

Evaluation title	Mid Term Evaluation of the McGovern-Dole International Food for Education and Child Nutrition Program 2018-2021 in Nepal	
Evaluation category and type	Decentralized – Midterm evaluation	
Post Hoc Quality Assessment (PHQA) – overall rating	Partly Satisfactory: 49%	
<p>The report provides well-substantiated findings in relation to the relevance and coherence of the reviewed programme. However, it shows considerable weaknesses when discussing programme effectiveness, efficiency, impact and sustainability due to a lack of critically examining the evaluated programme's theory of change, and clearly distinguishing between correlation and causation when discussing programme contributions to changes. The evaluation's focus on emphasizing programme strengths and equating planned with actual contributions to results create an impression of (unintended) positive bias. The report effectively mainstreamed GEWE-perspectives into data collection, analysis and reporting. While the report is generally written in clear language, its readability is negatively affected by the absence of clearly formulated key messages or findings statements. The conclusions constitute a summary of findings but do not formulate strategic forward-looking considerations. The recommendations do not clearly flow from the findings and conclusions and vary in the extent to which they are specific and realistic.</p>		
CRITERION 1: REPORT SUMMARY	Rating	Satisfactory
<p>The summary presents key evaluation features and relevant context information, as well as information on evaluation findings, conclusions and recommendations. While the presentation of findings suffers from the absence of clearly identified key messages for each of the main evaluation questions, compared to the main report, the conclusions are stronger in that they are formulated at a higher-level than those in the main report. The recommendations do not flow logically from the findings and conclusions, both of which strongly focus on strengths of the reviewed programme.</p>		
CRITERION 2: CONTEXT AND OVERVIEW OF EVALUATION SUBJECT	Rating	Partly Satisfactory
<p>The report describes the evaluation context and the subject of the evaluation. It reflects on gender equality issues of different contextual dimensions but does not discuss broader equity and human rights considerations. The evaluation describes the main features of the programme under review and positions the programme within the broader context of WFP programming in Nepal. The evaluation does not, however, present or comment on the programme's (explicit or explicit) theory of change and related underlying assumptions.</p>		
CRITERION 3: EVALUATION RATIONALE, OBJECTIVES, AND SCOPE	Rating	Satisfactory
<p>The report clearly describes the evaluation objectives, purpose, and scope in terms of the time period and geographic area covered. It identifies the intended users and uses of the evaluation (albeit in the context section rather than as part of describing key evaluation features), and notes that gender equality and women's rights, as well as accountability to affected populations, were mainstreamed in the evaluation.</p>		
CRITERION 4: METHODOLOGY	Rating	Partly Satisfactory
<p>The design chosen allowed the evaluation to do some comparisons between a treatment and a control group of schools to support findings. Gender equality was mainstreamed, and the evaluation consistently considered ethical standards. The evaluation methodology was only partially suited for tracing and presenting the programme's 'contribution story' especially in relation to higher-level objectives related to student attendance and performance, and to social norms change around gender equality. Key methodological weaknesses include the absence of a clearly articulated and critically examined theory of change; an oversimplified understanding of the time and complex change processes required to affect changes in areas such as learning outcomes and gender norms; and uncritically relying on stakeholder perceptions for identifying and quantifying progress made towards results. Application of the originally envisaged methodology was negatively affected by the COVID-19 pandemic, which limited the evaluation's ability to extensively use comparisons with available baseline data.</p>		

CRITERION 5: FINDINGS	Rating	Partly Satisfactory
<p>The evaluation addresses all the posed evaluation questions and most sub-questions. It provides strong evidence in relation to programme relevance and coherence, but is considerably weaker when discussing programme effectiveness, efficiency, impact, and sustainability. It frequently constitutes that positive changes have occurred and attributes these changes to programme interventions yet without providing compelling evidence beyond the perceptions of selected stakeholder groups. The evaluation places stronger emphasis on programme strengths than on weaknesses, which creates an overall impression of positive bias. Sources are provided for all quotes but are missing for most of the presented figures. The discussion is, in part, characterized by an overly strong focus on detail rather than on formulating bigger picture insights on the main evaluation questions. The report does not explicitly discuss unanticipated effects, including on human rights and gender equality.</p>		
CRITERION 6: CONCLUSIONS AND LESSONS	Rating	Partly Satisfactory
<p>The conclusions summarize the evaluation findings across all evaluation criteria and make reference to GEWE-related issues. The conclusions largely reformulate the findings, however, rather than synthesizing them at a higher, strategic level. While they include some forward-looking elements, these do not clearly derive from the evidence presented in the findings section. The conclusions do not reflect on broader inclusion and equity issues. The report does not include lessons but presents two examples of 'good practices', the credibility of which suffers from the weak evidence base of the findings from which they derive.</p>		
CRITERION 7: RECOMMENDATIONS	Rating	Partly Satisfactory
<p>The evaluation puts forward five recommendations, one of which explicitly addresses gender equality issues. All recommendations are appropriately grouped as either strategic or operational. However, they do not clearly flow from the evaluation findings and conclusions. They vary in the extent to which they are realistic and feasible, and several of them do not clearly take the implementation context into account. The recommendations are not addressed to specific lead actors responsible for their implementation and do not outline a clear timeframe for their implementation (beyond the broad categories of 'mid-term' or 'long-term'). All five recommendations are categorized as 'high' priority, which does not allow for prioritizing among them.</p>		
CRITERION 8: ACCESSIBILITY AND CLARITY	Rating	Unsatisfactory
<p>The report is written in clear language and uses visual aids including graphs, tables, and textboxes. It provides sources for all quotes and is only minimally longer than the recommended 45 pages. However, the report does not summarize or visually highlight key messages, which makes it difficult for readers to identify the main findings. Section 1 of the report slightly deviates from the structure outlined in the WFP template, and the actual content presented is not always aligned with the sub-heading under which it is positioned. The language used by the evaluation is not always precise and includes sometimes unnecessary jargon. The reviewed version of the report includes a few grammatical and spelling errors. For most figures, the evaluation does not provide sources or comprehensive axes labelling. In some cases, the report makes excessive use of direct quotes from consulted stakeholders. The evaluation includes some, but not all, of the required annexes, and the annexes considerably exceed the recommended maximum 40,000 words.</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: 7 points
<p>GEWE considerations are very well integrated into the report. The chosen mixed-method approach was based on deliberate considerations on how to effectively integrate GEWE, reflected in the evaluation matrix which includes dedicated sub-questions on gender. The evaluation drew upon a variety of data sources and processes, thereby facilitating inclusion, accuracy and credibility. Moreover, the findings include reflections on GEWE dimensions and one of the five recommendations addresses GEWE issues. However, the report would have benefited from commenting explicitly on the availability of monitoring data on GEWE-related indicators.</p>	

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.