



MANAGING EVALUATIONS IN THE FIELD: STEP-BY-STEP PLANNING GUIDANCE

ANALYSIS AND EVIDENCE

1. WHAT THIS GUIDANCE IS FOR AND HOW TO USE IT

How do we know if the International Committee of the Red Cross's (ICRC) humanitarian work is effective and successful? Evaluations are one of the key ways to answer this question. By pausing and taking stock – from a more objective perspective – of what is working or has worked and what needs to change, it is possible to gain structured insight into the overall performance of an intervention. Evaluations therefore contribute to better intervention accountability and learning.

Although there is no shortage of guidance for humanitarian organizations and aid workers on evaluations, none of these existing guides align with the ICRC's mandate and working methods, or with the nature of its partners' work. This guidance is designed for individuals or teams in ICRC delegations who manage evaluations in the field.¹ It will also prove useful for regional and headquarters staff, as well as for other humanitarian and development organizations managing these kinds of evaluations, such as components of the International Red Cross and Red Crescent Movement, the United Nations, and national and international non-governmental organizations (NGOs).

Field evaluations are managed directly by delegations, with oversight from the Evaluation Office. An evaluation is a step-by-step process that includes designing evaluation activities and outputs and agreeing upon these with key stakeholders. This document helps delegations to identify these steps and to develop a Terms of Reference (ToR) document. A ToR helps the delegation share key information about the evaluation with stakeholders, guides the evaluation team, and sets out the expectations that apply to the various phases of the evaluation and the resources that will be needed.

This guidance is useful for ICRC teams in the field managing different evaluation types that cover various interventions. Importantly, it is not an evaluation manual. Nor is it the only source to draw on when planning and managing an evaluation. It should be used in conjunction with other ICRC policies, standards and guidance on which the evaluation relies.

In this document, we use the term “evaluation” as a catch-all term to refer to evaluations, impact evaluations, evaluative reviews and learning workshops. Evaluations are determined by a level of objectivity and/or independence and are conducted according to clear lines of inquiry and a methodological approach (see [section 3](#)). Specifically, when we refer to an evaluation, we understand it as follows:

The systematic and objective assessment of an ongoing or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, developmental efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learnt into the decision-making process of both recipients and donors.²

Likewise, we use the term “intervention” to refer to the subject of the evaluation, including all the various types of work or efforts that may be evaluated (such as a project, programme, strategy, thematic area, or other activity or action).³

This document will also prove useful for those guiding design, strategic planning and results management at the start of the intervention. A good evaluation relies on effective monitoring, evaluation and learning systems within the intervention cycle. Gaining clarity on what success looks like at the design phase of an intervention helps to make the intervention assessable.

¹ Field evaluations differ from centralized evaluations, which are managed at the institutional level (e.g. by the Evaluation Office).

² Organisation for Economic Co-operation and Development (OECD), “DAC Glossary of Key Terms and Concepts”: <https://www.oecd.org/dac/dac-glossary.htm#Evaluation>, all web addresses accessed September 2021.

³ OECD, *Applying Evaluation Criteria Thoughtfully*, OECD, Paris, 2021a: https://www.oecd-ilibrary.org/development/applying-evaluation-criteria-thoughtfully_543e84ed-en.

2. WHAT SUPPORT IS AVAILABLE TO HELP MANAGE EVALUATIONS?

The Evaluation Office aims to help embed evaluation practice and culture in the organization. It does this by commissioning centralized, independent external evaluations in order to review progress toward the ICRC's strategic orientations, and by supporting colleagues in the organization to commission their own (decentralized) independent external evaluations in the field at specific points in time.

The Analysis & Evidence (A&E) team provides methodological expertise and technical support on collecting, managing and using data to inform decision-making across the project cycle. In other words, A&E encourages the systematic use of evidence to design, adapt and learn throughout a project. A&E may support internal evaluation processes as part of the regular project cycle, such as by helping to design methodologies or leading impact evaluations.

In delegations, A&E and evaluation work are complementary. An evaluation will use the data, analyses and interpretations that A&E has helped to develop during the course of an intervention, but may also look at other aspects. Conversely, an evaluation can help identify ways to strengthen existing A&E practices in a given setting.

In practical terms, the Evaluation Office provides real-time feedback on key products of an evaluation in order to support the process. This means that delegations conducting an evaluation can seek constructive advice on draft ToR, draft inception reports and draft evaluation reports. The feedback highlights the strengths of each document and offers suggestions and recommendations for further improvement. The quality criteria applicable to each document type – in the form of a checklist – are available to evaluation commissioners (the checklists for inception reports and evaluation reports should also be made available to external evaluators). These checklists, and the associated constructive feedback process, form the basis of the ICRC's evaluation quality assurance system.

3. EVALUATION TYPES

The appropriate evaluation type will depend in part on the steps involved in managing a given evaluation (see [section 4](#)). Evaluation types can be determined according to various factors, such as the scope of the exercise (e.g. project, programme or thematic evaluation), the level of results being evaluated (e.g. process or impact evaluation), the timing (e.g. mid-term evaluation or *ex-post* evaluation) and who is conducting the evaluation (e.g. participatory, internal or external evaluation). This guidance differentiates between four main evaluation types: external evaluations, impact evaluations (either external or internal), internal evaluative reviews and learning workshops.⁴

3.1 EVALUATION

An evaluation is a systematic examination of an intervention to determine its worth or significance. It creates an objective evidence base to support important decisions, draw lessons to improve interventions and practice, and enhance accountability (including transparency) to people affected by conflict or other violence, as well as to donors, staff, partners and/or authorities.

Evaluations are appraised against recognized quality criteria and are conducted by an external third party or individual with oversight from the Evaluation Office, which lends it a degree of objectivity and technical expertise. Evaluations are an integral part of the ICRC's Planning and Monitoring for Results system. They should be factored into the annual budget and resource planning exercise, which should include both planned evaluations and enough flexibility to conduct unplanned evaluations when the need arises. Evaluations should be systematically planned for, in addition to other evaluative activities.

