

|   |   |              |
|---|---|--------------|
| <b>Evaluation title</b>   | Evaluation of Bhutan WFP Country Strategic Plan 2019-2023 |              |
| <b>Evaluation category and type</b>   | Centralized – Country Strategic Plan                      |              |
| <b>Post Hoc Quality Assessment (PHQA) – overall rating</b>  | Satisfactory: 85%   |              |
| <p>Overall, the report of the evaluation of the CSP for Bhutan (2019-2023) contains clear and concise findings, conclusions, and recommendations, based on a sound methodology, which users can rely on for decision-making. Findings generally address all evaluation questions based on triangulation across a diverse set of data sources. There is general coherence between findings, conclusions, and recommendations, and the latter are prioritized, actionable and time bound. The report would have been strengthened with a clearer narrative and visual depiction of the CSP results framework, and a clearer depiction of CSP effectiveness in the achievement of outputs and outcomes indicators and targets. Additionally, evaluation findings could have more comprehensively addressed the extent to which gender equality and women empowerment and other inclusion dimensions) were integrated into CSP design. They would have also benefited from better describing how these dimensions were mainstreamed during implementation, and what results were achieved with regard to country capacity strengthening (CCS) activities.</p> |   |              |
| <b>CRITERION 1: REPORT SUMMARY</b>  | <b>Rating</b>   | Satisfactory |
| <p>The report summary is comprehensive in its coverage and the narrative flows logically from findings to recommendations. Graphics and charts are used appropriately. Main findings are summarized against the four evaluation questions from the terms of reference. The summary introduction would have benefited from the inclusion of an overview of the evaluation rationale and objectives, as well as a more complete description of CSP strategic outcomes, outputs, and activities. In addition, evaluation conclusions present some new information not included in the findings described in the summary.</p>   |   |              |
| <b>CRITERION 2: CONTEXT AND OVERVIEW OF EVALUATION SUBJECT</b>  | <b>Rating</b>   | Satisfactory |
| <p>The context section of the report provides a good overview of the general economic, political, and geographic situation of the country, the effects of COVID, as well as key trends and data regarding CSP sectors of intervention (nutrition, agriculture/food security, education/literacy, gender equality and disability). The changes brought to the current CSP design are clearly articulated, including transfer modalities, while the integration of gender in the CSP is also addressed. However, the report could have provided more detail on the external events which informed and shaped the design and evolution of the CSP between 2018 and 2022. The report could have been strengthened with a stronger analysis on how the CSP addresses the conditions of different social groups. It would have also benefited from a graphic description of the CSP results framework presenting the relationship among CSP strategic outcomes, outputs, and activities.</p>  |   |              |
| <b>CRITERION 3: EVALUATION RATIONALE, OBJECTIVES, AND SCOPE</b>   | <b>Rating</b>   | Satisfactory |
| <p>The report presents evaluation rationale, objectives, and users. While there is no specific objective focused on GEWE, this dimension is reflected in the evaluation framework related to CSP relevance to beneficiary needs, and effectiveness in results achievement. The evaluation scope in terms of geographic and programmatic coverage should have been made more explicit in the report.</p>   |   |              |
| <b>CRITERION 4: METHODOLOGY</b>   | <b>Rating</b>   | Satisfactory |
| <p>The CSP design and methodology are sound and sufficiently robust to answer the evaluation questions. The report provides a summary of CSP evaluation design, methods, and limitations, with appropriate reference to further details found in Annexes. The evaluation methods and sources of data are clearly described and include a diverse range of stakeholders, with consideration for gender and inclusion. The methodology discussion in the report could have been strengthened with a more detailed description of the sampling frame, strategy, and selection.</p>   |   |              |
| <b>CRITERION 5: FINDINGS</b>  | <b>Rating</b>   | Satisfactory |

Findings are clearly articulated, constructive and neutral in tone, presenting both strengths and challenges of CSP performance, and are appropriately supported by triangulated data from diverse primary and secondary data sources. Findings in the report are structured against the four evaluation questions and sub-questions articulated in the evaluation matrix. There is internal consistency among findings, and where information gaps prevent assessment, they are identified and explained. Findings present an appropriate level of detail to support key messages and explain strengths and challenges. They also discuss the impact of COVID on output delivery during the period under review, and discuss limitations of the CSP logframe, country office monitoring, and the availability of data. In addition, they assess how these challenges affected WFP's ability to track and report on CSP output to outcome achievement. Sub-questions presented in the evaluation matrix relating to cross-cutting issues could have been more comprehensively discussed in findings, as per lines of inquiry and indicators included in the evaluation matrix. The representation of diverse stakeholder voices in findings would have benefited from greater disaggregation by sex, age, disability where feasible. Finally, findings would have been strengthened with an explicit discussion of recent evaluation recommendations relating to the current CSP and to what extent they have been addressed.

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

Conclusions are balanced, exploring areas of strength and challenge for WFP in Bhutan. Conclusions cover key messages raised in findings bringing the analysis up to a more strategic level, foreshadowing areas for recommendations. Conclusions include a summary analysis of internal CSP coherence and logic, and challenges encountered with corporate indicators and the absence of a theory of change. The conclusions also address CSP alignment with national priorities and the SDGs. They are not structured against the evaluation questions, so the line of sight with findings is slightly hampered. Since all CSP outputs and outcomes relate to CCS, the conclusions section would have benefited from a greater focus on the integration of GEWE, equity and inclusion in country capacity strengthening and related results achievement.

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

Recommendations flow logically from findings and conclusions and are aligned with the evaluation objectives. They are realistic, feasible, targeted and easily actionable, with priority level and responsibility clearly articulated. However, recommendations on improving GEWE mainstreaming in the next CSP are unusually broad and not necessarily aligned with evaluation findings and conclusions on GEWE issues of the current CSP.

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

The report is written in clean and clear language, and information is appropriately sourced using footnotes. References to annexes are appropriately signposted. Good use is generally made of visual aids and key messages are highlighted in bold. The report would have benefited from a final edit to remove typos and grammatical errors and acronyms should have been consistently spelled out the first time used.

**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> | Meets requirements: 7 points |
|--|------------------------------|

GEWE is integrated in the evaluation scope of analysis. The evaluation framework was designed in a way that ensured GEWE-related data would be collected. Gender-responsive methods and tools, and data analysis techniques were selected to a certain extent, although details are limited. Hence, the methodology section could have been strengthened with a more comprehensive treatment of GEWE. The evaluation findings, conclusions, and recommendations reflect GEWE considerations. However, overall, the ER could have gone farther in its analysis of multi-dimensional poverty and the social, economic, religious, gender, age, and other intersectional inequities in the country, particularly as they affect CSP sectors of intervention (agriculture, food security, education outcomes, child marriage, climate vulnerability, etc.).

**Post Hoc Quality Assessment – Rating scale and definitions at overall report and criteria levels**

POST HOC QUALITY ASSESSMENT OF WFP EVALUATIONS

|                            |   |
|----------------------------|---|
| <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>  |