

Evaluation title	General Food Assistance and School Feeding Programmes, Libya 2017-2019	
Evaluation category and type	Decentralized evaluation	
Post Hoc Quality Assessment (PHQA) – overall rating	Satisfactory: 71%	
<p>The evaluation report is generally well written and well documented, providing credible findings that evaluation users can rely on with confidence for decision-making. The methodological approach involving mixed methods of data collection and the sampling, data gathering tools and sources produced reliable findings. The findings are very comprehensive and summaries at the end of each evaluation question help to identify the key issues. The report also differentiates between findings that are interesting but not fully verifiable and those that were fully triangulated. Conclusions and recommendations follow logically from the findings, although explicit linkages would have made the logic flow clearer. Lessons learned are useful, but many are written as conclusions, failing to demonstrate their more general relevance. Several required annexes are missing, and additional ones, such as the programme logic model and/or theory of change, would have provided a useful context for some of the findings. Gender and inclusion are well incorporated in the methodology, findings and recommendations.</p>		
CRITERION 1: REPORT SUMMARY	Rating	Satisfactory
<p>The summary includes the main evaluation features, context and overview, as well as findings and recommendations. However, it exceeds the requirements on length, making it difficult to find the key messages. The summary would have benefited from a more selective choosing of what to include and could have made the findings-conclusions-recommendations link more explicit.</p>		
CRITERION 2: CONTEXT AND OVERVIEW OF EVALUATION SUBJECT	Rating	Partly Satisfactory
<p>The country context is relatively well done, providing a useful overview of the humanitarian crisis with key information for understanding issues faced by WFP. However, national policies and the normative framework for gender and human rights are missing and there is only a brief reference to the treatment of minority/ethnic groups and refugees. For example, although CEDAW is in the list of acronyms, commitments under CEDAW are (not mentioned in the text. The report notes that WFP received a score of 4 (the highest) on GAM and has recently completed a gender analysis, but there is no preliminary information presented from the latter, and no explanation of what elements in the programme resulted in such a high GAM score. As such, the information contained is not sufficient to understand the specific context of women and other vulnerable groups in the programme areas. There is no reference to the SDGs or the National Voluntary Report, and little information on any previous research or evaluation. There is a good overview of the evaluation subject, particularly partners, transfer modalities, implementation period and geographic coverage. However, the capacity building activities are not mentioned, even though they are important in the findings. Additionally, the overview of the intervention contains little information about objectives, theory of change, logical framework or crosscutting issues and the annexes do not include the logical framework analysis or Theory of Change (even though the text notes that there is one) that would have more clearly outlined the logic of intervention and expected achievements of the programmes, and only global budget figures are provided (Annex 6). Although the text provides numbers reached versus planned, only the number of children in the SFP was disaggregated by sex.</p>		
CRITERION 3: EVALUATION RATIONALE, OBJECTIVES, AND SCOPE	Rating	Satisfactory
<p>The rationale, objectives and scope are generally complete, providing a good overview of what the evaluation sought to address. The introduction notes that this is the first evaluation since WFP started working in Libya in 2014 (from Tunisia), but does not explicitly state why it is being carried out at this specific point in time. In the case of the SFP, which is a pilot, the reason for the timing of the evaluation is clearer. Moreover, the rationale is clear, as are the geographic areas and time period covered by the evaluation. The main users/stakeholders are also briefly described.</p>		