3.2 IMPACT EVALUATION

An impact evaluation can be conducted as part of a full evaluation process or as a stand-alone exercise. It focuses on the achievement of objectives and/or the wider effects of an intervention, rather than on its management and delivery. These include intended and unintended, positive and negative, macro (sector) and micro (household, individual), short- and long-term, and direct and indirect impacts. An impact evaluation may also explore cause and effect, by assessing what impact is caused by the intervention compared to the many other factors at play.

An appropriate approach should be designed, and a data collection method (qualitative and/or quantitative) selected, depending on the scope of the impact evaluation. In terms of design, there are two options, each with its own advantages and disadvantages:

- A non-experimental design (such as a case study, or an analysis of outcome and impact indicators for the people who received assistance over time): This is the least demanding design, since there is usually no comparison group (i.e. people who did not receive assistance). While this approach can give some insight into the impact of an intervention, it does not always provide enough data to draw solid conclusions as to whether an impact is not caused by other changes in the environment.
- A quasi-experimental design (with a non-randomly selected comparison group) or an experimental design (such as a randomized control trial, in which people are randomly allocated to the assisted and control groups): These alternative options can provide clearer insights into the attribution of impact. However, please note that for ethical and practical reasons, it is often not feasible to use control groups in humanitarian contexts.

An impact evaluation can be conducted by the team in charge of the intervention, but should be rigorously planned for together with the A&E team at the intervention design stage. Alternatively, an impact evaluation can be supported by an external third party or individual.

⁴ For further reading on evaluation types, see: International Federation of Red Cross and Red Crescent Societies (IFRC), *Project/programme monitoring and evaluation (M&E) guide*, 2011, p. 77: <https://www.fsnnetwork.org/sites/default/files/ifrc-me-guide-8-2011.pdf>; ALNAP, *Evaluation of Humanitarian Action Guide*, ALNAP, London, 2016, p. 78: <https://www.alnap.org/system/files/content/resource/files/main/alnap-evaluation-humanitarian-action-2016.pdf>; Better Evaluation, "Types of evaluation", 2021: https://www.betterevaluation.org/themes_overview#Types.

3.3 EVALUATIVE REVIEW

An evaluative review is a structured opportunity for reflection that tends to focus on learning lessons and on accountability to affected people. This type of exercise is not intended to contribute to accountability to donors. Evaluative reviews help strengthen staff capacity and increase ownership among the implementation team, and are therefore perceived as more subjective.

The key feature of evaluative reviews (unlike evaluations) is that they are conducted by an internal third party, or by an individual from the region or headquarters who is not involved in managing the intervention. An evaluative review does not replace an external evaluation. But it is an important component of ongoing monitoring, programme management and learning, helping the delegation to take stock of where things are at, and of what knowledge it has gained through the experience of implementing the intervention. Because these reviews are conducted internally, they tend to be smaller in scope than evaluations owing to capacity, resource and time constraints. They typically follow similar guidelines and processes as evaluations (such as using recognized quality criteria), but are not bound to do so.

3.4 LEARNING WORKSHOPS

Learning workshops are an alternative approach to answering evaluative review questions. In these structured review or debriefing exercises, the team responsible for implementing an intervention discusses its objectives and achievements. Participants consider and reflect together on what happened and why it happened, and how to sustain strengths and improve on weaknesses. These workshops provide instant feedback on change that can be made to improve ongoing implementation or future replication, with an emphasis on immediate lesson-learning rather than impact evaluation or accountability. Learning workshops can take place at any point during or shortly after an intervention. They are usually facilitated by an internal third party, or by an individual from the delegation, region or headquarters who is not involved in managing the intervention. The A&E team provides methodological support.

For further guidance, see the following resource prepared by the A&E team:

[*Managing Learning Workshops in the Field: A Step-by-Step Planning Guidance*](#)

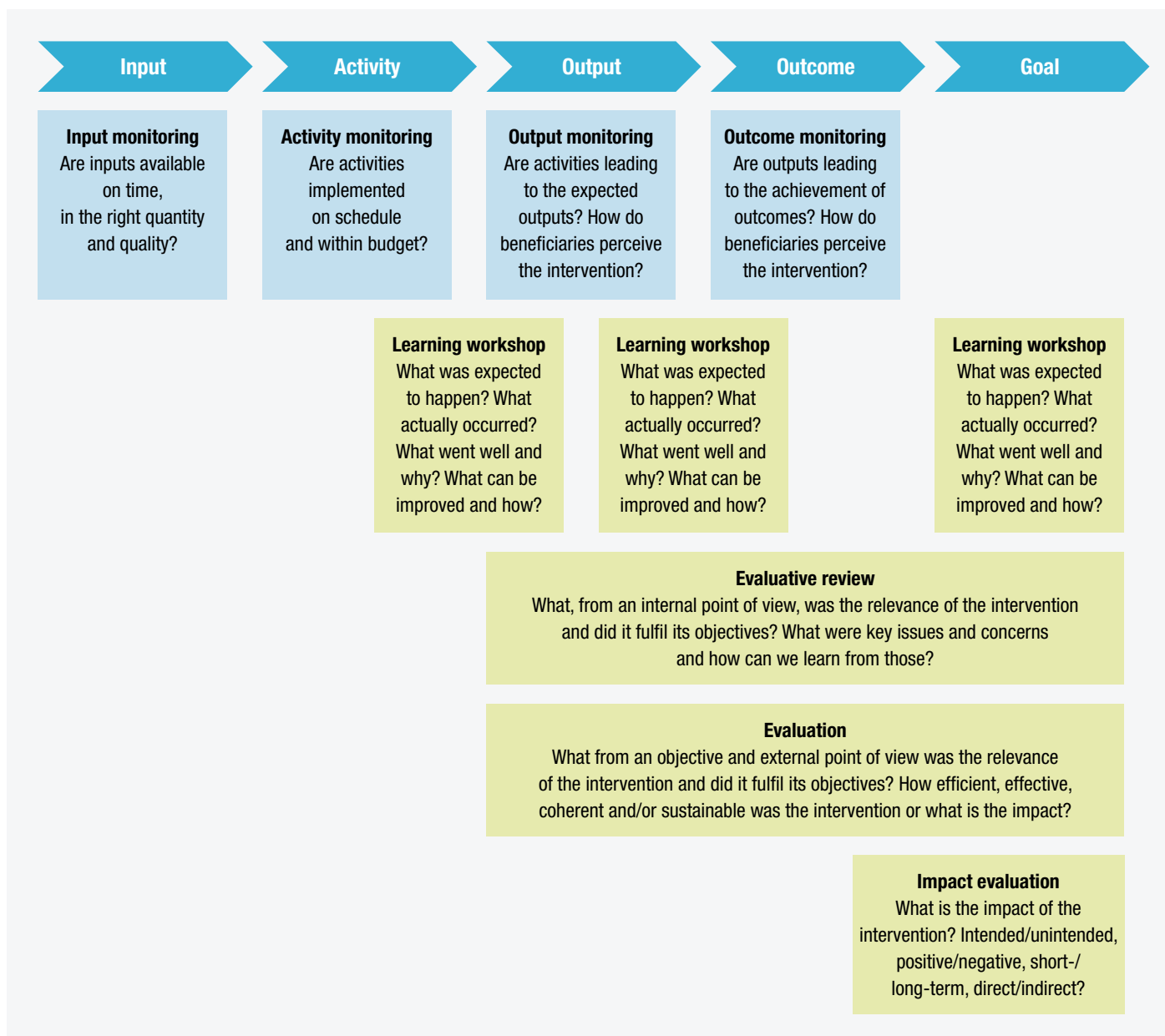
(internal ICRC document)

