

<b>Evaluation title</b>	<b>Mid-Term Evaluation of Strategic Outcome 2 (Sustainable Food Systems Programme, SO2) of the World Food Programme (WFP) Kenya Country Strategic Plan (CSP), in arid and semi-arid areas in Kenya, 2018-2023</b>	
<b>Evaluation category and type</b>	Decentralized - Activity	
<b>Post Hoc Quality Assessment (PHQA) – overall rating</b>	Partly Satisfactory: 59%	
<p>Overall, evaluation users of the Mid-Term Evaluation of Strategic Outcome 2 (Sustainable Food Systems Programme, SO2) of the WFP Kenya Country Strategic Plan (CSP), in arid and semi-arid areas in Kenya, 2018-2023 can rely on the report's findings despite some shortcomings in the information provided. The report's key strengths include the relevance and depth of its analysis, as well as its comprehensiveness in addressing evaluation questions across all evaluation criteria. In addition, report coherence is strong with good alignment among findings, conclusions, and recommendations. However, more relevant context analysis on the intervention zones should have been provided, as well as greater detail on the SO2 programme, specifying the geographic and programmatic scope, the range and quality of monitoring data available, and a more detailed description of the sampling frame, sample selection strategy and actual sample of respondents achieved, disaggregated by gender, age, location, beneficiary category. Moreover, the evaluation findings would have benefited from a more systematic assessment of SO2 progress and performance against its theory of change and underlying assumptions and should have included an assessment against the International Humanitarian Principles (IHP). Finally, the conclusions could have considered the contextual factors which supported/ hindered progress since 2018 while the timeframe for the recommendations could have been more realistic.</p>		
<b>CRITERION 1: REPORT SUMMARY</b>	<b>Rating</b>	Satisfactory
<p>The executive summary provides a clear and concise overview of evaluation findings, conclusions, and recommendations. The conclusions and recommendations are mostly well linked to the findings. However, the executive summary would have benefited from an expanded description of key stakeholders/users of the evaluation and a more detailed description of the context. The conclusions would also have been strengthened with an overall assessment of SO2 performance against its theory of change and the contextual factors that affected its progress.</p>		
<b>CRITERION 2: CONTEXT AND OVERVIEW OF EVALUATION SUBJECT</b>	<b>Rating</b>	Satisfactory
<p>The context section of the report provides a comprehensive description of the national policy context with appropriate reference to normative international instruments and goals. However, it would have been strengthened with more information on the general context in the intervention zones as well as more detailed analysis of intersectionality as it relates to agriculture, nutrition and food security in the ASALs, particularly given their diversity. The overview of the SO2 programme is sufficiently detailed in terms of CSP objectives, theory of change, key assumptions, types of transfer modalities, funding, initial assessments undertaken, and adaptations to SO2 relative to contextual changes but would have benefited from more detail on planned to actual transfers of in-kind and cash inputs and some mention of gender equality or social inclusion considerations.</p>		
<b>CRITERION 3: EVALUATION RATIONALE, OBJECTIVES, AND SCOPE</b>	<b>Rating</b>	Satisfactory
<p>The rationale for the mid-term evaluation and its relevance to monitoring and the final evaluation are well explained, with gender considerations mainstreamed. The evaluation objectives and timeframe of the intervention covered are clearly identified, although the scope of the evaluation could have been more explicitly articulated in terms of geographic and activity or intervention coverage of SO2.</p>		
<b>CRITERION 4: METHODOLOGY</b>	<b>Rating</b>	Satisfactory

The methodology section refers to OECD/DAC evaluation criteria and an evaluation matrix is annexed, which meets WFP requirements for content. There are three sub-questions in the matrix related to GEWE under relevance, effectiveness and impact criteria. Gender dimensions are addressed explicitly in the methodology section of the main report while limitations and associated mitigation strategies are summarized in the main report and expanded in an annex. The report would have benefited from a more comprehensive discussion of monitoring data available, including GEWE, human rights, inclusion dimensions and results. The report would have benefited from a clearer explanation of the sampling frame, sample selection strategy and actual sample of respondents achieved, disaggregated by gender, age, location, beneficiary category. The ethical standards section would have benefited from more detail on child safeguarding measures during data collection.

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

The findings are balanced and systematically address the evaluation criteria and questions complemented by annexes containing credible evidence based on qualitative and quantitative data collection from multiple sources. However, the report could have been strengthened with more frequent reference to the extent to which SO2 design and implementation was informed by and responded to the lessons and recommendations from previous studies and reviews. In addition to weaknesses identified with respect to assessing SO2 progress and performance against the theory of change components, logical framework (including outputs) and underlying assumptions, the findings do not provide an assessment of performance against the IHP which would have been applicable - and expected as stated in the TOR. Findings would have been strengthened by more systematic sex disaggregation of primary evaluation and monitoring data.

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

The conclusions logically flow from the findings, are balanced, and bring analysis up to a strategic level which serves to connect findings across evaluation criteria. However, the conclusions could have been strengthened by providing analysis of the contextual factors which supported/hindered progress since 2018, and how adept WFP was in adapting to contextual changes.

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

Recommendations are appropriately targeted and focused on improving SO2 programme performance. There is a logical line of sight between findings, conclusions, and recommendations. Recommendations could have been strengthened by more comprehensively addressing equity and inclusion considerations. Given that all the recommendations are identified as strategic in nature - so that, taken together, they imply considerable redesign of SO2 - the specified timeframes for their implementation appear very short and somewhat unrealistic.

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

The report includes all of the required lists, sources data and quotes appropriately, uses clear language, and includes adequate visual aids and relevant annexes. However, the report and its annexes exceed WFP length requirements, and the report could have signposted data from the annexes with more detail as well as made more frequent use of visual aids in the findings section, while highlighting best 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**

<b>UN-SWAP EPI – individual evaluation score</b>	<b>Approaches requirements: 4 points</b>
--	--

GEWE is mainstreamed in the evaluation while monitoring data is analysed in terms of its attention to GEWE. There are findings as well as one recommendation specifically focused on GEWE. However, not all interview protocols in annex contain gender-related questions, while the GEWE questions that are included do not focus on assessing whether the programme addressed barriers to access and control over resources and/or reduced gender inequalities. The findings could have paid more systematic attention to sex and intersectional disaggregation of data and evidence of triangulation, as well as included a more comprehensive and systematic analysis of GEWE and inclusion results across the SO2 programme. The discussion of unanticipated effects does not include sex-disaggregated data and is not analysed in terms of gender equality, intersectionality or human rights.

Post Hoc Quality Assessment – Rating scale and definitions at overall report and criteria levels	
<b>Highly Satisfactory</b>	<i>Definition at overall report level: 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>
	<i>Definition at criterion level: The criterion is addressed without any gaps or limitations.</i>
<b>Satisfactory</b>	<i>Definition at overall report level: Evaluation users can rely on the quality and credible evaluation findings provided and can use it with confidence for decision-making.</i>
	<i>Definition at criterion level: There are no significant gaps or limitations in addressing the criterion.</i>
<b>Partly Satisfactory</b>	<i>Definition at overall report level: 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>
	<i>Definition at criterion level: There are some significant gaps or limitations in addressing the criterion.</i>
<b>Unsatisfactory</b>	<i>Definition at overall report level: 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>
	<i>Definition at criterion level: There are critical gaps or limitations in addressing the criterion. Most of the required parameters are not met.</i>