EVALUATION

A project or programme evaluation is normally carried out at a significant time. This might be during a project or programme, at the end, or sometime after. An evaluation may also be carried out when a project or programme is about to move into a new phase, or in response to a critical issue. There are many different types of evaluation, and they are applied in many different circumstances.

A project or programme evaluation is normally carried out at a significant time. This might be during a project or programme, at the end, or sometime after. An evaluation may also be carried out when a project or programme is about to move into a new phase, or in response to a critical issue. There are many different definitions of evaluation. The most widely used is *"the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results"* (OECD 2010).

Evaluation differs from monitoring in several ways, although there are often exceptions to these rules. For example:

- an evaluation is normally a short, time-bound exercise, whilst monitoring is more likely to be on-going;
- an evaluation usually focuses on outcomes and impact, rather than activities and outputs; and
- an evaluation is often carried out by an external person or team, whilst monitoring tends to be carried out by project or programme staff.

Within the CSO world, evaluations are most often carried out within projects and programmes. However, evaluations may also focus on different levels. For example (see Bakewell et al. 2003):

- a CSO involved in a number of projects within the same sector (e.g. water and sanitation, health) may want to evaluate all their activities in this sector over a period;
- a CSO might wish to evaluate a strategy designed to achieve its overall goals, such as working in a particular region or adopting an overall approach to its work; and
- some large NGOs regularly evaluate regional or country programmes of work.

In fact it is possible to evaluate almost any intervention or set of interventions, including training courses, events, policies, funding mechanisms, relationships and partnerships. However, these kinds of evaluations tend to be carried out by larger agencies, such as International NGOs or large Southern NGOs.

The different stages of an evaluation

There are many different types of evaluation, and they can be applied in many different circumstances. Consequently, it is not possible to provide a comprehensive step-by-step approach to carrying out an evaluation. Nonetheless, it is possible to outline some of the broad stages that are applied in most evaluations. Note that in different circumstances some of these stages may be omitted, some may be added, and some may be carried out in a different order.

STAGE ONE

It is always very important for different stakeholders to **agree the purpose** or purposes of an evaluation at the start. This means being clear about who will be expected to make decisions based on the findings of the evaluation. Evaluations can serve many different purposes, and many evaluations are carried out for more than one purpose. Some of these purposes are described below (see DFID 2013).

- Evaluations carried out during a project or programme are often designed to generate information and recommendations which can be used to improve the performance of the project or programme being evaluated.
- Evaluations may also generate lessons that can be applied within future projects or programmes, or by other agencies operating similar projects and programmes in other times and places.
- Evaluations are often designed to enable organisations to account for the use of resources to funding agencies, governments, International NGOs (INGOs), multilateral agencies and/or taxpayers.
- Sometimes an evaluation can be used to help decide whether or not to extend a project or programme into a new phase.
- The findings of evaluations may be used to inform policy decisions. These may be internal (e.g. policies developed by the institutions managing an evaluation, such as donors, governments or NGOs) or external, such as policies developed by government agencies.

- Evaluations are often used to generate information on achievements which can then be communicated to wider audiences for marketing or fundraising purposes.
- Some evaluations are specifically designed to improve the development process within the project or programme concerned. Examples could include building trust and understanding between different stakeholders, and ensuring that different voices are heard within projects and programmes, especially those of marginalised or excluded groups.

STAGE TWO

The next stage is to **decide on the type of evaluation**. There are many different types of evaluation. Some are designed to achieve one or more of the different purposes described above, whilst others are used within specific types or sectors of work. The type of evaluation has implications for how the work of the evaluation is taken forward, who takes it forward, and who is involved at different stages. Some of the different types of evaluation are described in a separate paper in this section of the M&E Universe.

STAGE THREE

It is always important to **establish who should be involved** in an evaluation from the beginning, and how they should be involved at each stage of the process. There are three key questions that need to be addressed: who needs to be involved in the evaluation, at what stage do they need to be involved, and what purpose their participation will serve.

Participation of different stakeholders needs to be considered at each stage of an evaluation process. For example, different stakeholders might be involved in establishing the purpose of the evaluation, deciding on the type of evaluation, contributing to data collection and analysis, supporting reporting and communication, and using the findings.

Decisions on stakeholder participation – and especially beneficiary participation – should be made based on two criteria. The first is whether or not those stakeholders can help produce better evaluation findings (e.g. contribute better information, support better analysis, or improve the uptake of findings). The second criterion is whether they have the *right* to be involved. This is important for any organisation involved in rights-based programming, or which believes that people have a right to be involved in informing decisions that will affect them.

