Frame the boundaries for an evaluation

Framing an evaluation involves being clear about the boundaries of the evaluation.

Why is the evaluation being done? What are the broad evaluation questions it is trying to answer? What are the values that will be used to make judgments about whether it is good or bad, better or worse than alternatives, or getting better or worse?

Identify primary intended users

It is important to identify the people who are intended to actually use the evaluation, and to engage them in the evaluation in some way if possible.

This increases the likelihood that the evaluation will be done in ways that will be appropriate and that will actually be used.

Your primary intended users are not all those who have a stake in the evaluation, nor are they a general audience. They are the specific people, in a specific position, in a specific organization who will use the evaluation findings and who have the capacity to effect change (for example, change policies and procedures, improve management strategies). Who they are will depend on your evaluation.

Research into how evaluation findings are used shows the importance of the 'personal factor'. The personal factor, a specific person or group of people who care about the evaluation findings, is the single most important predictor of evaluation finding use:

'The personal factor is the presence of an identifiable individual or group of people who personally care about the evaluation and the findings it generates. Where such a person or group was present, evaluations were used; where the personal factor was absent, there was a correspondingly marked absence of evaluation impact.'

The tasks of identifying primary intended users and deciding the purposes of an evaluation are interconnected. You might begin by identifying the intended users, who will then decide the purpose of the evaluation. Or the purpose of an evaluation may have already been prescribed,which helps you to identify intended the users.

Resources

- Identifying the intended user(s) and use(s) of an evaluation

  This guideline from the International Development Research Centre (IDRC) highlights the importance of identifying the primary intended user(s) and the intended use(s) of an evaluation and outlines a variety of methods that can be used to achieve this in

- Utilization-Focused Evaluation: 4th edition

  Useful for practitioners and students alike this book is both theoretical and practical. Features include follow-up exercises at the end of each chapter and a utilization-focused evaluation checklist.

- Utilisation-focused evaluation (U-FE) checklist

  Composed by Michael Quinn Patton in 2002 and updated in 2013, this is a comprehensive checklist for undertaking a utilisation-focused evaluation.

Decide purposes

It is important that key stakeholders agree on the main purpose or purposes of evaluation, and be aware of any possible conflicts between purposes.

The purposes of an evaluation will inform (and be informed by) the evaluation timelines, resources, stakeholders involved and choice of evaluation options for describing implementation, context and impact.

It is not enough to state that an evaluation will be used for accountability or for learning.

Evaluations for accountability need to be clear about who will be held accountable to whom for what and through what means. They need to be clear about whether accountability will be upwards (to funders and policymakers), downwards (to intended beneficiaries and communities) or horizontal (to colleagues and partners).

Evaluations for learning need to be clear about who will be learning about what and through what means. Will it be supporting ongoing learning for incremental improvements by service deliverers or learning about 'what works' or 'what works for whom in what circumstances' to inform future policy and investment?

It may be possible to address several purposes in a single evaluation design but often there needs to be a choice about where resources will be primarily focused.

Methods

Using findings

- Contribute to broader evidence base

  Inform future policy and practice by others outside the organisation.

- Inform decision making aimed at improvement (formative)

  Changing or confirming policies and practices.

- Inform decision making aimed at selection, continuation or termination (summative)

  Identifying best value for money.

- Lobby and advocate

  Justify expenditure and demonstrate achievements.

Using process

- Build trust and legitimacy across stakeholders

  Develop better understandings of each other and demonstrate that expectations are being met.

- Ensure accountability

  Holding someone to account to someone for something.

- Ensure diverse perspectives are included, especially those with little voice

  Make explicit the experiences and values of key stakeholders, especially intended beneficiaries.

Resources

- Exploding the myth of incompatibility between accountability and learning

  This chapter from Capacity Development in Practice (archived link) examines the conflict in the field of Monitoring and Evaluation (M&E) between the need for 'accountability' and the desire to ensure 'learning'.

- Purposes of assessment - Keystone guide

  This webpage from Keystone Accountability outlines the six major reasons that social organizations monitor, assess and report their performance and results.

  The reasons identified include:

- Seeking Surprise: Rethinking monitoring for collective learning in rural resource management

  This PhD Thesis from Irene Guijt draws on her extensive knowledge and experience in the field of rural resource management in Brazil.

- Utilization-Focused Evaluation: 4th edition

  Useful for practitioners and students alike this book is both theoretical and practical. Features include follow-up exercises at the end of each chapter and a utilization-focused evaluation checklist.

Specify the key evaluation questions

Key Evaluation Questions (KEQs) are the high-level questions that an evaluation is designed to answer - not specific questions that are asked in an interview or a questionnaire.

Having an agreed set of Key Evaluation Questions (KEQs) makes it easier to decide what data to collect, how to analyze it, and how to report it.

KEQs usually need to be developed and agreed on at the beginning of evaluation planning - however sometimes KEQs are already prescribed by an evaluation system or a previously developed evaluation framework.

