

Technical Note

Evaluation Questions and Criteria

Version August 2017

1. Introduction

1. This Technical Note is intended to guide Evaluation Managers to identify well defined and appropriate evaluation questions that are based on international evaluation criteria, and that will result in a high quality and credible evaluations.
2. Developing appropriate and well-defined **evaluation questions** is fundamental to any evaluation, as they frame the evaluation and support the selection of appropriate evaluation approaches and methods, indicators and data collection tools. The evaluation questions should reflect **international evaluation criteria**¹, and be drawn from the programme Theory of Change, considering the purpose of the evaluation, and how it will be used.
3. All WFP evaluations assess WFP interventions' quality in relation to multiple international evaluation criteria. The international evaluation criteria are: **relevance, effectiveness, efficiency, impact and sustainability**. Evaluations of humanitarian and emergency programming, or interventions in a humanitarian setting, should also apply the criteria of **appropriateness, connectedness, coherence, and coverage**. Evaluation questions are normally developed with reference to these criteria to ensure quality, credibility and relevance of WFP evaluations and their accessibility by stakeholders. The criteria help structure the evaluation questions, and help the evaluation manager to ensure the appropriate aspects of the intervention are included in the evaluation.

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

4. International evaluation criteria set the *values* and the *framework* that will guide an evaluation. They provide a frame that helps to select the main evaluation questions, undertake analysis and organise conclusions drawing from findings presented. Ideally an evaluation (including impact evaluations) will address all 5 main criteria but may give different weighting to each criterion and use different methods when incorporating them to the evaluation. Depending on the subject and the evaluation questions of interest to the commissioner, a specific set of criteria might be more or less dominant in a particular evaluation. The TOR (or IR) should justify why specific criteria are chosen, and how they will be applied in the different phases of the evaluation. Table 1 (page 2) sets out the definitions for each criteria².
5. Evaluation questions need to be relevant to the intervention and to the needs of the users. The international criteria provide a value-based framework in which to situate the evaluation questions. An evaluation can also ask additional important evaluation questions that are outside the framework (for example, on equity and human-rights).

¹ OECD-DAC Criteria for Evaluating Development Assistance (<http://www.oecd.org/dac/evaluation/49756382.pdf>)

² Adapted from OECD-DAC, 2000, DAC Criteria for Evaluating Development Assistance: Standard definitions for *Relevance, Effectiveness, Efficiency, Impact, Sustainability* and ALNAP.

Table 1: Definitions for international evaluation criteria

Criteria	Definition
Relevance	<i>Extent to which the objectives of an intervention are consistent with the most vulnerable group's needs, country needs, organisational priorities and partners' policies and practice</i>
Effectiveness	<i>The extent to which the intervention's objectives as defined are achieved, and the extent to which outputs have led (or are expected to lead) to expected outcomes as planned.</i>
Efficiency	<i>Measures the outputs – qualitative and quantitative – in relation to inputs – funds, expertise, time etc. This generally requires comparing alternative approaches to achieving the same outputs</i>
Impact	<i>Wider effects of the project- social, economic, technical, environmental – on individuals, gender- and age groups, communities and institutions. Impacts can be intended or unintended, positive or negative, macro (sector) or micro (household).</i>
Sustainability	<i>The continuation of benefits from an intervention after assistance has been completed, or the probability of long term benefits.</i>
Further criteria for interventions in Humanitarian settings:	
Appropriateness	<i>The tailoring of activities to local needs and context, thereby increasing ownership, accountability and cost-effectiveness. This replaces the OECD-DAC criteria of Relevance.</i>
Coverage	<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>
Coherece	<i>The relationship between the subject of the evaluation, and the political, security, developmental, trade and military context as well as humanitarian policies. In particular, the degree to which these take into account humanitarian and human-rights considerations, principles and standards.</i>
Connectedness	<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/replacing sustainability above.</i>

3. Setting appropriate evaluation questions for WFP evaluations

6. Identifying and agreeing appropriate evaluation questions is one of the key tasks for an evaluation manager during the evaluation design and preparation. The evaluation will be structured around the evaluation questions, which will dictate the appropriate methods and approaches, indicators, data collection and analysis. The evaluation manager needs to consider which evaluation criteria and questions are the most appropriate and useful for the evaluation. She/he should set out the 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. Table 3 (Annex 1) shows how the international evaluation criteria can support the development of evaluation questions.
7. Good evaluation questions should be:
 - **well-defined** and chosen depending on the scope and objectives of an evaluation, and how the evaluation findings will be used. For example, an evaluation primarily seeking to inform replication of the intervention in a different context will emphasize very different questions from an evaluation that is seeking to demonstrate whether the intervention covered all vulnerable groups, or whether an intervention was cost-effective.
 - drawn from the **Theory of Change** for the intervention, to ensure specificity to the context and the intervention. In the absence of a Theory of Change, the evaluation manager may consider investing some time with the programme colleagues to reconstruct the ToC.
 - go beyond assessing if intended results were achieved (e.g. how many people were reached with what quantity of food), by seeking to **explain the reasons for the results** or for the lack of them. These types of questions promote lesson-learning.

