

Evaluation title	Evaluation of Ethiopia WFP Country Strategic Plan 2020-2025	
Evaluation category and type	Centralized - CSPE	
Post Hoc Quality Assessment (PHQA) – overall rating	Satisfactory: 85%	
The Evaluation of Ethiopia WFP Country Strategic Plan 2020-2025 is a satisfactory report that stakeholders can use with confidence for decision-making. It clearly describes the evaluation purpose, objectives, questions, context, and methodology. Drawing on relevant primary and secondary data sources, the report articulates clear findings for all evaluation questions and sub-questions. It presents conclusions and recommendations that, for the most part, logically derive from the findings. The report makes good use of figures and tables and highlights key insights. The report could have been further improved by sharpening the focus of the topline findings to emphasize key analytical insights, while reducing the amount of descriptive detail. Additionally, the conclusions would have benefited from a more strategic orientation to set the stage for more specific and actionable recommendations. Finally, it should have more consistently used a neutral, non-judgmental language.		
CRITERION 1: REPORT SUMMARY	Rating	Highly Satisfactory
The report summary offers a clear, accurate, and concise overview of the evaluation. It effectively functions as a standalone document that communicates key issues, findings, and recommendations to readers unfamiliar with the full report. It includes an introduction summarizing evaluation features and main conclusions with supporting findings. There are minor inconsistencies in the titles and references of the figures included in the section that should have been addressed.		
CRITERION 2: CONTEXT AND OVERVIEW OF EVALUATION SUBJECT	Rating	Highly Satisfactory
The report provides a concise but detailed description of the country context that affected WFP programming design and implementation. It clearly describes the evaluation subject's strategic focus, internal logic, evolution over time and key operational aspects. The overview of the evaluation subject could have benefited from providing more information on how the CSP design and implementation addressed gender and broader inclusion considerations.		
CRITERION 3: EVALUATION RATIONALE, OBJECTIVES, AND SCOPE	Rating	Satisfactory
The report clearly presents the evaluation purpose, objectives, and scope, and provides information on the intended users and stakeholders. Human rights and gender equality considerations were mainstreamed in the evaluation purpose and objectives. It might have benefited from explicitly referencing gender equality and/or human rights in the primary purpose statement or the main evaluation questions, thereby marking them as key evaluation objectives and ensuring their relevance to the entire evaluation process.		
CRITERION 4: METHODOLOGY	Rating	Highly Satisfactory
The evaluation's mixed-methods and theory-based design and the methods of data collection and analysis employed were appropriate for answering the evaluation questions thoroughly, comprehensively, and in an unbiased way. The evaluation was informed by an evaluability assessment and deliberately integrated gender and broader inclusion considerations into data collection and analysis. Ethical standards were considered throughout the evaluation. The methodology annex could have benefited from summarizing key findings from the evaluability assessment, especially regarding the availability and quality of data on gender equality and inclusion issues. The report could also have mentioned how the evaluation ensured participation of people with disabilities.		
CRITERION 5: FINDINGS	Rating	Satisfactory

The report presents analytical findings for all evaluation questions and sub-questions that draw on robust evidence. It triangulates and synthesizes information from document and data reviews, interviews, focus group discussions, and analysis of contextual data. Findings present strengths and weaknesses, reflect the voices of different stakeholder groups, and are transparent about evidence gaps and data limitations. Topline finding statements could have omitted descriptive detail and focused on the main analytical insights. In some instances, the section could have used a more neutral language, and it should have commented on unanticipated results of CSP implementation. Finally, findings could have better disaggregated and referenced the voice of different informant (sub)groups.

CRITERION 6: CONCLUSIONS AND LESSONS	Rating	Satisfactory
---	---------------	--------------

The conclusions are substantiated by and logically flow from the findings and do not introduce new information. While some of them synthesize insights across evaluation findings and questions, others vary in their degree of analytical abstraction and could have offered a more strategic take on the findings and criteria. A specific conclusion explicitly addresses gender equality dimensions.

CRITERION 7: RECOMMENDATIONS	Rating	Satisfactory
-------------------------------------	---------------	--------------

The report makes six recommendations that address key areas for improvement identified in the findings and conclusions. Sub-recommendations provide additional operational guidance. The recommendations address both accountability and learning objectives. They identify responsible actors, are prioritized, structured as 'strategic' or 'operational', and accompanied by implementation timeframes. Some recommendations could have been strengthened by making them more specific and actionable, and addressing resource implications and contextual constraints.

CRITERION 8: ACCESSIBILITY AND CLARITY	Rating	Satisfactory
---	---------------	--------------

The report maintains a consistent narrative progression from context to findings to recommendations, using appropriate technical language and generally avoiding grammatical errors. It effectively incorporates visual aids like figures, tables, and charts, while using footnotes and annexes to provide supplementary information without disrupting flow. Summary findings are visually highlighted. However, the report exceeds recommended length and could be improved by reducing descriptive detail; consistently linking findings to the theory of change framework; using bold text more systematically to emphasize key insights; and incorporating more visual aids in the findings section to showcase good practices.

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
--	------------------------------

The mixed-method approach and evaluation methodology deliberately integrated GEWE and inclusion, with the evaluation matrix including a gender equality sub-question. Multiple data sources enhanced inclusion, accuracy, and credibility. Gender analysis appears in context and findings sections. The recommendations could have addressed gender equality and inclusion issues with more specificity, with one sub-recommendation addressing GEWE. The report could have benefited from: analyzing how intersecting identities create compounded vulnerabilities; attributing views to specific stakeholder groups; and examining whether CSP implementation had unintended effects on GEWE.

Integration of disability considerations in the evaluation report based on the UN Disability Inclusion Strategy (UN-DIS) scorecard

UN-DIS individual evaluation score	Misses requirements: 2 points
---	-------------------------------

While the evaluation terms of reference do not explicitly mention disability inclusion, the topic is referred to in the evaluation matrix and the findings provide relevant related data. Disability inclusion is, however, not addressed in the conclusions and recommendations. There was insufficient information to assess the evaluation team's knowledge of disability inclusion and whether stakeholder mapping and data collection methods involved persons with disabilities.

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.