STAGE FOUR

The next stage is to **develop a Terms of Reference**. An evaluation Terms of Reference (ToR) is a brief document that outlines the purpose, scope and objectives of an evaluation. It can also include information on work plans, methodologies, logistics and budgets. A ToR should be formally agreed between different stakeholders, including those funding an evaluation, those managing the evaluation, those primarily responsible for implementing the evaluation, and the organisation or organisations responsible for the intervention that is being evaluated. A ToR is normally produced before the detailed work of an evaluation begins, and helps shape both design and implementation. It makes clear from the start what is expected from those involved in the evaluation.

Some key elements of a typical ToR for a straightforward project or programme evaluation are shown in the box on the following page (see World Vision 2007, Tearfund u.d.).

STAGE FIVE

Once the ToR has been agreed, the next step is normally to **put together an evaluation team** or specify an individual who will carry out the work. Some teams or individuals are appointed to carry out an evaluation because of their position, knowledge or expertise. Other teams are appointed based on competitive tendering.

Evaluations can be run by an external evaluator or evaluation team, and this is still the most common situation. However, they can also be run by a separate unit of an implementing organisation. For example, many United Nations (UN) bodies have semi-independent evaluation units that carry out evaluations on UN programmes. Evaluations may also be run by internal staff within a development intervention, or (more rarely) by those involved in an intervention, such as the intended beneficiaries.

An evaluation team may be comprised of both internal and external people. This is particularly important when there are language or cultural barriers that make it difficult for outsiders to conduct interviews or observe proceedings. In these cases it is common to supplement external experts with local evaluators who either have an in-depth knowledge of the intervention being evaluated, or at least a strong understanding of local cultures and language. Evaluation teams may also be selected to ensure balance, for example by making sure at least one team member has expertise in an area such as gender or social inclusion.

STAGE SIX

Once a team or individual has been chosen, the next step is to **design the evaluation** itself. The design will be heavily influenced by decisions made during earlier stages, and will be based on the ToR. It may also be influenced by an initial literature review of relevant material, which is usually conducted after the ToR has been developed but before a final evaluation plan has been developed.

Once the evaluation has been designed a formal evaluation plan is often produced, either as a standalone document or as an extension to the ToR. In large evaluations, an inception report is often produced at this stage, which needs to be approved before the evaluation can go ahead. The evaluation plan may include:

- details of who will be observed, surveyed or interviewed;
- details of how an evaluation will be conducted;

Elements of a Typical Terms of Reference (ToR)

Basic information: the project / programme title and period covered by the evaluation; the main project / programme objectives and key working approaches; the background to the project / programme, including the political and socio-economic context; and summaries of previous reports or evaluations.

Evaluation partners: a list and description of the key stakeholders within the evaluation. This could include the organisation commissioning the evaluation, the main proposed users of the evaluation, and key stakeholders expected to be involved in the evaluation.

The purpose and objectives of the evaluation: the main purpose of the evaluation, along with any secondary purposes. This section should also include the key evaluation questions to be answered through the evaluation, and the information that will be needed to answer them.

Methodology: if the methodology is defined before the evaluation team is in place then the ToR should include a description of the data collection and analysis methods to be used. Otherwise it should specify who should define the methodology – the organisation commissioning the evaluation, the evaluation team, or the project stakeholders. Other issues to cover include how different stakeholders will be involved in data collection, analysis and use, and key limitations (time, budgets, politics, etc.) that might affect the usefulness of the evaluation.

Conduct of the evaluation: an outline of the key personnel required within the evaluation team, and their roles and areas of expertise. This includes any key qualifications and/or experiences required, especially for the team leader or any technical advisors.

Schedules: dates and schedules for procuring, selecting or forming the evaluation team, and dates, duration and logistics for each evaluation event. This might include briefing / debriefing sessions, travel, fieldwork, submission of final reports and presentations, or any other relevant activity.

Expected outputs: the products, including the final report, which will result from the evaluation. Expected outputs might also include details of how data and results will be recorded, and how feedback will be provided to partners and beneficiaries. This section might also cover the format of the final report required, and should clarify the scope of any required recommendations – in other words, it should specify which groups will be expected to act on any recommendations.

Documents: a list of key documents that should be reviewed as part of the evaluation.

Management of the evaluation: the person or people responsible for commissioning and approving the work, or resolving issues as they arise. This section should also outline the person or people responsible for practical arrangements, e.g. travel, accommodation.

Review of the evaluation: clarification of who will provide feedback on how the evaluation process went, and when they will provide this feedback. This section should also include details of who will review and comment on the report, and who will provide quality assurance.

