



## Technical Note

# Evaluation Questions and Criteria

1

Introduction

2

International Evaluation Criteria as a value framework to guide evaluation

3

Setting appropriate evaluation questions for WFP evaluations

Annex I

Using the International Evaluation Criteria to develop evaluation questions

## 1. Introduction

This Technical Note is intended to guide Evaluation Managers (EMs) in the formulation of clear and utility-focussed evaluation questions. These should be directly deriving from the purpose and objectives of the evaluation (“What does the evaluation aim to find out? And for what purpose?”) and draw from the intervention’s Theory of Change. Developing appropriate and relevant questions is fundamental to deliver high quality, credible and useful evaluations. The evaluation questions will in turn inform the selection of appropriate evaluation approaches and methodologies.

Evaluation questions should be structured around international evaluation criteria. “Each criterion is a different lens or perspective through which the intervention can be viewed. Together, they provide a more comprehensive picture of the intervention, the process of implementation, and the results.”<sup>1</sup>

For further support in the development of evaluation questions of a decentralized evaluation, EMs are encouraged to consult their Regional Evaluation Officer (REO) if they are based in a country office or OEV Decentralized Evaluation Helpdesk if they are based in a Headquarter Division.<sup>2</sup>

<sup>1</sup> OECD/DAC 2019 Better Criteria for Better Evaluation: Revised Evaluation Criteria, Definitions and Principles for Use. <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf>

<sup>2</sup> OEV Decentralized Evaluation Helpdesk can be contacted at the following email: [wfp.decentralizedevaluation@wfp.org](mailto:wfp.decentralizedevaluation@wfp.org)

## 2. International Evaluation Criteria as a value framework to guide evaluation

The Economic Co-operation and Development's (OECD) Development Assistance Committee (DAC) first outlined a set of evaluation criteria in 1991 and reviewed them in 2019 taking stock of experience and learning. These international evaluation criteria form a core reference for evaluating development and humanitarian interventions and include **relevance, coherence, effectiveness, efficiency, impact** and **sustainability**. In addition, evaluations of humanitarian interventions should also apply the criteria of **appropriateness, connectedness** and **coverage**.

International evaluation criteria set the *values* and the *framework* that will guide an evaluation. They provide a frame that helps to develop the main evaluation questions, undertake the analysis and present the conclusions. The following principles should guide the use of the evaluation criteria to prevent a mechanistic approach:

- **Principle One:** The criteria should be applied thoughtfully to support high quality, useful evaluations. 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 analyzed.
- **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 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.<sup>3</sup>

If a particular criterion is not applied, the TOR, inception and evaluation reports should provide an explanation and justification for the exclusion. Table 1 sets out the definitions for each criterion, while Annex I provides additional information in relation to each criterion to explain the concepts.

---

<sup>3</sup> Evaluability assessments (carried out before an evaluation begins) can be useful in setting realistic expectations for what information the evaluation can provide, what evidence can be gathered, and how the evaluation will answer questions.

**Table 1: Definitions for international evaluation criteria**

<b>Criteria</b>	<b>Definition<sup>4</sup></b>
<b>Relevance:</b> <i>Is the intervention doing the right thing?</i>	<i>The extent to which the intervention objectives and design respond to beneficiaries<sup>5</sup>, global, country and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.</i>
<b>Coherence:</b> <i>How well does the intervention fit?</i>	<i>The compatibility of the intervention with other interventions in a country, sector or institutions.</i>
<b>Efficiency:</b> <i>How well are resources used?</i>	<i>The extent to which the intervention delivers, or is likely to deliver, results in an economic<sup>6</sup> and timely<sup>7</sup> way.</i>
<b>Effectiveness:</b> <i>Is the intervention achieving its objectives?</i>	<i>The extent to which the intervention achieved, or is expected to achieve, its objectives and its results, including any differential results across groups.</i>
<b>Impact:</b> <i>What difference is the intervention making?</i>	<i>The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.</i>
<b>Sustainability:</b> <i>Will the benefits last?</i>	<i>The extent to which the net benefits of the intervention continue, or are likely to continue.</i>
<b>Further criteria for interventions in Humanitarian settings<sup>8</sup>:</b>	
<b>Appropriateness</b>	<i>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.</i>
<b>Coverage</b>	<i>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 analysis of differential coverage/ targeting, inclusion and exclusion impacts on population sub-groups (gender, ethnicity, location, family circumstance).</i>
<b>Connectedness</b>	<i>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 as part of or replacing the criterion of Sustainability.</i>

<sup>4</sup> OECD/DAC 2019 Better Criteria for Better Evaluation: Revised Evaluation Criteria, Definitions and Principles for Use.