Try not to have too many Key Evaluation Questions - a maximum of 5-7 main questions will be sufficient. It might also be useful to have some more specific questions under the KEQs.

Key Evaluation Questions should be developed by considering the type of evaluation being done, its intended users, its intended uses (purposes), and the evaluative criteria being used. In particular, it can be helpful to imagine scenarios where the answers to the KEQs being used - to check the KEQs are likely to be relevant and useful and that they cover the range of issues that the evaluation is intended to address. (This process can also help to review the types of data that might be feasible and credible to use to answer the KEQs).

The following information has been taken from the New South Wales Government, Department of Premier and Cabinet Evaluation Toolkit, which BetterEvaluation helped to develop.

Key evaluation questions for the three main types of evaluation

Process evaluation

How is the program being implemented?

How appropriate are the processes compared with quality standards?

Is the program being implemented correctly?

Are participants being reached as intended?

How satisfied are program clients? For which clients?

What has been done in an innovative way?

Outcome evaluation (or impact evaluation)

How well did the program work?

Did the program produce or contribute to the intended outcomes in the short, medium and long term?

For whom, in what ways and in what circumstances? What unintended outcomes (positive and negative) were produced?

To what extent can changes be attributed to the program?

What were the particular features of the program and context that made a difference?

What was the influence of other factors

Economic evaluation (cost-effectiveness analysis and cost-benefit analysis)

What has been the ratio of costs to benefits?

What is the most cost-effective option?

Has the intervention been cost-effective (compared to alternatives)?

Is the program the best use of resources?

Appropriateness, effectiveness and efficiency

Three broad categories of key evaluation questions are often used to assess whether the program is appropriate, effective and efficient .

Organising key evaluation questions under these categories, allows an assessment of the degree to which a particular program in particular circumstances is appropriate, effective and efficient. Suitable questions under these categories will vary with the different types of evaluation (process, outcome or economic).

Appropriateness

To what extent does the program address an identified need?

How well does the program align with government and agency priorities?

Does the program represent a legitimate role for government?

Effectiveness

To what extent is the program achieving the intended outcomes, in the short, medium and long term?

To what extent is the program producing worthwhile results (outputs, outcomes) and/or meeting each of its objectives?

Efficiency

Do the outcomes of the program represent value for money?

To what extent is the relationship between inputs and outputs timely, cost-effective and to expected standards?

Example

The Evaluation of the Stronger Families and Communities Strategy used clear Key Evaluation Questions to ensure a coherent evaluation despite the scale and diversity of what was being evaluated – an evaluation over 3 years, covering more than 600 different projects funded through 5 different funding initiatives, and producing 7 issues papers and 11 case study reports (including studies of particular funding initiatives) as well as ongoing progress reports and a final report.

The Key Evaluation Questions were developed through an extensive consultative process to develop the evaluation framework, which was done before advertising the contract to conduct the actual evaluation.

1. How is the Strategy contributing to family and community strength in the short-term, medium-term, and longer-term?
2. To what extent has the Strategy produced unintended outcomes (positive and negative)?
3. What were the costs and benefits of the Strategy relative to similar national and international interventions? (Given data limitations, this was revised to ask the question in ‘broad, qualitative terms’
4. What were the particular features of the Strategy that made a difference?

5. What is helping or hindering the initiatives to achieve their objectives? What explains why some initiatives work? In particular, does the interaction between different initiatives contribute to achieving better outcomes?
6. How does the Strategy contribute to the achievement of outcomes in conjunction with other initiatives, programs or services in the area?
7. What else is helping or hindering the Strategy to achieve its objectives and outcomes? What works best for whom, why and when?
8. How can the Strategy achieve better outcomes?

   CIRCLE (2008) Stronger Families and Communities Strategy 2000-2004: Final Report. Melbourne: RMIT University.

The KEQs were used to structure progress reports and the final report, providing a clear framework for bringing together diverse evidence and an emerging narrative about the findings.

The Managers' Guide

Coming at this from a manager or commissioner's perspective? Step 2: Scope the evaluation in our Managers' Guide has some specific information geared towards making decisions about what the evaluation needs to do, including how to develop agreed key evaluation questions.

Resources

- Practical guide for engaging stakeholders in developing evaluation questions

  This guide from the Robert Wood Johnson Foundation was designed to support evaluators engage their stakeholders in the evaluation process.

- Looking back, moving forward: Sida evaluation manual

  This manual from the Swedish International Development Cooperation Agency (SIDA) is aimed at supporting staff in conducting evaluations of development interventions.

- Evaluation questions

  This site provides a step-by-step guide on how to identify appropriate questions for an evaluation.

- Stakeholders' interest in potential evaluation questions

  This worksheet from Chapter 5 of the National Science Foundation's User-Friendly Handbook for Mixed Method Evaluations provides a template for developing evaluation questions which engage stakeholders interest in the process.

- Prioritize and eliminate questions

  This worksheet from Chapter 5 of the National Science Foundation's User-Friendly Handbook for Mixed Method Evaluations provides a template which allows the organisation and selection of possible evaluation questions.