CRITERION 4: METHODOLOGY	Rating	Satisfactory
<p>The evaluation methodology is clearly described, taking into account the limitations of a conflict zone and COVID-19, as well as limitations on available data, qualitative indicators, and cultural barriers to reaching women beneficiaries. The evaluation used mixed methods of data collection consisting of document review, key informant interviews (KIIs), including beneficiaries, site visits and a case study addressing the issue of conflict sensitivity (CS). The use of multiple approaches was particularly appropriate for triangulation, given the lack of qualitative data, and the heavy impact of conflict in the context. The evaluation matrix is clear and complete with all the required columns and information. Gender is explicitly addressed in three evaluation questions. OECD/DAC and humanitarian criteria were used effectively, especially given the UN move towards the New Way of Working, essentially linking development, humanitarian and peace actions.</p>		
CRITERION 5: FINDINGS	Rating	Satisfactory
<p>The findings are very comprehensive, with all questions satisfactorily and transparently addressed, making for a very interesting narrative that will surely be of great use to the WFP CO and perhaps stimulate new questions and issues to address. Use of evidence is excellent, and findings appear to be impartial and balanced. Of particular use are the summary boxes at the end of the findings for each evaluation question, which help to emphasize the findings that are most important. However, the findings, while comprehensive, are much too long, making it difficult to differentiate between minor observations and significant findings. In addition, there seems to be considerable repetition/overlap between findings related to some evaluation questions, and much data that the evaluation team notes could not be verified. Moreover, no unexpected results are identified.</p>		
CRITERION 6: CONCLUSIONS AND LESSONS	Rating	Partly Satisfactory
<p>Conclusions and lessons learned appear to be logical and draw together findings across criteria. While the conclusions can be useful for strategic decision-making, they may not be specific enough for accountability. The conclusions are very positive until the final one which provides a list of shortcomings. It is not clear how much these shortcomings impinge on the positive conclusions listed in the other 9 of 10 conclusions provided. Under conclusion 10, the first, third, fifth, sixth and ninth points seem to undermine this conclusion. Conclusions do not make any explicit links to the findings, and it is therefore not easy to see the flow from the findings. Some form of specific linkage between the conclusions and the findings would have been useful and would have ensured that the linkages were comprehensive. While lessons learned are useful, a number of them are written more as conclusions without clarifying the potential for their wider application. In addition, it is unusual for the conclusions in the text to be slightly shorter than the ones in the Executive Summary.</p>		
CRITERION 7: RECOMMENDATIONS	Rating	Satisfactory
<p>There are 13 recommendations, exceeding the limit of 10. The recommendations, with a few exceptions, are logically derived from the findings and conclusions, and are mostly realistic/ feasible, take into account the context, are actionable and directed. Some, especially operational ones, are not necessarily very actionable: R 11 and 12 clearly depend on the cooperation of or are the responsibility of other actors who are not defined. The recommendations provide useful suggestions for future programming. Although they are derived logically (intuitively) from the findings, the report could have been strengthened with clearer linkages drawn between findings/ conclusions/ recommendations explicit through a table, or cross-referencing relevant paragraphs. Also, the recommendations should have been prioritized and provided with timelines for their implementation, with some more specifically targeted to responsible actors.</p>		
CRITERION 8: ACCESSIBILITY AND CLARITY	Rating	Satisfactory
<p>The report is well written and very well documented, but its accessibility would have been improved by moving many of the details to the annexes to significantly reduce the length of the report which largely exceeds WFP requirements. Several required annexes are missing, along with others that are not required but would have been useful.</p>		

Integration of Gender Equality and Women's Empowerment (GEWE) considerations in the evaluation report based on the UN System-Wide Action Plan (UN-SWAP) Evaluation Performance Indicator (EPI) scorecard

UN-SWAP EPI – individual evaluation score Meets Requirements

While gender is not included as a specific evaluation objective, gender is well integrated into the evaluation scope, criteria and evaluation questions including in some evaluation questions. The report notes most available data is sex-disaggregated. However, WFP does not monitor qualitative gender indicators and the evaluation team did not generate questions and indicators to approximate the aspect of empowerment. Despite this limitation, as well as those imposed by political conflict and COVID-19, the methodology was effective in ensuring women's voices were reflected, including women interviewers who could more appropriately approach women, and men's cooperation was sought to avoid possible conflict within the family. The findings are strong in transparency, data disaggregation and explicit triangulation of the voices of different social groups, particularly women. The recommendations include specific measures for GEWE, including identifying and monitoring the qualitative data that has been identified as a gap. While the country context includes a paragraph on gender and refers to women and vulnerable groups in other paragraphs, it does not adequately explain the context for gender in Libya or the impact of ethnicity, conflict, and displacement on women and girls in the project zones.

Post Hoc Quality Assessment – Rating scale and definitions at overall report and criteria levels	
Highly Satisfactory	<i>Definition at overall report level:</i> Evaluation users can rely on the credible and useful evaluation findings provided and can use the evaluation with a high degree of confidence for decision-making. The report is considered an excellent example.
	<i>Definition at criterion level:</i> The criterion is addressed without any gaps or limitations.
Satisfactory	<i>Definition at overall report level:</i> Evaluation users can rely on the quality and credible evaluation findings provided and can use it with confidence for decision-making.
	<i>Definition at criterion level:</i> There are no significant gaps or limitations in addressing the criterion.
Partly Satisfactory	<i>Definition at overall report level:</i> Evaluation users can rely on the evaluation findings provided and may use it for decision-making noting that there are some gaps/shortcomings in the information provided.
	<i>Definition at criterion level:</i> There are some significant gaps or limitations in addressing the criterion.
Unsatisfactory	<i>Definition at overall report level:</i> Evaluation users can use some of the learning from the evaluation, noting that there are significant gaps/ shortcomings in the evaluation findings provided. The report may still contribute to decision making but should be used with caution.
	<i>Definition at criterion level:</i> There are critical gaps or limitations in addressing the criterion. Most of the required parameters are not met.