4. HOW DO EVALUATIONS RELATE TO MONITORING AND AUDITS?

Evaluations differ from monitoring and audits in terms of timing and focus:

- Monitoring is a continuous process that tracks progress (process, activity and results monitoring) or changes in the environment or context (situation monitoring) against specific indicators. Monitoring data informs ongoing implementation by tracking outputs, outcomes, budgets and alignment with procedures, and allows for adjustments to be made at short notice.
- Audits assess or verify compliance with established rules, regulations, procedures or mandates. An audit can be distinguished from an evaluation or review by its emphasis on assurance and compliance, rather than on making a judgement of worth. Audits are conducted by an internal or external third party or individual. At the ICRC, audits are conducted by the Ethics, Risk and Compliance Office.
- Evaluations are conducted at specific points in time to assess how well an intervention was managed and what difference it made. Evaluations may also inform implementation (e.g. a mid-term evaluation), but they are less frequent and examine larger changes that require more methodological rigour and analysis beyond existing monitoring frameworks. Monitoring typically provides data for evaluation, and aspects of evaluation occur when monitoring.

FIGURE 1: MONITORING AND EVALUATION TYPES ACROSS THE LOGICAL FRAMEWORK



5. MANAGING AN EVALUATION: A STEP-BY-STEP GUIDE

The ICRC follows a uniform process for managing all evaluations and reviews. This process necessary involves nine steps, with a corresponding set of mandatory deliverables. The level of rigour and detail for each step varies, however, according to the evaluation type.

STEP 1: EVALUABILITY ASSESSMENT

For further guidance, see the following resource prepared by the Evaluation Office:
[Evaluability Checklist](#) (internal ICRC document)

Step 1.1: Decide whether an evaluation is needed and what the objectives will be

Evaluations and evaluative reviews in the field are managed directly by the delegation, with oversight from the Evaluation Office. An evaluation might be triggered according to the ICRC's Evaluation Framework.

Decide on the objectives of the evaluation and whether the evaluation is needed for accountability or learning purposes:

- Accountability refers to our responsibility to report to and get feedback from others, such as donors, the affected population, other staff members including management and authorities.⁵ Typically, evaluations help to increase an organization's accountability.
- An evaluation can contribute to learning by examining what worked, what didn't and how performance can be improved.⁶ Here, the focus is on the initial implementation phase of an intervention. Typically, an evaluative review or learning workshop contributes to learning.

Usually, an evaluability assessment is carried out to provide a reasoned basis for proceeding with an evaluation. It assesses whether an evaluation can be conducted, and ensures that the evaluation will be both feasible and useful. Aspects to consider when assessing evaluability include the:

- overall level of ambition and the type of questions that the evaluation should answer
- programme design and intervention logic
- availability of data
- conduciveness of the context.⁷

Evaluations are not the only (or indeed the most cost-effective) way to promote accountability or organizational learning. Some alternative approaches are outlined below:⁸

- Ongoing dialogue, consultation and feedback mechanisms may be more effective ways of achieving accountability to affected people.
- Audits may be more appropriate, suitable approach if the focus is on managerial, financial or contractual accountability.
- Learning workshop is a faster and more cost-effective way to promote learning than a full evaluative review exercise.
- Remote evaluation options should be considered when security issues make it difficult or impossible to reach the affected people – and, as a general rule, other risks and potential solutions should be considered in challenging humanitarian contexts.⁹

⁵ ALNAP, 2016, p. 27.

⁶ ALNAP, 2016, p. 27.

⁷ ALNAP, 2016, pp. 53–54.

⁸ For alternative accountability and learning processes to consider, see: ALNAP, 2016, pp. 43, 46 and 49.

⁹ For evaluation challenges and solutions to consider, see: ALNAP, 2016, pp. 32 and 283.

Step 1.2: Identify the users of your evaluation and what they need to know

Identify the primary intended users of the evaluation and how the results will be used.¹⁰ For example, open-ended evaluations can help to improve an intervention by assessing its strengths and weaknesses, while also contributing to wider learning within the organization. Evaluations that focus on providing rapid and ongoing feedback in emergent and complex environments are useful for guiding direct adaptation of an intervention. And evaluations conducted after an intervention often aim to judge its impact and to decide whether to maintain funding.¹¹

List the people, units or organizations who will use the findings and who have the capacity to make decisions, act, and change policies or strategies. The intended users should be engaged throughout the evaluation process.