<sup>5</sup> OECD/DAC 2019 OECD/DAC 2019 Better Criteria for Better Evaluation: Beneficiaries is defined as, "the individuals, groups, or organisations, whether targeted or not, that benefit directly or indirectly, from the development intervention." Other terms, such as rights holders or affected people, may also be used.

<sup>6</sup> "Economic" is the conversion of inputs (funds, expertise, natural resources, time, etc.) into outputs, outcomes and impacts, in the most cost-effective way possible, as compared to feasible alternatives in the context.

<sup>7</sup> "Timely" delivery is within the intended timeframe, or a timeframe reasonably adjusted to the demands of the evolving context. This may include assessing operational efficiency (how well the intervention was managed).

<sup>8</sup> Adapted from Evaluation of Humanitarian Action Guide, ALNAP 2016.

### 3. Setting appropriate evaluation questions for WFP evaluations

Defining appropriate evaluation questions is a key step during the evaluation design and preparation, since evaluation questions guide every step of the evaluation design. The evaluation questions should be developed in line with the purpose and objectives of the evaluation, the commissioning office's needs, and the interests of key stakeholders. The appropriate criteria for the evaluation will need to be selected in alignment with the evaluation questions.

The evaluation manager should set out intended questions in the TOR, but expect that they will be refined and unpacked during inception together with the evaluation team when data availability and context become clearer. Annex I shows how the international evaluation criteria can support the development of evaluation questions.

Good evaluation questions should be:

- **Well-defined and specific** to the timing, objectives of the evaluation, and to how the evaluation findings will be used, and by whom. For example, an evaluation primarily seeking to inform replication of an intervention in a new context will have different questions from an evaluation that asks whether the intervention covered all vulnerable groups, or whether an intervention was cost-effective.
- **Drawn from the intervention Theory of Change (ToC)**, to ensure specificity to the context and the intervention. In the absence of a ToC, the evaluation manager may consider investing time with programme colleagues to reconstruct the ToC, or alternatively require the evaluation team to re-construct it as a first step during the inception phase.
- Go beyond assessing if intended results were achieved (e.g. how many people were reached with what quantity of food), by seeking to **explain why and how the project achieved or did not achieve its results**, to promote lesson-learning.
- **Relevant to the intervention and to users' needs**. An evaluation might also ask additional important evaluation questions that are outside the framework (for example, on equity and human rights).
- **Prioritised**, given that evaluations are limited in time and resources. A few, strategically designed and well-defined evaluation questions are better than many questions that may duplicate or contradict each other.
- Developed applying a **gender** lens by considering **gender equality and empowerment of women** within each of the criteria or as a standalone criterion, and in the evaluation questions, to ensure that the evaluation assesses the inclusion of gender dimensions in the intervention design and implementation<sup>9</sup>.

Collectively, the evaluation questions should be designed to give evaluation users the information they need to make strategic and/or operational decisions, take action, or learn from the intervention. The EM should discuss the evaluation questions with stakeholders to ensure shared

---

<sup>9</sup> In accordance with "Integrating Human Rights and Gender Equality in Evaluation Towards UNEG Guidance, 2011", page 30, table 2.4. More information on how to do this can be found in the [Technical Note on integrating gender in WFP evaluations, Checklist](#), and [Quick Guide](#).

understanding on what the evaluation can and will answer, often by seeking input from the Evaluation Reference Group (ERG)<sup>10</sup> on draft TORs and, later, the draft Inception Report (IR).

The EM and the ERG should ensure throughout the evaluation that the evaluation team is focused on the evaluation criteria and on answering the evaluation questions, and that the methods, data collection tools, and analysis are systematically linked to the evaluation questions. This can be effectively aided by using the Evaluation Matrix (see [Technical Note on Evaluation Matrix](#)). Ultimately, the evaluation **findings and conclusions should be reported against the evaluation questions and criteria**.

**Box 1: Further reading – Evaluation Criteria and Questions**

- [OECD/DAC Network on Development Evaluation, Webinar 2020: Evaluation Criteria, What has changed, why and why does it matter?](#)
- [OECD/DAC 2019 Better Criteria for Better Evaluation: Revised Evaluation Criteria, Definitions and Principles for Use](#)
- ALNAP, Cosgrave J. and Buchanan-Smith M., 2016, [Evaluating Humanitarian Action Guide](#)
- Better Evaluation: [Specify the Key Evaluation Questions](#)
- Better Evaluation: [Develop agreed key evaluation questions](#)
- CDC, [Good Evaluation Questions: A Checklist to Help Focus Your Evaluation](#)
- UNEG, 2018, [Integrating Gender Equality and Human Rights in Evaluation - UN-SWAP Guidance, Analysis and Good Practices](#)

---

<sup>10</sup> The composition and exact title of the Evaluation Reference Group vary from one type of evaluation to another.