- **prioritised** given evaluations are often limited in terms of budget, time, and resources. It is better to have few, strategically designed and well-defined evaluation questions, instead of a large number of questions that cover everything and that may contradict each other.
 - developed applying a **gender** lens by considering **gender equality and women empowerment** within each of the criteria and in the evaluation questions, to ensure that the evaluation assesses the inclusion of gender dimensions in the intervention design and implementation.³
8. Collectively, the evaluation questions should be designed to give evaluation users the information they need, to make strategic or operational decisions, take action, or understand and learn from the intervention. It is also useful to discuss the evaluation questions with stakeholders to ensure shared understanding on what the evaluation can and will answer. This would be typically done by seeking inputs from the Evaluation Reference Group to the draft TORs and subsequently the draft Inception Report.
 9. The evaluation manager and the evaluation reference group should ensure throughout the evaluation that the evaluation team is focused on the evaluation criteria and answering the evaluation questions, and that the analytical approach, methods, data collection and any triangulation are systematically structured to flow from the evaluation questions. The evaluation **findings should be reported against the evaluation questions and criteria**. Stemming logically from the findings, the conclusions can also be organized around the evaluation criteria used in the evaluation.

Box 1: Further (including DEQAS) reading – Evaluation Criteria and Questions

- ALNAP, Cosgrave J. and Buchanan-Smith M., 2016, [Evaluating Humanitarian Action Guide](#)
- IASC, 2010, [International Humanitarian Norms & Principles - Guidance Materials](#)
- OECD, 2007, [Principles for Good International Engagement in Fragile States and Situations](#)
- OECD-DAC, 2010, [Evaluating Development Co-Operation - Summary of Key Norms and Standards](#)
- OECD-DAC, 2000, [DAC Criteria for Evaluating Development Assistance](#)
- OECD-DAC, 1999, [Guidance for Evaluating Humanitarian Assistance in Complex Emergencies](#)
- OECD-DAC, 1991, [Principles for Evaluation of Development Assistance](#)
- UNEG, 2014, [UN-SWAP Evaluation Performance Indicator Technical Note](#)
- UNEG, 2011, [Integrating Human Rights and Gender Equality in Evaluation](#)
- WFP, [Online Evaluation Glossary](#)

³ 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](#).

4. Further reading

Annex 1: Table 3 – Using the International Evaluation Criteria to develop evaluation questions

Criterion ⁴	Includes analysis of:	Potential evaluation questions
Relevance	<ul style="list-style-type: none"> • Relevance of the intervention design in view of the needs of the most vulnerable groups. • Continued relevance of the objectives over the life of the intervention (ability to adapt to new needs). • Alignment with government, partners, donors' policies and interventions; or alignment and coherence with WFP policies. • Consistency of project design and logic. • Extent to which design and implementation were gender-sensitive, based on gender analysis, and addressed diverse needs. 	<p>To what extent:</p> <ul style="list-style-type: none"> • was the design of the intervention relevant to the wider context? • is the intervention in line with the needs of the most vulnerable groups (men and women, boys and girls)? • is the intervention aligned with the needs of the government? • is the intervention aligned with WFP, partners, UN agencies and donor policies and priorities? • was the intervention based on a sound gender analysis? • was the design and implementation of the intervention gender-sensitive?
Effectiveness	<ul style="list-style-type: none"> • Achievement of objectives (or likelihood that the objectives be achieved). • Main results including positive, negative, intended and unintended outcomes. • Outputs and outcomes for men, women, boys and girls, and other relevant socio-economic categories. • Potential constraints and facilitating factors to achievements. 	<p>To what extent:</p> <ul style="list-style-type: none"> • were (are) the outputs and outcomes of the intervention achieved (likely to be) achieved; and what were the major factors influencing the achievement or non-achievement of the outcomes? • is the achievement of outcomes leading to/likely to lead to achievement of objectives of the intervention; what were the major factors influencing this? • did the intervention deliver results for men and women, boys and girls? • were the relevant assistance standards met?
Efficiency	<ul style="list-style-type: none"> • Costs per recipient for different implementation mechanisms/mode of transfer -food/cash/ voucher. • timeliness of delivery, compliance with intended timeframes, comparison of channels of delivery (e.g. schools/ health systems versus community-based). • comparison of different institutional arrangements (e.g. continuity of supplies and use of local partners / systems / procurement where feasible). 	<ul style="list-style-type: none"> • Was the intervention cost-efficient? • Was the intervention implemented in a timely way? • Was the intervention implemented in the most efficient way compared to alternatives? • Did the targeting of the intervention mean that resources were allocated efficiently?
Impact <i>As defined as a Criterion. Impact has a nuanced meaning when used in the context of 'Impact Evaluation'.</i>	<ul style="list-style-type: none"> • Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended. • Wider effects of the project- social, economic, technical, environmental – on individuals, gender- and age groups, communities and institutions. 	<ul style="list-style-type: none"> • What were the effects of the intervention on recipients' lives? • Did a specific part of the intervention achieve greater impact than another? • Were there unintended (positive or negative) effects for recipients and non-recipients of assistance? • What were the gender-specific impacts? Did the intervention influence the gender context?