Step 1.3: Decide on the scope of the evaluation

Decide on the scope by identifying the focus of the evaluation (such as a project, programme, policy, strategy or thematic area, or technical assistance, policy advice, or another activity or action). An evaluation can focus on one or more of these aspects at once. Also specify the geographic focus, such as site-level, delegation-level, regional or global.

Step 1.4: Decide on the timing of the evaluation in relation to implementation

Decide whether the evaluation is intended to influence the intervention (e.g. mid-term) or whether it will take place after the intervention has finished. Typically, a learning-oriented evaluation will take place in the initial stages or during implementation, while an accountability-oriented evaluation will focus on later stages or occur after implementation.

Step 1.5: Pinpoint the level of results of interest

Decide whether you want to evaluate processes, outcomes or impact, and how far down the results chain you want to go. A useful starting point is the theory of change, which explains how the activities undertaken by the intervention contribute to a chain of results that lead to the intended or observed results.¹² Alternatively, other programme logic tools can be used, such as the conceptual framework, logical framework or results chain.

The theory of change helps to identify the evaluation questions and key indicators for monitoring and evaluation, to observe gaps in available data, and to prioritize data collection, analysis and reporting. In some circumstances, the original theory of change might need to be reviewed or even updated as part of the evaluation, such as where the context has changed significantly, where understanding of how the intervention works has evolved since the original theory of change was developed or where there is little evidence to support the current theory of change.

STEP 2: ESTABLISH AN EVALUATION REFERENCE GROUP

Identify who should carry out the evaluation.¹³ With the exception of impact evaluations, they should always be conducted by an individual or third party who is not involved in managing the intervention. Evaluations with a focus on accountability intend to bring an objective perspective and are carried out by an external individual or third party with no involvement in implementation, with oversight from the ICRC's Evaluation Office. Evaluations focused on lesson-learning, such as evaluative reviews and learning workshops, often produce internal documents. These are conducted or facilitated by an internal third party or individual (e.g. from the region or headquarters). An impact evaluation,

¹⁰ For a step-by-step approach to identifying, understanding and engaging intended users of the evaluation, see: ALNAP, 2016, p. 65; also see: Better Evaluation, "Identify who are the primary intended users of the evaluation and what will they use it for", 2021: <https://www.betterevaluation.org/en/node/5281>.

¹¹ International Development Research Centre (IDRC), *Identifying the Intended User(s) and Use(s) of an Evaluation*, IDRC, Ottawa, 2012: <https://www.betterevaluation.org/sites/default/files/idrc.pdf>.

¹² Better Evaluation, "Describe the theory of change", 2021: https://www.betterevaluation.org/en/managers_guide/step_2/describe_theory_of_change; ALNAP, 2016, p. 91.

¹³ ALNAP, 2016, p. 156; Better Evaluation, "Establish decision-making processes", 2021: https://www.betterevaluation.org/en/rainbow_framework/manage/establish_decision_making_processes; Better Evaluation, "Decide who will conduct the evaluation", 2021: https://www.betterevaluation.org/en/rainbow_framework/manage/decide_who_will_conduct_the_evaluation.

meanwhile, can be conducted by the team in charge of the intervention but should follow rigorous standards and requires involvement from the A&E team.

Clarify who will be involved in the various decisions involved in an evaluation, what their roles will be and what kind of decision-making authority they have based on the RACI matrix (which stands for “responsible, accountable, consulted and informed”).

Usually, an advisory group is set up to keep primary stakeholders engaged in the evaluation. This is typically a steering and/or management group consisting of the evaluation or review manager/team and several other colleagues, whose role is to guide an evaluation through the key stages (e.g. drafting the ToR, the inception report and the final report). For larger evaluations, it is helpful to have a reference group (including senior management) to advise on practical issues associated with the evaluation and on the feasibility of the resulting recommendations.¹⁴

Other key stakeholders of an evaluation are listed below:

- The evaluation or review team/manager manages the evaluation or evaluative review from the beginning to the end, ensures that the various stakeholders are involved, and forms a bridge between internal ICRC colleagues and external evaluators.
- The A&E specialist acts as the technical lead and provides support for impact evaluations.
- Senior management guides the direction of the evaluation (coordinator, delegation management, head of sector, etc.) and acts on recommendations from the evaluation, often as part of a reference group.
- The ICRC’s Evaluation Office provides oversight for the evaluation process, including technical advice, support and endorsement.
- A community consultation committee or key informants from the community should be involved at key stages in the evaluation process, such as the decision on whether, what and how to evaluate; they should also be consulted during the evaluation process and on validating the findings.¹⁵
- The intervention implementation team facilitates the evaluation process and acts on recommendations from the evaluation.

STEP 3: DEVELOP THE EVALUATION TOR

For further guidance, see the following resource prepared by the Evaluation Office:

[Checklist \(quality criteria\) for Evaluation Terms of Reference](#)

Also see this resource prepared by the A&E team:

[Review and Evaluation Terms of Reference \(ToR\) Template](#) (both internal ICRC documents)

Step 3.1: Develop agreed key evaluation questions

Evaluation questions derive from the purpose and objectives of the evaluation and draw on the intervention’s theory of change. The evaluation questions should be based on what the primary intended users need to know (i.e. what would make a difference in their work).¹⁶

In order to make the evaluation more targeted, it is useful to think practically about the specific questions the evaluation needs to answer before reviewing how these fit with the various criteria. Once the evaluation questions have been established, they can then be linked to the evaluation criteria, which provide a lens or perspective through which the intervention can be viewed (see [Annex 1](#)).¹⁷

14 ALNAP, 2016, p. 148; World Food Programme (WFP), *Technical Note: Evaluation Committee*, WFP, Rome, 2018: <https://docs.wfp.org/api/documents/WFP-0000003174/download/>; WFP, *Technical Note: Evaluation Reference Group*, WFP, Rome, 2018: <https://docs.wfp.org/api/documents/WFP-0000003175/download/>.