## Annex I: Using the International Evaluation Criteria to develop evaluation questions

Criterion	Includes analysis of:	Potential evaluation questions
<b>Relevance</b>	<ul style="list-style-type: none"> <li>• Relevance of the intervention design to the needs and priorities of the most vulnerable groups.</li> <li>• Continued relevance of the objectives over the life of the intervention, or its ability to adapt to new needs if circumstances change.<sup>11</sup></li> <li>• Alignment with government, partners, donors' policies and interventions; alignment and coherence with WFP policies.</li> <li>• Consistency of intervention design and logic.</li> <li>• Extent to which design and implementation were gender-sensitive, based on gender analysis, and addressed diverse needs.</li> <li>• Extent to which the design and implementation of the intervention were sensitive to the capacities in place.</li> <li>• Differences and trade-offs between different priorities or needs.</li> </ul>	<p>To what extent:</p> <ul style="list-style-type: none"> <li>• Was the design of the intervention relevant to the wider context?</li> <li>• Is the intervention in line with the needs and priorities of the most vulnerable groups (men and women, boys and girls)?</li> <li>• Is the intervention design and objectives aligned with the needs of the government?</li> <li>• Is the intervention aligned with WFP, partners, UN agencies and donor policies and priorities?</li> <li>• Was the intervention based on a sound gender analysis?</li> <li>• Was the design and implementation of the intervention gender-sensitive?</li> <li>• Did the design and implementation of the intervention consider the available capacities?</li> </ul>
<b>Coherence</b>	<ul style="list-style-type: none"> <li>• Contextual factors and how they influenced the design/ implementation of the subject.</li> <li>• Links to the food security and nutrition policies and programmes of other actors.</li> <li>• Consideration of humanitarian and human rights principles and standards, including gender equality and women empowerment and wider equity issues.</li> <li>• The extent to which other interventions (particularly policies) support or undermine the intervention, and <i>vice versa</i>.</li> <li>• The 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).</li> <li>• The consistency of the intervention with other actors' interventions in the same context (external coherence).<sup>12</sup></li> </ul>	<ul style="list-style-type: none"> <li>• To what extent were context factors (political stability/instability, population movements, etc.) considered in the design and delivery of the intervention?</li> <li>• To what extent was WFP's intervention coherent with policies and programmes of other partners operating within the same context?</li> <li>• To what extent was the intervention design and delivery in line with humanitarian principles?</li> <li>• What have been the synergies between the intervention and other WFP interventions?</li> </ul>
<b>Efficiency</b>	<ul style="list-style-type: none"> <li>• Costs per recipient for different implementation mechanisms/mode of transfer -food/cash/voucher.</li> <li>• Timeliness of delivery, compliance with intended timeframes or budgets, comparison</li> </ul>	<ul style="list-style-type: none"> <li>• Was the intervention cost-efficient?</li> <li>• Was the intervention implemented in a timely way?</li> <li>• Was the intervention implemented in the most efficient way compared to alternatives?</li> </ul>

<sup>11</sup> The objectives and design of the intervention should be sensitive to the economic, environmental, equity, social, political economy, and capacity conditions in which it takes place.

<sup>12</sup> This includes complementarity, harmonisation and co-ordination with others, and the extent to which the intervention is adding value while avoiding duplication of effort.

Criterion	Includes analysis of:	Potential evaluation questions
	<ul style="list-style-type: none"> <li>of channels of delivery (e.g. schools/health systems versus community-based).</li> <li>Comparison of different institutional arrangements (e.g. continuity of supplies and use of local partners / systems / procurement where feasible).</li> </ul>	<ul style="list-style-type: none"> <li>Did the targeting of the intervention mean that resources were allocated efficiently?</li> </ul>
<b>Effectiveness</b>	<ul style="list-style-type: none"> <li>Achievement of objectives (or likelihood that the objectives will be achieved), taking account of the relative importance of the objectives or results.</li> <li>Main results including positive, negative, intended and unintended outcomes.</li> <li>Outputs and outcomes for men, women, boys and girls, and other relevant socio-economic categories.</li> <li>Potential constraints and facilitating factors to achievements.</li> </ul>	<p>To what extent:</p> <ul style="list-style-type: none"> <li>Were (are) the outputs and outcomes achieved (likely to be achieved);</li> <li>What major factors influenced the achievement or non-achievement of the outcomes?</li> <li>Were there unintended (positive or negative) outcomes of assistance for participants and non-participants?</li> <li>Is the achievement of outcomes leading to/likely to lead to meeting intervention objectives? What major factors influenced this?</li> <li>Were results delivered for men, and women, boys and girls?</li> <li>Were relevant assistance standards met?</li> </ul>
<b>Impact</b>	<ul style="list-style-type: none"> <li>The 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).</li> <li>The 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).</li> <li>Impact as criterion should not be confused with '<a href="#">Impact Evaluation</a>'.<sup>13</sup></li> </ul>	<ul style="list-style-type: none"> <li>What were the effects of the intervention on participants' lives (intended and unintended)?</li> <li>Did a specific part of the intervention achieve greater impact than another?</li> <li>Were there any gender-specific impacts? Did the intervention influence the gender context?</li> <li>Were there impacts on institutions?</li> <li>Did the intervention contribute to long-term intended results?</li> </ul>
<b>Sustainability</b>	<ul style="list-style-type: none"> <li>Capacity building/development results.</li> <li>Institutional/systemic changes.</li> <li>Integration of intervention elements into national systems and processes.</li> <li>The financial, economic, social, environmental, and institutional capacities of</li> </ul>	<ul style="list-style-type: none"> <li>To what extent did the intervention implementation consider sustainability, such as capacity building of national and local government institutions, communities and other partners?</li> <li>To what extent did intervention benefits continue after WFP's work ceased? <i>OR</i> To what extent is it likely that the benefits of the</li> </ul>