Annexes: A ToR might also contain annexes containing more detailed information. This could include some of the following.

- A logical framework or equivalent planning tool for the project or programme.
- A detailed budget for the evaluation, broken down according to different criteria.
- A person specification for different team members.
- A specification and breakdown of the working days required from different evaluation team members.
- A timeline for the evaluation, outlining detailed activities to be carried out.

- a description of methodologies for data collection and analysis;
- a description of processes used to analyse data;
- a proposed format for any reports;
- proposed methods for the dissemination of findings to different stakeholders;
- a schedule of activities; and
- a budget.

STAGE SEVEN

Once the ToR has been completed and the evaluation designed to everyone's satisfaction the next task is to **implement the evaluation**. Implementation is likely to involve a range of different activities such as planning and organising logistics (e.g. vehicles, accommodation, events, meetings); collecting primary data and reviewing secondary data; analysing the data; identifying major findings, lessons and recommendations; preparing a report or reports; presenting preliminary and final results to different stakeholders in different ways; and incorporating feedback into evaluation findings and recommendations.

It is common during the implementation stage of an evaluation to have some kind of formal oversight over the conduct of the evaluation. This may be an individual or a committee. This is considered important as it helps ensure that the evaluation is carried out to appropriate standards.

STAGE EIGHT

Most evaluations require the **production of a formal report** or reports. In the end, the stakeholder who commissioned the report, or the management oversight, needs to make a judgement on whether the report is good enough. INTRAC's experience over many years is that an evaluation report typically needs to perform four key functions (see box below).

Four Key Functions of an Evaluation Report

1. A good evaluation report needs to answer the four universal M&E questions:
   - Has the project, programme or organisation done what it said it would do?
   - Has it done it well?
   - What difference has it made?
   - Has it done the right things?
2. The report needs to fulfil the evaluation's ToR, and should answer the key evaluation questions set out in those ToR.
3. It needs to be digestible – in other words it needs to be an appropriate length with a logical structure, and with a style and language that is appropriate for its intended readership.
4. The report needs to facilitate learning and enable decision-makers to make better decisions about the project, programme or organisation.

STAGE NINE

The whole purpose of an evaluation is to produce information that is used. Sadly, this is often not the case, and there is a well-known tendency for CSOs to produce an *'elephants' graveyard'* of evaluation reports that remain unread and unused after their immediate publication.

The final stage of an evaluation should therefore be to **act to maximise the possibility of the evaluation findings being used**. This can include the following (see World Vision 2007):

- getting agreement from different stakeholders on the evaluation's recommendations, and how they will be acted upon;
- deciding on the main person(s) responsible for following up key recommendations;
- circulating the report's findings to different stakeholders in different ways to best suit the audience (e.g. reports, summaries, case studies, presentations, workshops, events, feedback sessions in the field);
- disseminating the findings of the evaluation to actors in other projects, programmes or organisations who might benefit;
- ensuring that recommendations for the future are captured and stored in such a way that they can be easily retrieved at a future date; and
- making sure that the raw data captured during the evaluation is readily available so that it can be re-examined, if necessary, in the future.

In general, the more these actions have been considered at an earlier stage – for instance when designing the evaluation or producing the ToR – the easier it is to ensure that the findings of the evaluation will be used.

Evaluation criteria and standards

There are many sets of evaluation standards and criteria that have been developed to support the better use of evaluations. The most well-known set of criteria is the OECD DAC evaluation criteria. It consists of five criteria for assessing development evaluations, with an additional four that are specifically designed to address evaluations in the humanitarian sector. The criteria are shown in the box on the following page (see OECD 2010, Alnap 2016). They are basically a list of different aspects of a project or programme that an evaluation ought to cover. They are designed to be a checklist to ensure that key issues are considered in each evaluation, although not all criteria are designed to be applied in every evaluation. The criteria are not intended to be a blueprint for evaluation, and do not replace the need to develop individual questions for each evaluation.

In addition to documents covering evaluation criteria, there is a large and growing number of documents covering evaluation standards. Some of these are very detailed and prescriptive. Others, however, are more concerned with broader principles. For example, the UK Evaluation Society (UKES 2013) has produced a set of guidelines for evaluators and those commissioning evaluations, including measures to ensure that the rights of all evaluation participants are upheld. The guidelines are basically intended to help commissioners, practitioners and participants establish good practice in the conduct of evaluations. Many of the guidelines emphasise the need to be open and transparent about the expectations and requirements of different stakeholders. These guidelines are referred to in the further reading section of this paper.