⁴ Definitions from WFP's Operations Evaluation Guidance: Technical Note on Evaluation Criteria

Criterion ⁴	Includes analysis of:	Potential evaluation questions
<i>Refer to the TN on impact evaluation for further guidance.</i>	<ul style="list-style-type: none"> Impact evaluations seek to establish a causal relationship between the intervention and resulting changes, recognising that any intervention only contributes to any long-term observed impacts. 	<ul style="list-style-type: none"> Impacts on institutions? Contribution of an intervention to long-term intended results.
Sustainability	<ul style="list-style-type: none"> Capacity building/development results. Institutional/systemic changes. Integration of intervention elements into national systems and processes. 	<ul style="list-style-type: none"> To what extent did the intervention implementation arrangements include considerations for sustainability, such as capacity building of national and local government institutions, communities and other partners? To what extent did the benefits of the intervention continue after WFP's work ceased? OR To what extent is it likely that the benefits of the intervention will continue after WFP's work ceases? Has the intervention made any difference to gender relations in the medium or longer term?
Appropriateness	<ul style="list-style-type: none"> Extent to which WFP inputs were tailored to needs. Extent to which they were adapted to respond to the changing demands of unstable environments. Extent to which design and implementation were gender-sensitive, based on gender analysis. 	<ul style="list-style-type: none"> Was the chosen intervention approach the best way to meet the food security and nutrition needs of affected populations and intended beneficiaries? Were adopted transfer modalities the best way of meeting recipient's needs? Were protection issues considered in the design and implementation? To what extent was the intervention based on a sound gender analysis? To what extent was the design and implementation of the intervention gender-sensitive i.e. considered gender equality and women empowerment issues?
Coverage	<ul style="list-style-type: none"> Extent to which different groups are targeted or included Impact of exclusion on sub-groups (gender, ethnicity, location, family circumstance). Differentiation of targeting and forms/amount of assistance provided. 	<ul style="list-style-type: none"> Were the humanitarian needs of key target groups (men and women, boys and girls) met by the intervention? Was WFP's assistance provided proportionally according to the needs within the context? OR different geographical areas or groups of populations affected differently receive assistance according to their needs? Were relevant assistance standards met? Was WFP's assistance provided consistent with that provided by others (duplication/gaps)?
Coherence	<ul style="list-style-type: none"> Contextual factors and how they influenced the design/ implementation of the subject. Links to the food security and nutrition policies and programmes of other actors. Consideration of humanitarian and human rights principles and standards, including gender equality and women empowerment. 	<ul style="list-style-type: none"> To what extent were context factors (political stability/instability, population movements etc) considered in the design and delivery of the intervention? To what extent was WFP's intervention coherent with key policies and programmes of other partners operating within the same context? To what extent was the intervention design and delivery overall in line with humanitarian principles?
Connectedness	<ul style="list-style-type: none"> Consistency between short-term activities and other development interventions/ goals etc. that address contextual problems. Presence of transition-focused analyses, including stakeholder consultations and existence of transition strategy. 	<ul style="list-style-type: none"> What have been the linkages between the intervention and any other WFP interventions in relief/recovery/development? To what extent did the intervention link to any transition strategies in the context or development goals?