

<b>Evaluation title</b>	End-Term Evaluation of WFP School-Feeding USDA McGovern Dole Grant 2017-2020 in Bangladesh	
<b>Evaluation category and type</b>	Decentralized – Activity	
<b>Post Hoc Quality Assessment (PHQA) – overall rating</b>	Partly Satisfactory: 51%	
<p>The report provides a wealth of data providing evidence of different project activities and achievements. The design that included systematic baseline, midline and endline data collection and a control group allowed making some well-substantiated claims of attributing selected observed changes to project interventions. The evaluation effectively mainstreamed GEWE-perspectives into data collection, analysis, and reporting. However, the report shows considerable weaknesses in terms of critically examining the evaluated project's (implicit) theory of change and presenting a plausible contribution story for more complex envisaged outcomes and overall impact. While the report is written in clear language, findings are negatively affected by the absence of clearly formulated key messages, and an overly strong focus on activity- and output-level details that distract the reader from the bigger 'so what?' questions. The conclusions section summarizes findings but does not formulate strategic forward-looking considerations. Several of the recommendations do not logically flow from the presented findings, are not targeted, and lack specificity. Readability would have benefited from using more visual aids such as graphs and using bold font more sparingly.</p>		
<b>CRITERION 1: REPORT SUMMARY</b>	<b>Rating</b>	<b>Unsatisfactory</b>
<p>The report summary provides an overview of the evaluation subject, context, and describes key evaluation features. It presents several evaluation findings and reflects the evaluation recommendations, which however do not logically flow from the findings reported. However, the summary is overly detail-oriented, which results in it being considerably longer than suggested. The presentation of findings suffers from the absence of clearly identified key messages for each of the evaluation questions. The summary does not present evaluation conclusions.</p>		
<b>CRITERION 2: CONTEXT AND OVERVIEW OF EVALUATION SUBJECT</b>	<b>Rating</b>	<b>Partly Satisfactory</b>
<p>The report presents relevant dimensions that constitute the backdrop for the programme under review and provides some basic information on the evaluation subject. The context section includes reflections on how different population groups (women/girls, refugee populations, poorer families) are impacted. The section is overly long, however, and not all provided details are clearly relevant to the evaluation. The evaluation subject is only partly described, and it remains unclear what role WFP played in terms of programme implementation beyond the distribution of micronutrient fortified biscuits. There is no information on other WFP work, or on changes in the work of WFP and other actors. The report would have benefited from describing implications of the Rohingya refugee crisis in more detail.</p>		
<b>CRITERION 3: EVALUATION RATIONALE, OBJECTIVES, AND SCOPE</b>	<b>Rating</b>	<b>Satisfactory</b>
<p>The report clearly outlines the evaluation learning and accountability objectives, rationale and includes explicit reflections on why and how gender and (to a lesser extent) human rights considerations were mainstreamed in the evaluation. It defines the time period, geographic areas, and target groups covered by the evaluation. However, the report does not explicitly identify its intended users and would have benefited from providing more explicit information on the evaluation scope in terms of activities explored beyond the distribution of fortified biscuits.</p>		
<b>CRITERION 4: METHODOLOGY</b>	<b>Rating</b>	<b>Partly Satisfactory</b>
<p>The methodology was suited for collecting relevant information on most of the evaluation questions. Quantitative data were complemented by qualitative information derived from consultations with diverse stakeholder groups. Gender equality was mainstreamed, and the use of mixed method data collection approaches was appropriate for addressing GEWE considerations. The report includes a detailed evaluation matrix, clearly notes limitations and mitigation strategies, and comments on ethical standards considered. The methodology was however less suited for plausibly tracing the initiative's "contribution story" in relation to observed competency and behaviour changes among students or</p>		

communities. Weaknesses are due to the absence of a clear and critically examined theory of change; an oversimplified understanding of the change processes required for improvements in learning outcomes and behaviours; and solely relying on self-reported changes to identify positive changes.

<b>CRITERION 5: FINDINGS</b>	<b>Rating</b>	<b>Partly Satisfactory</b>
------------------------------	---------------	----------------------------

The evaluation addresses most of the posed evaluation questions. Evidence is generally presented transparently and clearly, providing sources for most presented data, and largely using a neutral tone. The report discusses WFP contributions to results, taking effects of the COVID-19 pandemic into account, and reflects the voices of diverse stakeholder groups. Findings on effectiveness, efficiency and impact are structured by indicator, which leads to a focus on activities and outputs, while neglecting 'bigger picture' synthesis and analysis. In some cases, the report implies an (unintended) positive bias by attributing changes to project interventions without providing compelling evidence. The report does not systematically reflect on likely contributions of other actors, such as to changes in student reading and writing abilities.

<b>CRITERION 6: CONCLUSIONS AND LESSONS</b>	<b>Rating</b>	<b>Partly Satisfactory</b>
---	---------------	----------------------------

The conclusions summarize the evaluation findings across all evaluation criteria and make reference to a couple of GEWE-related issues. The conclusions do not synthesize the findings at a higher, strategic level, such as by noting strategic implications of the evaluation findings for the future and/or for similar interventions. The conclusions do not reflect on broader inclusion and equity issues and the report does not explicitly identify key lessons learned despite noting the intention to do so.

<b>CRITERION 7: RECOMMENDATIONS</b>	<b>Rating</b>	<b>Partly Satisfactory</b>
-------------------------------------	---------------	----------------------------

The report makes a limited number of recommendations that are categorized as either 'strategic' or 'operational'. Two of the recommendations explicitly address GEWE or broader equity and inclusion dimensions. Most of the recommendations do not clearly flow from the evaluation findings and conclusions. The recommendations vary in the extent to which they are realistic and feasible, and several of them do not clearly take the implementation context and WFP's comparative advantage into account. The recommendations are neither prioritized nor addressed to specific lead actors responsible for their implementation.

<b>CRITERION 8: ACCESSIBILITY AND CLARITY</b>	<b>Rating</b>	<b>Satisfactory</b>
---	---------------	---------------------

The report uses clear language without jargon, with most of the acronyms being spelled out the first time they are used. It clearly identifies sources for all data and quotes, and includes cross-references to direct readers to information that can be found in other parts of the report. Readability and clarity of the report are negatively affected by the absence of clearly identified higher-level findings for each (sub)section, a very detailed and micro-level-oriented narrative that makes it difficult to identify and prioritize key messages, and the comparatively scarce use of visual aids such as tables or graphs. While it includes most prescribed annexes, some annexes are missing, and not all are referenced in the main report.

**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**

<b>UN-SWAP EPI – individual evaluation score</b>	<b>Meets requirements: 9 points</b>
--	-------------------------------------

The evaluation approach and mixed-methods methodology, which drew upon a variety of data sources and processes, were gender-responsive and based on deliberate considerations on how to integrate GEWE dimensions in data collection and analysis. While there is no stand-alone question or sub-question on GEWE, gender and human-rights related dimensions are addressed through explicit indicators for some questions. The report provides explicit reflections on the availability of data on GEWE collected by WFP. Ethical standards were consistently considered, and all stakeholder groups treated with respect for confidentiality and integrity. Evaluation findings and conclusions, and recommendations reflect a gender analysis that, at least in some places, goes beyond questions of numerical equality towards exploring the (potential) relevant effects of the observed changes for broader equity and equality issues.

Post Hoc Quality Assessment – Rating scale and definitions at overall report and criteria levels	
<b>Highly Satisfactory</b>	<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.
<b>Satisfactory</b>	<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.
<b>Partly Satisfactory</b>	<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.
<b>Unsatisfactory</b>	<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.