

Evaluation title	Evaluation of China WFP Country Strategic Plan 2022 - 2025	
Evaluation category and type	Centralized	
Post Hoc Quality Assessment (PHQA) – overall rating	Highly Satisfactory: 92%	
<p>The <i>Evaluation of China WFP Country Strategic Plan 2022–2025</i> presents credible, well-substantiated findings that users can rely on confidently for decision-making. The report provides a clear and up-to-date description of the country context, identifies relevant vulnerable groups, and appropriately links CSP components to national strategies, the previous CSP, and UN frameworks. The purpose, objectives, users, and stakeholders are clearly outlined, and gender and wider equity and inclusion considerations are well integrated. Evaluation questions are well formulated, and the theory-based mixed-methods design is appropriate and applied with sufficient rigor. Findings are balanced, evidence-based, and demonstrate triangulation. Conclusions synthesize insights across questions and are logically connected to the evidence presented. Recommendations are actionable, forward-looking, and aligned with both learning and accountability objectives. Nevertheless, there are a few minor shortcomings. The reconstructed theory of change is only briefly presented and could have benefitted from a fuller narrative. In addition, private sector actors—important CSP partners—are not identified as evaluation users and some findings lack sex-disaggregated primary data that limits understanding of how different social groups’ perspectives are reflected. The findings-conclusions-recommendations mapping could have been clearer by linking each conclusion to relevant findings, and some sub-recommendations remain broad and could have been more precisely formulated.</p>		
CRITERION 1: REPORT SUMMARY	Rating	Satisfactory
<p>The summary evaluation report (SER) provides a concise introduction outlining the evaluation purpose, scope, methodology, and context. Key findings effectively support the main conclusions and messages, and recommendations are fully aligned with those in the main report. Visual aids are clear and strengthen the narrative. However, the SER cannot be fully used in isolation of the main report given a few shortcomings. The SER does not provide a clear picture on the rationale for the CSP or the evaluation and the evidence underpinning some key messages is not explicitly provided.</p>		
CRITERION 2: CONTEXT AND OVERVIEW OF EVALUATION SUBJECT	Rating	Highly Satisfactory
<p>The report provides a clear and concise description of the country context, drawing on data across relevant sectors and clearly identifying vulnerable groups targeted by WFP interventions. CSP components are well described with clear articulation of the strategic and operational shifts between the two CSP periods. The inclusion of a timeline of WFP interventions from 2015–2024—situated within national strategies and UN frameworks—usefully enhances the reader’s understanding of the programme’s evolution. However, the presentation of the theory of change could have benefitted from additional narrative explanation to more fully elaborate the components shown in the figure.</p>		
CRITERION 3: EVALUATION RATIONALE, OBJECTIVES, AND SCOPE	Rating	Satisfactory
<p>The report clearly identifies the evaluation purpose, objectives, primary users, and stakeholders, appropriately mainstreaming gender and equity considerations. However, Chinese private sector organizations are not mentioned as evaluation users, despite their relevance under CSP Activity 3 on Partnerships. In addition, no explanation is provided on why the evaluation period begins in November 2020 despite CSP implementation starting in 2022.</p>		
CRITERION 4: METHODOLOGY	Rating	Highly Satisfactory
<p>The evaluation questions and sub-questions are well formulated and aligned with the evaluation’s context and objectives. The evaluation matrix is comprehensive, linking questions, sub-questions, dimensions of analysis, lines of inquiry, indicators, data sources, and data collection and analysis methods. The theory-based mixed-methods design is appropriate and applied rigorously, with strong emphasis on triangulation, validation, and inclusion of diverse stakeholder and beneficiary perspectives. Ethical principles are explicitly addressed, and the evaluability assessment conducted during inception provides a coherent link between identified challenges and the methodological choices</p>		

made. However, the distinction between lines of inquiry and indicators in the matrix is not always clear, and some indicators could have been more appropriately formulated.		
CRITERION 5: FINDINGS	Rating	Highly Satisfactory
The findings are clearly derived from in-depth analysis and well supported by diverse sources, demonstrating strong triangulation across primary and secondary data. They are presented in a balanced and objective manner, with appropriate caveats where evidence is limited. The findings comprehensively address all evaluation questions and sub-questions, with a good balance between detail and synthesis. The use of shaded text boxes to highlight key findings for each sub-question enhances clarity and shows full coverage of the evaluation framework. However, while quantitative data are often sex-disaggregated, it is difficult to assess the extent to which qualitative evidence reflects the perspectives of different social groups, as primary qualitative data are not disaggregated by sex.		
CRITERION 6: CONCLUSIONS AND LESSONS	Rating	Satisfactory
The conclusions draw effectively on multiple findings and are logically connected to the evidence presented, without introducing new information. They appropriately reflect CSP performance on gender and equity dimensions and demonstrate alignment with the evaluation questions. However, some conclusions do not clearly cut across findings from all relevant evaluation questions, and the findings–conclusions–recommendations mapping could have more clearly indicated how each conclusion links to specific findings, to improve transparency and traceability.		
CRITERION 7: RECOMMENDATIONS	Rating	Highly Satisfactory
The main recommendations are actionable, well-structured, and coherently supported by corresponding sub-recommendations. They are forward-looking, balanced between learning and accountability, and clearly derived from the evaluation’s findings and conclusions. Responsibilities are clearly assigned—primarily to the WFP Country Office with relevant contributing entities identified—and the recommendations are realistic, internally consistent, and prioritized with appropriate deadlines. However, while several sub-recommendations are specific and actionable, others remain overly general and could have benefitted from further refinement to enhance their practical utility.		
CRITERION 8: ACCESSIBILITY AND CLARITY	Rating	Highly Satisfactory
The report is well written, well structured, and maintains internal consistency. Visual aids are used effectively to support the narrative, and the report makes strong use of footnotes. Key messages are clearly highlighted through bold text and shaded boxes, enhancing readability.		

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	Meeting requirements	
GEWE and equity considerations are mainstreamed in the evaluation, with evaluation question two directly addressing these issues and the evaluability assessment clearly linking GEWE data limitations to methodological choices. The mixed-methods approach is appropriate, using sex-disaggregated samples and seeking diverse perspectives. Triangulation and ethical safeguards are clearly described. Vulnerable groups and some intersecting vulnerabilities are identified. However, qualitative evidence is not sex-disaggregated, making it difficult to assess how different social groups are reflected in the findings.		

Integration of disability considerations in the evaluation report based on the UN Disability Inclusion Strategy (UN-DIS) scorecard		
UN-DIS individual evaluation score	Meeting requirements	
Disability inclusion is integrated throughout the evaluation design and implementation. The evaluation includes a sub-question on disability inclusion, and the evaluation matrix integrates specific dimensions of analysis, lines of inquiry, and indicators. The sampling frame included persons with disabilities, and the team conducted a visit to a project focusing		

on disability inclusion, allowing interviews with women and men with disabilities and an organization of persons with disabilities. Findings and conclusions explicitly reference persons with disabilities among key vulnerable groups.

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.