maximise different forms of evaluation use.

# 7

## ANALYSE, COMPILE, ASSEMBLE AND ORDER THE EVIDENCE



### QUESTIONS TO ASK

- What does each of the data sources show?
- How robust is the data?
- Do we have enough data?
- How confident can we be about each source of data?
- How much weight should we put on different sources of data?

### GUIDANCE

- Now it's time to analyse the data. At this stage a clear set of evaluation criteria will provide a useful analytical and anchoring framework.
- The first step is to systematically analyse and interpret each individual data source relative to the evaluation criteria. (This is similar to how pre-coded categories might be used in qualitative data analysis). This process doesn't have to wait until all the data has been collected. It can happen on a continuous basis if data is being collected in parallel, or sequentially.
- The evaluative reasoning process should be sufficiently transparent for key stakeholders to see the data that contributes to the evidence. They should also see how that data has been compiled and interpreted using the evaluation criteria.
- After a number of different data sources are analysed, it becomes clearer where there is lots of evidence and where gaps exist. Gaps may occur in data coverage (enough sources of data for each criterion) or quality (data of high enough quality for use).
- If the message emerging from several sources of the data seems consistent, it may be that there is enough data – even that saturation level has been reached.
- On the other hand extra data may be needed. This occurs when a wide range of inconsistent messages emerge or the data seems inadequate in terms of coverage or quality.
- Data that doesn't fit the criteria should be kept separate and reviewed during the evaluation. This data may be important for identifying unexpected or emergent issues or unintended outcomes. It may help identify where further criteria could be added or existing criteria adjusted.



### QUESTIONS TO ASK

- Who should be part of the evaluative reasoning process?
- Now that we've considered the evidence, what do we think about performance (quality, value and importance), and why do we think it?
- What is the significance of this for the strategy, programme or project?
- How transparent is our evaluative reasoning?
- How confident are we that our judgements will be seen as credible and sound?

### GUIDANCE

- Synthesis (combining the elements to form a cohesive whole) is a task that is critical to evaluation. It's where the analysed evidence is combined with the evaluative criteria and the performance framework, and this is used to reach judgements about quality, value and importance and to answer the evaluation questions.
- Evaluation synthesis is a systematic way to take account of the good and not-so-good performance. It helps evaluators reach transparent, credible evaluative conclusions about the quality and value of something. The synthesis process entails assessing evidence using definitions of "quality" and "value" to come to defensible conclusions about performance. This is the crux of evaluative reasoning.
- Michael Quinn Patton (2012) observes that many people assume evaluation is about methods and measurement. Whereas he argues that "It's all about reasoning" and that "...valuing is fundamentally about reasoning and critical thinking".<sup>15</sup>

*Evaluative reasoning is a requirement of all evaluations, irrespective of purpose, method or design.*<sup>16</sup>

- The use of criteria and a performance framework to make judgements from the evidence may be perceived by some stakeholders as being too subjective or less robust than a purely measurement-based system. This results from a misunderstanding about the relationship between measurement and evaluation: measurement can provide evidence for evaluation, but evaluative reasoning is still needed.
- An evaluative judgement cannot simply be made by an algorithm, as it involves weighing multiple pieces of evidence – some of which may be ambiguous or contradictory. The criteria and standards are used to guide the process, with the aim of making transparent and defensible judgements with a clear rationale.<sup>17</sup>
- Evaluation stakeholders are often interested in discussing the evidence and talking about strengths and weaknesses. The synthesis step pushes this one more step to "the evaluative interpretation of evidence".<sup>18</sup> It involves assessing how good, bad, strong or weak the results were and why. This requires a credible justification of the basis for the judgement.
- In the synthesis phase, a collaborative sense-making process enhances the transparency and credibility of evaluation findings for stakeholders.

<sup>15</sup> Patton, M. Q. (2012). Contextual Pragmatics of Valuing, in *Promoting Valuation, New Directions in Evaluation*, Volume 2012, Issue 133, Spring 2012, pp97-108.

<sup>16</sup> Davidson, E. J. (2014). Evaluative Reasoning, *Methodological Briefs: Impact Evaluation 4*, UNICEF Office of Research, Florence.

<sup>17</sup> King, J., & Oxford Policy Management (2018). *The OPM approach to assessing value for money*. A guide. Oxford. Oxford Policy Management Ltd.

<sup>18</sup> See <http://genuineevaluation.com/minirubrics/>



## REPORT, USE AND SOCIALISE RESULTS

### QUESTIONS TO ASK

- Who is this report for?
- What's the best way to frame the evaluation findings and conclusions?
- What are the decisions that will be made as a result of the evaluation?
- Who needs to be influenced by the evaluation?
- How do we share the results in a useful and meaningful way?
- What format(s) will help ensure the evaluation is used and socialised?
- How else might the important messages be communicated?

### GUIDANCE

- This final, reporting stage of an evaluation should be considered right from the start, as it is key to achieving the purpose of the evaluation.
- The plan for reporting and communication of evaluation results should be guided by the agreed use and intended audiences.
- It's important to think about how the findings might be used to influence decisions and changes. It is useful to develop a communication strategy at this stage. This ensures that findings are produced in a manner that is “transparent, understandable and meaningful to stakeholders”.<sup>19</sup>
- There is always a tension putting too little or too much detail in a report. Evaluation reports should cut to the chase and directly answer the evaluation questions. Present the most important points first.
- A good evaluation report tells a compelling performance story. It focuses on the aspects of performance that matter – as defined by the criteria – and this influences the structure. It presents a clear judgement about the level of performance – as defined by the performance framework.
- The evaluation report should give clear answers to important questions. It should: get straight to the point, present transparent evidence and be transparent about the basis on which judgements are made.

Accordingly, evaluation reports should include:

- A summary of the key findings, and clear, concise answers to the key evaluation questions
- A systematic account of the basis upon which each judgement was made, including the evidence, criteria and performance standards
- Extra details, like methods and detailed data analysis presented separately, in appendices or companion reports.
- Finally, reports can take many forms. Consider feeding back results in accessible ways, such as PowerPoint presentations, videos, song, or poetry. After all, reporting is about effective communication and this can sometimes be better served by other media and medium than Word based reports.

<sup>19</sup> Kusters, C., van Vugt, S., Wigboldus, S., & Williams, B. (2011) *Making evaluations matter: A practical guide for evaluators*. Centre for Development Innovation, Wageningen University & Research centre, Wageningen, The Netherlands.



## SUMMARY

This guide presents the Kinnect Group's approach to evaluation. Developed through ten years of collaborative work, it has been well tested. While there are many other guides to evaluation and different ways to do it, our approach emphasises the place of evaluative reasoning in evaluation. The guide draws on the work of many evaluation theorists, as well as our practice-based body of knowledge.

We hope that using the building blocks as explained in the guide will help you in your endeavour to do credible and useful evaluation. We find this approach works for us. Our clients tell us it gives them meaningful and insightful findings that they can use to take action.

Our view to intellectual property is simple. What you read is yours to use. We just ask that you cite the guide when you draw from it. For those of you who engage with this guide and use it in your practice, we look forward to hearing from you. We are particularly interested in feedback where you find aspects of the guide work well for you. In that way we can all continue to build and evolve our practice.

McKegg, K., Oakden, J., Wehipeihana, N., King, J. (2018). *Evaluation Building Blocks: A Guide*. The Kinnect Group: [www.kinnect.co.nz](http://www.kinnect.co.nz)