15 ALNAP, 2016, p. 268.

16 ALNAP, 2016, p. 104.

17 WFP, *Technical Note: Evaluation Questions and Criteria*, WFP, Rome, 2021: <https://docs.wfp.org/api/documents/WFP-0000003173/download/>.

Evaluation questions are initially defined as high-level questions, which can be further unpacked into sub-questions (including the questions that are asked to interviewees, focus groups and survey subjects in interviews, topic guides and survey instruments). An evaluation matrix¹⁸ is usually developed to aid this process.

The evaluation questions have a significant effect on the quality of the evaluation. They also guide what type of evaluation should be chosen, and its design, methodology and budget, as well as the shape of the recommendations from the evaluation.

Good high-level evaluation questions are usually limited in number (around seven), phrased as open questions, and specific enough to help focus the evaluation but broad enough to be broken down further into more detailed sub-questions to guide data collection.

The recommended approach is to work with the primary intended users (as well as with other key stakeholders of the evaluation) to develop an agreed list of key evaluation questions. Colleagues will typically propose a large number of questions, especially when the ToR is circulated for wider feedback. To avoid this, colleagues could be asked to stick to evaluation objectives in their proposals. These could then be further unpacked by the evaluation team in the inception report.

Step 3.2: Consider the evaluation framework and criteria

Frame the evaluation according to normative frameworks and criteria.¹⁹ It can be helpful to break the evaluation down into smaller tasks that are easier to manage and provide more structure. This exercise also determines the merit or worth of an intervention and serves as the basis upon which evaluative judgements are made, as well as providing a framework that helps with developing the main evaluation questions, undertaking the analysis and presenting the conclusions.

The most commonly used criteria are those developed by the OECD's DAC: relevance, coherence, effectiveness, efficiency, impact and sustainability.²⁰ Additional criteria (appropriateness, connectedness and coverage) are sometimes used. Other frameworks can be used on top of, or even in place of, the OECD criteria if appropriate. The objectives, priorities, scope and context of the intervention and the evaluation will shape the relative focus on different criteria.

The following principles should guide the use of the evaluation criteria to prevent a mechanistic approach:²¹

- Principle one: The criteria should be applied thoughtfully to support high-quality, useful evaluation. They should be contextualized – understood in the context of the individual evaluation, the intervention being evaluated and the stakeholders involved. The evaluation questions (what you are trying to find out) and what you intend to do with the answers should inform how the criteria are specifically interpreted and analysed.
- Principle two: The use of the criteria depends on the purpose of the evaluation. The criteria should not be applied mechanistically. Instead, they should be covered according to the needs of the relevant stakeholders and the context of the evaluation. More or less time and resources may be devoted to the evaluative analysis for each criterion depending on the evaluation purpose. Data availability, resource constraints, timing and methodological considerations may also influence how (and whether) a particular criterion is covered.

18 WFP, *Technical Note: Evaluation Matrix*, WFP, Rome, 2020: <https://docs.wfp.org/api/documents/WFP-0000003176/download/>.

19 ALNAP, 2016, p. 101; OECD Development Assistance Committee (DAC), "Evaluation Criteria", 2021b: <https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>.

20 OECD, 2021a.

21 OECD, 2021a.

Teams can identify relevant criteria by considering the questions in Figure 2 below:

FIGURE 2: QUESTIONS TO HELP ESTABLISH THE INITIAL EVALUATION CRITERIA²²

QUESTION	CRITERION
Did the intervention take place in a complex and sensitive environment?	Relevance
Did the intervention take place in a crowded institutional space?	Coherence
Should the evaluation measure the delivery of the intervention's objectives?	Effectiveness
Is it valuable to establish how the intervention's funding was utilized?	Efficiency
Is it worthwhile establishing the intervention's (un)intended higher-level effects?	Impact
Is it possible to establish if the effects of the intervention will continue?	Sustainability

Step 3.3: Develop the ToR document and budget

At this stage, the ToR should be drafted based on the outcomes of the previous steps.²³ This document forms the basis of the contract between the ICRC team commissioning the evaluation and the evaluation team. As such, the quality of an evaluation depends not only on the technical skills of the evaluator, but also on the decisions made when drafting the ToR.

In most cases, the ToR document will be further discussed and refined with the evaluators. Alternatively, the evaluation team may use the inception report to determine how they will address the ToR.

Importantly, even internal reviews require a ToR document, since it clearly sets out the objective, purpose and scope of the review and prevents misunderstandings from arising later in the process.

A key driver of the ToR is the budget, since the resources available for the evaluation will influence the objectives, scope and design of the evaluation, and vice versa. A good budget usually covers the preparation, inception, fieldwork and post-fieldwork phases.

The team drafting the ToR will need to engage with the key stakeholders of the evaluation. If an external evaluator is being sought, consideration should be given to the tendering process (including advertising the ToR, selecting an evaluator and preparing the contract).

STEP 4: RECRUIT AN APPROPRIATE EVALUATION TEAM OR EVALUATORS

For further guidance, see the following resource prepared by the Evaluation Office:
[International Humanitarian Evaluation Profession Standards](#) (internal ICRC document)

Recruit an evaluation team or evaluators with the appropriate structure according to the evaluation type and ToR. The degree of independence of the evaluator(s) should be determined (external to the organization, or internal to the organization but external to the intervention) in order to avoid potential conflicts of interest, or the risk that independence, impartiality or neutrality could be compromised. The contract with the evaluator(s) must follow the ICRC's procurement guidelines and procedures and must include a termination clause should performance be unsatisfactory at the inception phase.

²² OECD, 2021a.

²³ WFP, "Decentralized Evaluation Terms of Reference" (template), 2016: <https://docs.wfp.org/api/documents/4970fed797bc4d1099df259a92c632ee/download/>; ALNAP, 2016, p. 118; Better Evaluation, "Develop the Terms of Reference (ToR)", 2021: https://www.betterevaluation.org/en/commissioners_guide/step3; Better Evaluation, "Engage the evaluation team", 2021: https://www.betterevaluation.org/en/commissioners_guide/step4.