- CDC: Checklist to help focus your evaluation

This checklist, created by the Centers for Disease Control and Prevention (CDC), helps you to assess potential evaluation questions in terms of their relevance, feasibility, fit with the values, nature and theory of change of the program, and the level

- Evaluation questions checklist for program evaluation

  Created by Lori Wingate and Daniala Schroeter, the purpose of this checklist is to aid in developing effective and appropriate evaluation questions and in assessing the quality of existing questions.

- Evaluation question examples: Evaluation at country level, regional level, sector or thematic global evaluation

  This document contains example questions, many of which are drawn from country, regional, sector or thematic global evaluations undertaken by the Evaluation Unit.

Determine what ‘success’ looks like

Evaluation is essentially about values, asking questions such as : What is good, better, best? Have things improved or got worse? How can they be improved?

Therefore, it is important for evaluations to be systematic and transparent in the values that are used to decide criteria and standards.

Criteria

Criteria refer to the aspects of an intervention that are important to consider when deciding whether or not, and in what ways, it has been a success or a failure, or when producing an overall judgement of performance. There are different types of criteria:

Positive outcomes and impacts: for example, should childcare be judged in terms of its success in supporting early childhood development or in supporting parents to engage in education or work? If it is both, how should they be weighted?

Negative outcomes and impacts: for example, an infrastructure development might produce negative unintended effects (e.g. soil erosion caused by a new road) as well as positive intended effects)

Distribution of costs and benefits: for example, is it important for everyone to receive some benefit or the same benefit or for the intervention to be targeted so that the most disadvantaged receive more benefit?

Resources and timing: for example, is there a need for results to be achieved within a certain timeframe?

Processes: for example, use of recyclable materials; providing access to groups with restricted mobility

Standards

Standards refer to the levels of performance required for each of the criteria. For example, if a project aims to reduce maternal mortality, what level of performance is needed for it to be considered successful? Any reduction? A reduction of at least xx%? A reduction of at least xx in absolute terms? A reduction to a rate of x.x that matches other similar regions, or matches official targets?

Criteria and standards need to be agreed on in order to identify the data that need to be gathered for an evaluation.

In addition, these data need to be combined to form an overall judgement of success or failure, or to rank alternatives against each other. For example, if a road project achieves its economic objectives but produces environmental damage, should it be considered a success overall? How much damage, and at whose cost, would be enough to outweigh the positive impacts? These issues are addressed under the task Synthesise data from a single evaluation.

Methods

Formal statements of values

- Some options are used to identify possible criteria and standards that could be used in an evaluation, drawing on formal and informal sources, and some options are used to negotiate which should be used and how they should be weighed.

- Standards, evaluative criteria and benchmarks

  Standards, evaluative criteria, or benchmarks refer to the criteria by which an evaluand will be judged during an evaluation.

- Stated goals and objectives

  Evaluations can use the program's stated objectives and goals to assess program success or failure.

Articulate and document tacit values

- Hierarchical card sorting

  Hierarchical card sorting (HCS) is a participatory card sorting method designed to provide insight into how people categorise and rank different phenomena.

- Open space

  Open Space Technology (OST) is a group facilitation approach for small and large gatherings in which a central purpose, issue, or task is addressed, but which begins with a purposeful lack of any formal initial agenda.

- Photovoice

  Photovoice is a participatory photography method that seeks to empower marginalised people to share their experiences through digital storytelling.

- Rich pictures

  A rich picture is a way to explore, acknowledge and define a situation and express it through diagrams to create a preliminary mental model and can help to open discussion and come to a broad, shared understanding of a situation.

- Stories of change

Stories of change show what is valued through the use of specific narratives of events.

Structured with a beginning, middle and end, they focus on the change that has taken place due to the program.

- Values clarification interviews

  Values Clarification Interviews involve interviewing key informants and intended beneficiaries to identify what they value.

- Values clarification public opinion questionnaires

  Seeking feedback from large numbers of people about their priorities through the use of questionnaires.

Negotiate between different values

- Concept mapping

  A concept map shows how different ideas relate to each other - sometimes this is called a mind map or a cluster map.

- Delphi study

  The Delphi technique is a quantitative option to generate group consensus through an iterative process of answering questions.

- Dotmocracy

  Dotmocracy is an established facilitation method for collecting and recognizing levels of agreement on written statements among a large number of people.

- Open space

  Open Space Technology (OST) is a group facilitation approach for small and large gatherings in which a central purpose, issue, or task is addressed, but which begins with a purposeful lack of any formal initial agenda.

- Public consultations

  Public consultations are usually conducted through public meetings to provide an opportunity for the community to raise issues of concern and respond to options.

Approaches

- Critical systems heuristics

  This chapter provides a detailed introduction to critical systems heuristics and the use of its central tool boundary critique.

- Participatory evaluation

Participatory evaluation is an approach that involves the stakeholders of a programme or policy in the evaluation process.