Good practice in evaluations

Many evaluations only ever reach and influence a small group of people, often limited to the staff and donors of the project or programme being evaluated. Potentially useful learning and recommendations often remain unused. Based on extensive experience over the years, there are a few actions which INTRAC believes can promote the wider and better use of evaluation learning. These are as follows.

- Clearly identify the people who are expected to use the evaluation findings, and engage them in the process at the earliest possible stage. This means that the findings and recommendations are more likely to be relevant, owned and used.
- Time the evaluation to coincide with key moments of planning, reflection or strategy development. This will also provide greater opportunity for the findings to be used.
- Secure the support and commitment of senior managers, and build in strong follow-up and accountability mechanisms. This will help ensure that recommendations are likely to be implemented.
- Design the evaluation outputs and dissemination strategies for the intended users, so that findings are more likely to reach those who can make good use of them. This may mean being more innovative in finding different ways to communicate findings to different groups. A lengthy report is not always the right answer!
- Pay attention to ethical standards throughout the evaluation process. This means ensuring that the evaluation methodology is conducive to ethical good practice throughout, from planning through to data collection, analysis and use.
- Plan an evaluation at the first possible opportunity, ideally as soon as a project or programme begins. The method used for an evaluation often has implications for how baseline information should be collected. Many decisions taken at the start of a project or programme have implications for future evaluations.

OECD DAC Evaluation Criteria	
Relevance	Relevance means the extent to which a development intervention was suited to the priorities and policies of the target group, recipient and donor.
Effectiveness	Effectiveness is a measure of the extent to which a development intervention has attained its objectives.
Efficiency	Efficiency is an economic term which signifies that the development intervention used the least costly resources possible to achieve the desired results.
Impact	Impact includes the positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended.
Sustainability	Sustainability is concerned with assessing whether the benefits of an intervention are likely to continue (or have continued) after donor funding has been withdrawn.
Additional Criteria for Humanitarian Evaluations	
Coverage	Coverage means the extent to which major population groups facing life-threatening suffering were reached by the humanitarian action.
Coherence	Coherence is the extent to which security, developmental, trade, and military policies, as well as humanitarian policies, are consistent and consider humanitarian and human rights considerations.
Coordination	Coordination means the extent to which the interventions of different actors were harmonised with each other, promoted synergy, and avoided gaps, duplication and resource conflicts.
Connectedness	Connectedness assesses the extent to which activities of a short-term emergency nature were carried out in a context that took longer-term and interconnected problems into account.

Further reading and resources

This section of the M&E Universe contains papers on monitoring, impact assessment, review, research, and different types of evaluation. Other sections of the M&E Universe cover data collection, analysis and use.

There are many different documents available that contain evaluation guidelines. INTRAC usually uses the version produced by the UK Evaluation Society (UKES 2013). This is freely available from the internet at https://www.evaluation.org.uk/guidelines-new-edition/

The Better Evaluation website (www.betterevaluation.org) contains the largest set of resources in the world covering evaluation. The site offers step-by-step guidance for those managing or implementing evaluations. Experienced evaluators or those with an interest in evaluation are recommended to go to that site and search through the different materials. The Better Evaluation site also contains an online tool to help generate Terms of Reference. This can be accessed through the landing page.

References

- Alnap (2016). *Evaluation of Humanitarian Action Guide*. London.
- Bakewell, O; Adams, J and Pratt, B (2003). *Sharpening the Development Process; A practical guide to monitoring and evaluation*. INTRAC, UK.
- DFID (2013). *International Development Evaluation Policy.* Department for International Development (DFID), UK, May 2013.
- OECD (2010). *Evaluating Development Cooperation: Summary of key norms and standards*, 2nd edition. OECD DAC Network on Development Evaluation.
- Tearfund (u.d.). *Guidelines for Terms of Reference*, sourced on p46 of Bakewell et. al. (2003)
- UKES (2013). *Guidelines for Good Practice in Evaluation*.
- World Vision (2007). *Learning Through Evaluation With Accountability and Planning (LEAP)*, 2nd edition. World Vision.

Author(s):
INTRAC

Contributor(s):
Anne Garbutt, Alison Napier, Vera Scholz, Nigel Simister

INTRAC is a not-for-profit organisation that builds the skills and knowledge of civil society organisations to be more effective in addressing poverty and inequality. Since 1992 INTRAC has provided specialist support in monitoring and evaluation, working with people to develop their own M&E approaches and tools, based on their needs. We encourage appropriate and practical M&E, based on understanding what works in different contexts.