STEP 5: OVERSEE THE INCEPTION PHASE

For further guidance, see the following resource prepared by the Evaluation Office:
[Checklist \(quality criteria\) for Inception Reports](#) (internal ICRC document)

The evaluation methodology is a road map for implementing the evaluation, setting out the building blocks for how data will be collected and analysed. A good methodology will help to assess what implications data collection and analysis will have for implementation of the intervention, whether data collection can still be integrated from the beginning of the programme, and how the evaluation questions will be operationalized. A clear methodology also helps to ensure that all key stakeholders are on the same page, even when the evaluation is managed internally.

If the evaluation is conducted by an external evaluator, these aspects are usually captured in the inception report.²⁴ It is important to oversee the inception phase as part of the evaluation process and its alignment with the guidelines and quality criteria set out by the Evaluation Office. An inception report is produced following the first round of document review and discussion with the reference group and other evaluation stakeholders. The inception phase secures a consensus on the scope, objectives, methodology and potential limitations/risks of the exercise and documents any departures or changes from the ToR. The draft inception report is reviewed through the evaluation quality assurance mechanism. The inception report, which usually captures various topics, is validated by the reference group, confirming that there is a common understanding of the assignment.

An inception report normally includes an evaluation matrix, which sets out a plan for how each of the evaluation questions will be answered. A typical evaluation matrix contains several columns, including the evaluation questions, sub-questions and/or criteria, indicators, data collection methods and data sources.²⁵

One of the methodological aspects that the inception report will cover is the evaluation design,²⁶ possibly including experimental designs for impact evaluations, which usually involve the selection of control groups at the design stage or of a counterfactual group. While experimental approaches have advantages, they are also bound by certain constraints that limit their use in the contexts in which the ICRC (see [section 3.2](#)). As a result, most evaluations are non-experimental in design, although these still capture information about cause and effect to some degree.

Quantitative and qualitative data-collection methods are normally selected, depending on the objectives and approaches of the evaluation. A mixed-method approach is recommended to provide insights from different angles, including the use of both primary and secondary data.

STEP 6: OVERSEE THE EVALUATION PHASE

An evaluation plan should be developed, specifying the sequence of evaluation activities and clearly spelling out the roles and responsibilities of the evaluator (or the various members of the evaluation team) and everyone else involved in the evaluation process.²⁷ The evaluation manager will oversee the implementation of the evaluation plan and will help to problem-solve if any issues come up, including communicating with all stakeholders.

It is important to agree on how unforeseen issues that could affect the implementation of the evaluation and challenge milestones and/or deliverables will be raised, escalated and resolved.

²⁴ ALNAP, 2016, p. 145.

²⁵ WFP, 2020.

²⁶ WFP, [Technical Note: Evaluation Methodology](https://docs.wfp.org/api/documents/704ec01f137d43378a445c7e52dcf324/download/), WFP, Rome, 2016: <https://docs.wfp.org/api/documents/704ec01f137d43378a445c7e52dcf324/download/>; ALNAP, 2016, p. 194.

²⁷ Better Evaluation, “Manage implementation of the evaluation”, 2021: https://www.betterevaluation.org/en/commissioners_guide/step7.

STEP 7: HAVE THE REFERENCE GROUP VALIDATE THE FINAL EVALUATION REPORT

For further guidance, see the following resource prepared by the Evaluation Office:
[Checklist \(quality criteria\) for Evaluation Reports](#) (internal ICRC document)

The outputs of an evaluation are usually the inception report, the debriefing workshop reports, the advice provided by the evaluation team directly in the field and the evaluation report. The evaluation report contains the findings, conclusions and recommendations of the evaluation. A good report is succinct and clearly communicates findings and conclusions. Enough time should be set aside for the draft report to be circulated among key stakeholders for comment, which can help to build ownership of the results. How the evaluation findings are actually used will depend on how well the report meets the information needs of the primary intended users. It is therefore important to discuss the evaluation report in the initial planning phase, as well as in the final stages of the evaluation.

The final evaluation report needs to be validated by the reference group (in line with the guidelines and quality criteria set out by the Evaluation Office). A documented evaluation management response from the parties responsible for the recommendations is attached to the report. The final report is included in a centralized report library, with the management response attached, so that anyone can read it.

STEP 8: REVIEW PROGRESS ON IMPLEMENTING THE RECOMMENDATIONS

Following up on the organization's response to an evaluation is key to the use and take-up of the findings.²⁸ This is often the responsibility of management and/or the staff managing the intervention. Time should be built into the evaluation plan and budget for the evaluator to provide support beyond just delivering the evaluation report. Some concrete examples of ways to support the use of evaluation's findings are set out below:

- Involve key users in key steps of the evaluation process, including formulating the recommendations, in order to build ownership (e.g. through workshops).
- Identify champions within the organization to support implementation of the recommendations.
- Include findings from the evaluation in new or existing guidance, ToR, templates and training packages.
- Allocate responsibility for taking up the recommendations through an evaluation response matrix.

An evaluation response matrix helps to foster discussion among management and other intended users of the evaluation (e.g. staff managing the intervention) about how to implement the recommendations in practice. It also ensures accountability. It is good practice to agree upon the use of an evaluation response matrix in the evaluation ToR, to identify a focal point to coordinate the evaluation response, and to establish an agreed timeline and format for management and staff managing the intervention to provide their formal response to the evaluation.

A good evaluation response matrix denotes whether management accepts or rejects each recommendation and identifies what actions will be taken to implement it, including deadlines, roles and responsibilities. If a recommendation is rejected, the reason should be given in the matrix. The evaluation response matrix should also entail a monitoring plan, indicating a focal point and a timeline for monitoring implementation of the proposed actions (e.g. after six months or one year). The matrix should be disseminated alongside the evaluation report. A sample is given in Figure 3 below.

²⁸ ALNAP, 2016, p. 346.