<sup>13</sup> The [WFP Evaluation Policy \(2016-2021\)](#) defines impact evaluation as evaluations that “assess the positive and negative, direct or indirect, intended or unintended changes in the lives of affected populations in receipt of WFP interventions”. In addition, [WFP Impact Evaluation Strategy \(2019-2026\)](#) defines that impact evaluations require a credible counterfactual to establish a cause-and-effect relationship between the intervention and the changes attributed to a specific programme. For this, impact evaluations estimate what would have happened in the absence of the intervention, and attribute the difference to the programme's impact. Impact evaluations seek to answer questions such as: “*What is the impact on participants' lives which can be attributed to the programme intervention?*”, “*What were the effects of the intervention on participants' lives which would have not materialized in the absence of the intervention?*”, *What difference does the intervention make*”.

Criterion	Includes analysis of:	Potential evaluation questions
	<p>the systems needed to sustain net benefits over time.</p> <ul style="list-style-type: none"> <li>Resilience, risks and potential trade-offs.<sup>14</sup></li> </ul>	<p>intervention will continue after WFP's work ceases?</p> <ul style="list-style-type: none"> <li>Has the intervention made any difference to gender relations in the medium or longer term?</li> </ul>
<b>Appropriateness</b>	<ul style="list-style-type: none"> <li>Extent to which WFP inputs were tailored to needs.</li> <li>Extent to which they were adapted to respond to the changing demands of unstable environments.</li> <li>Extent to which design and implementation were gender-sensitive, based on gender analysis.</li> </ul>	<ul style="list-style-type: none"> <li>Was the chosen intervention approach the best way to meet the food security and nutrition needs of affected populations and intended beneficiaries?</li> <li>Were adopted transfer modalities the best way of meeting recipient's needs?</li> <li>Were protection and ethics issues considered in design and implementation?</li> <li>To what extent was the intervention based on a sound gender analysis?</li> <li>To what extent was the design and implementation of the intervention gender-sensitive i.e. considered gender equality and women empowerment issues?</li> </ul>
<b>Coverage</b>	<ul style="list-style-type: none"> <li>Extent to which different groups are targeted or included</li> <li>Impact of exclusion on sub-groups (gender, ethnicity, location, family circumstance).</li> <li>Differentiation of targeting and forms/amount of assistance provided.</li> </ul>	<ul style="list-style-type: none"> <li>Were the humanitarian needs of key target groups (men and women, boys and girls) met by the intervention?</li> <li>Was WFP's assistance provided proportionally according to the needs within the context? OR Did different geographical areas or groups of populations affected differently receive assistance according to their needs?</li> <li>Were relevant assistance standards met?</li> <li>Was WFP's assistance provided coordinated with that provided by others (duplication/gaps)?</li> </ul>
<b>Connectedness</b>	<ul style="list-style-type: none"> <li>Consistency between short-term activities and other development interventions/ goals etc. that address contextual problems.</li> <li>Presence of transition-focused analyses like stakeholder consultations, and the existence of a transition strategy.</li> </ul>	<ul style="list-style-type: none"> <li>What have been the linkages between the intervention and any other WFP interventions in relief/recovery/development?</li> <li>To what extent did the intervention link to any transition strategies in the context or to development goals?</li> </ul>

For more information, visit our [external](#) and [internal](#) webpages or contact OEV Cap/Qual Unit at: [wfp.decentralizedevaluation@wfp.org](mailto:wfp.decentralizedevaluation@wfp.org)

<sup>14</sup> Depending on the timing of the evaluation, this may involve an analysis of the actual flow of net benefits or an estimation of the likelihood of net benefits continuing over the medium and long-term.