FIGURE 3: EVALUATION RESPONSE MATRIX²⁹

RECOMMENDATION	FURTHER FUNDING REQUIRED?	MANAGEMENT RESPONSE	COMMENT	ACTION TO BE TAKEN	TIMING	RESPONSIBLE PERSON/UNIT
Recommendation 1	Yes/No	Accept/ partially accept/ reject				
Recommendation 2	Yes/No	Accept/ partially accept/ reject				

STEP 9: DISSEMINATE THE EVALUATION REPORT IN ACCORDANCE WITH THE ICRC'S POLICIES AND OBJECTIVES

The evaluation and its key findings and recommendations should be communicated to key stakeholders as part of a dissemination strategy. The findings can also be shared with individuals and entities other than key stakeholders of the evaluation, such as staff or organizations working in the same field. It may be worthwhile sharing some of the findings through articles or stories to attract wider attention.

The reports of all external evaluations are published on the ICRC's website in accordance with the organization's Access to Information Policy. In exceptional cases, the Evaluation Office will publish the executive summary of an evaluation report in lieu of the full document, if the content threatens the rights and security of individuals, or compromises the safety and integrity of the ICRC's operations.

6. FURTHER READING

The document is based on the following key documents, which can be accessed for further reading:

- ALNAP, [Evaluation of Humanitarian Action Guide](#), ALNAP, London, 2016
- Better Evaluation, ["Manage implementation of the evaluation"](#), 2016
- ICRC, [Analysis & Evidence Toolkit](#) (internal ICRC resource)
- ICRC, [ICRC Evaluation Strategy 2022–2024](#), 2022
- ICRC, [ICRC guidance on the Evaluation Community Site](#) (internal ICRC resource)
- IFRC, [Project/programme monitoring and evaluation \(M&E\) guide](#), 2011
- OECD, [Applying Evaluation Criteria Thoughtfully](#), OECD, Paris, 2021
- WFP, [Decentralized Evaluation: Guidance for Process and Content](#), WFP, Rome, 2021
- WFP, [Technical Note: Evaluation Approaches, Methods and Data Collection Tools for Decentralized Evaluations](#), WFP, Rome, 2021

²⁹ Adapted from ALNAP, 2016, p. 348.

ANNEX 1: EVALUATION QUESTIONS ORGANIZED BY CRITERION

POTENTIAL EVALUATION QUESTIONS ORGANIZED BY CRITERION³⁰

CRITERION AND DEFINITION	INCLUDES ANALYSIS OF	POTENTIAL EVALUATION QUESTIONS
<p>Relevance: Is the intervention doing the right thing?</p> <p>The extent to which the intervention objectives and design respond to beneficiaries', global, country and partner/institution needs, policies and priorities, and continue to do so if circumstances change.</p>	<p>Relevance of the intervention design to the needs and priorities of the most vulnerable groups.</p> <p>Continued relevance of the objectives over the life of the intervention, or its ability to adapt to new needs if circumstances change.</p> <p>Alignment with government, partner and/or donor policies and interventions; alignment and coherence with the organization's policies.</p> <p>Consistency of intervention design and logic.</p> <p>Extent to which design and implementation were gender-sensitive, based on gender analysis and addressed diverse needs.</p> <p>Extent to which the design and implementation of the intervention were sensitive to the capacities in place.</p> <p>Differences and trade-offs between different priorities or needs.</p>	<p>How was the design of the intervention relevant to the wider context?</p> <p>How was the intervention in line with the needs and priorities of the most vulnerable groups (men and women, boys and girls)?</p> <p>Was the affected population included in intervention design, and how was their feedback sought throughout?</p> <p>How were the intervention design and objectives aligned with the needs of the government?</p> <p>How was the intervention aligned with partners', UN agencies' and donors' policies and priorities?</p> <p>Was the intervention based on a sound gender analysis? How?</p> <p>Was the design and implementation of the intervention gender-sensitive? How?</p> <p>How did the design and implementation of the intervention consider the available capacities?</p>
<p>Coherence: How well does the intervention fit?</p> <p>The compatibility of the intervention with other interventions in a country, sector or institution.</p>	<p>Contextual factors and how they influenced the design/implementation of the subject.</p> <p>Links to relevant policies and programmes of other actors.</p> <p>Consideration of humanitarian and human rights principles and standards, including gender equality and women's empowerment and wider equity issues.</p> <p>Extent to which other interventions (particularly policies) support or undermine the intervention, and vice versa.</p> <p>Synergies and interlinkages between the intervention and other interventions carried out by the same institution/government, as well as the consistency of the intervention with the relevant international norms and standards to which that institution/government adheres (internal coherence).</p> <p>Consistency of the intervention with other actors' interventions in the same context (external coherence).</p>	<p>To what extent were context factors (political stability/instability, population movements, etc.) considered in the design and delivery of the intervention?</p> <p>To what extent was the organization's intervention coherent with policies and programmes of other partners operating within the same context?</p> <p>To what extent was the intervention design and delivery in line with humanitarian principles?</p> <p>What have been the synergies between the intervention and other interventions of the organization?</p>

³⁰ Adapted from WFP, 2021; OECD, 2021b; ALNAP, 2016.

CRITERION AND DEFINITION	INCLUDES ANALYSIS OF	POTENTIAL EVALUATION QUESTIONS
<p>Efficiency: How well are resources used?</p> <p>The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.</p>	<p>Costs per recipient for different implementation mechanisms/modes of intervention.</p> <p>Timeliness of delivery, compliance with intended timeframes or budgets, comparison of channels of delivery (e.g. schools/health systems versus community-based).</p> <p>Comparison of different institutional arrangements (e.g. continuity of supplies and use of local partners/systems/procurement where feasible).</p>	<p>To what extent was the intervention cost-efficient?</p> <p>Was the intervention implemented in a timely way? How?</p> <p>Was the intervention implemented in the most efficient way compared to alternatives? How?</p> <p>Did the targeting of the intervention mean that resources were allocated efficiently? How?</p>
<p>Effectiveness: Is the intervention achieving its objectives?</p> <p>The extent to which the intervention achieved, or is expected to achieve, its objectives and its results, including any differential results across groups.</p>	<p>Achievement of objectives (or likelihood that the objectives will be achieved), taking account of the relative importance of the objectives or results.</p> <p>Main results including positive, negative, intended and unintended outcomes.</p> <p>Outputs and outcomes for men, women, boys and girls, and other relevant socio-economic categories.</p> <p>Potential constraints and facilitating factors to achievements.</p>	<p>Were (are) the outputs and outcomes achieved (likely to be achieved)? How?</p> <p>What major factors influenced the achievement or non-achievement of the outcomes?</p> <p>Were there unintended (positive or negative) outcomes of assistance for participants and non-participants? Why?</p> <p>Is the achievement of outcomes leading to likely to lead to meeting intervention objectives? What major factors influenced this?</p> <p>How were results delivered for men, women, boys and girls?</p> <p>How were relevant assistance standards met?</p>
<p>Impact: What difference is the intervention making?</p> <p>The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.</p>	<p>Extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects (e.g. holistic and enduring changes in the systems or norms, and potential effects on people's well-being, human rights, gender equality and the environment).</p> <p>Ultimate significance and potentially transformative effects of the intervention (e.g. social, environmental and economic effects that are longer-term or broader in scope than those that are already captured under the effectiveness criterion).</p>	<p>What were the effects of the intervention on participants' lives (intended and unintended)?</p> <p>Did a specific part of the intervention achieve greater impact than another? Why?</p> <p>Were there any gender-specific impacts? Did the intervention influence the gender context? How?</p> <p>Were there impacts on institutions? How?</p> <p>How did the intervention contribute to long-term intended results?</p>
<p>Sustainability: Will the benefits last?</p> <p>The extent to which the net benefits of the intervention continue or are likely to continue.</p>	<p>Capacity strengthening/development results.</p> <p>Institutional/systemic changes.</p> <p>Integration of intervention elements into national systems and processes.</p> <p>The financial, economic, social, environmental and institutional capacities of the systems needed to sustain net benefits over time.</p> <p>Resilience, risks and potential trade-offs.</p>	<p>To what extent did the intervention implementation consider sustainability, such as capacity strengthening of national and local government institutions, communities and other partners?</p> <p>To what extent did intervention benefits continue after the organization's work ceased?</p> <p>To what extent is it likely that the benefits of the intervention will continue after the organization's work ceases?</p> <p>Has the intervention made any difference to gender relations in the medium or longer term? How?</p>

In addition, evaluations of humanitarian interventions also use the criteria of appropriateness, connectedness and coverage.

POTENTIAL EVALUATION QUESTIONS ORGANIZED BY CRITERION³¹

CRITERION AND DEFINITION	INCLUDES ANALYSIS OF	POTENTIAL EVALUATION QUESTIONS
<p>Appropriateness The extent to which humanitarian activities are tailored to local needs, increasing ownership, accountability and cost-effectiveness accordingly. If used, this criterion replaces the OECD DAC criterion of Relevance.</p>	<p>Extent to which the organization's inputs were tailored to needs.</p> <p>Extent to which they were adapted to respond to the changing demands of unstable environments.</p> <p>Extent to which design and implementation were gender-sensitive, based on gender analysis.</p>	<p>How was the chosen intervention approach the best way to meet the needs of affected populations and intended beneficiaries?</p> <p>Were the adopted procedures and arrangements the best way of meeting recipients' needs? Why?</p> <p>How were protection and ethics issues considered in design and implementation?</p> <p>To what extent was the intervention based on a sound gender analysis?</p> <p>To what extent was the design and implementation of the intervention gender-sensitive (i.e. gender equality and women's empowerment issues were considered)?</p>
<p>Coverage The degree to which major population groups facing life-threatening suffering, wherever they are, have been provided with impartial assistance and protection, proportionate to need. Requires an analysis of differential coverage/targeting, and inclusion and exclusion impacts on population sub-groups (gender, ethnicity, location, family circumstance).</p>	<p>Extent to which different groups are targeted or included.</p> <p>Impact of exclusion on sub-groups (gender, ethnicity, location, family circumstance).</p> <p>Differentiation of targeting and forms/amount of assistance provided.</p>	<p>How were the humanitarian needs of key target groups (men and women, boys and girls) met by the intervention?</p> <p>How was the organization's support provided proportionally according to the needs within the context?</p> <p>How did different geographical areas or groups of populations affected differently receive assistance according to their needs?</p> <p>How were relevant assistance standards met?</p> <p>How was the organization's support coordinated with that provided by others (duplication/gaps)?</p>
<p>Connectedness The degree to which activities of a short-term emergency nature are carried out in a way that takes longer-term and interconnected problems into account (e.g. refugee/host community issues; relief and resilience). Can be applied alongside or instead of the OECD DAC criterion of Sustainability.</p>	<p>Consistency between short-term activities and other development interventions/goals that address contextual problems.</p> <p>Presence of transition-focused analyses such as stakeholder consultations, and existence of a transition strategy.</p>	<p>What linkages have there been between the intervention and any other organizational interventions?</p> <p>To what extent did the intervention link to any transition strategies in the context or to development goals?</p>





³¹ Adapted from WFP, 2021; OECD, 2021b; ALNAP, 2016.

The ICRC helps people around the world affected by armed conflict and other violence, doing everything it can to protect their lives and dignity and to relieve their suffering, often with its Red Cross and Red Crescent partners. The organization also seeks to prevent hardship by promoting and strengthening humanitarian law and championing universal humanitarian principles.

People know they can count on the ICRC to carry out a range of life-saving activities in conflict zones and to work closely with the communities there to understand and meet their needs. The organization's experience and expertise enables it to respond quickly and effectively, without taking sides.



International Committee of the Red Cross
19, avenue de la Paix
1202 Geneva, Switzerland
T +41 22 734 60 01
shop.icrc.org
© ICRC, June 2022

 ecosecanalysis@icrc.org
 facebook.com/icrc
 twitter.com/icrc
 instagram.com/icrc