Evaluation title	Evaluation of WFP's Enterprise Risk Management Policy (2018)
Evaluation category and type	Centralized
Post Hoc Quality Assessment (PHQA) – overall rating	Satisfactory: 88%

The *Evaluation of WFP's Enterprise Risk Management Policy (2018)* provides credible evidence to support decision-making. It offers a clear and comprehensive overview of the Enterprise Risk Management (ERM) Policy, its implementation context, and the evaluation's purpose, scope, and methodology. Findings are well-structured, directly address all evaluation questions, and are supported by well triangulated primary and secondary data. The report establishes plausible causal linkages between implementation efforts and results, with appropriate consideration of contextual factors. Conclusions synthesize findings into higher-level insights, and recommendations are clear, actionable, and well grounded. The report is logically structured, uses visual aids effectively, and is written in accessible, professional language. However, the report could have been further strengthened by providing additional context on how enterprise risk management intersects with protection and inclusion, expanding the discussion of methodological limitations and mitigation strategies, and explicitly acknowledging potential resource implications and implementation challenges for some recommendations.

CRITERION 1: REPORT SUMMARY	Rating	Highly Satisfactory

The summary evaluation report (SER) offers a clear, accurate, and useful synthesis of the evaluation's key features, context, subject, main messages and recommendations. As such, it can be read as a stand-alone document.

CRITERION 2: CONTEXT AND OVERVIEW OF EVALUATION SUBJECT	Rating	Satisfactory

The report provides a comprehensive and well-structured overview of the ERM Policy and its context. It effectively describes the challenging operational environment in which WFP operates and situates the ERM Policy within the broader UN system. The policy's strategic direction, goals, and objectives are clearly articulated, supported by a reconstructed theory of change that outlines expected pathways from inputs to impact and illustrates the implementation framework and institutional arrangements. The context and subject sections could have been strengthened by explaining how enterprise risk management—and the ERM Policy specifically—intersects with protection and inclusion considerations, particularly how risk vulnerabilities may differ by gender, age, disability, or ethnicity. Furthermore, budget-related information could have been presented under this section, rather than deferring this to the findings section.

CRITERION 3: EVALUATION RATIONALE, OBJECTIVES, AND SCOPE	Rating	Satisfactory

The report clearly outlines the evaluation purpose, objectives, scope, and primary users and stakeholders. The report notes that gender equality considerations are mainstreamed into the evaluation objectives and scope; however, additional information on how mainstreaming was implemented could have provided a more complete overview.

CRITERION 4: METHODOLOGY	Rating	Satisfactory

The evaluation uses an appropriate mixed-methods, theory-based approach with strong alignment between questions, data collection tools, and analytical methods. The evaluation matrix clearly links sub-questions and indicators to data sources, supporting systematic analysis of the theory of change. Country case study sampling is transparent and well justified, and the report clearly explains the use of triangulation to strengthen validity. The report could have been strengthened by providing more detail on the evaluability assessment—particularly data availability and quality issues affecting gender equality and inclusion analysis—and by offering a fuller discussion of methodological limitations beyond the two noted, including relevant data gaps.

CRITERION 5: FINDINGS	Rating	Satisfactory

The report presents analytical findings that respond directly to all evaluation questions and sub-questions, synthesizing evidence from diverse sources into a coherent narrative. It highlights both strengths and weaknesses of the ERM Policy and its implementation, analyzes planned versus actual results, and establishes plausible causal linkages while

consistently considering contextual influences on outcomes. Given the constraints of evaluating an organizational policy, the report demonstrates a reasonable level of attention to gender and inclusion. However, the evaluation could have been strengthened by further addressing these dimensions—for example by examining how risk identification, risk assessment, and strategic planning processes incorporate gender considerations or whether risk management trainings incorporate content on gender-sensitive and inclusive approaches.

CRITERION 6: CONCLUSIONS AND LESSONS	Rating	Highly Satisfactory

The conclusions go beyond summarizing findings to provide higher-level insights that draw meaningful connections across multiple evaluation questions. They flow logically from the evidence and do not introduce new information. The conclusions could have been strengthened by framing identified gaps in gender equality and broader inclusion as strategic weaknesses requiring focused attention, rather than simply noting that these dimensions were underplayed in the current ERM Policy.

CRITERION 7: RECOMMENDATIONS	Rating	Satisfactory

The recommendations provide a clear and structured roadmap for strengthening WFP's enterprise risk management. They specify the required changes, outline concrete actions linked to identified gaps, and are appropriately prioritized with realistic timeframes. Responsibilities are clearly assigned, supporting accountability. The recommendations balance ambition with practicality, addressing policy design, organizational culture, systems and tools, resources, and partnerships, and include sub-recommendations explicitly targeting gender equality and broader inclusion. Some recommendations could have been strengthened by more clearly acknowledging potential resource implications and implementation challenges, and by ensuring sufficiently precise language when outlining specific operational actions.

CRITERION 8: ACCESSIBILITY AND CLARITY	Rating	Satisfactory

The report is clearly structured, follows a logical flow, and makes effective use of visual aids. It documents sources consistently, provides strong cross-referencing between the main report and annexes, and is written in clear, professional, and accessible language. The report follows standard WFP templates and includes all required annexes. However, readability could have been improved in some instances through simplification of long and complex sentences and replacement of passive voice with active voice.

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

The evaluability assessment did not address gender and inclusion-related data gaps, which are only discussed later in the methodology section. GEWE is mainstreamed through the evaluation's objectives, criteria, indicators, and methodological elements, supported by an appropriate mixed-methods design, strong triangulation, and adherence to UNEG norms and ethics. However, the introductory sections do not explain how ERM relates to inclusion and protection, and the report does not disaggregate internal perspectives. Evaluation findings include explicit attention to gender equality issues, and these dimensions are also addressed under two sub-recommendations. However, unanticipated effects are not identified, including any linked to GEWE.

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

The evaluation TOR includes explicit reference to mainstreaming of disability considerations in the approach and methodology. Disability inclusion considerations are captured under evaluation sub-question 1.1 through indicator 1.1.5. However, there is limited evidence that persons with disabilities or representative organizations participated in stakeholder mapping or data collection. The findings note that cross-cutting issues, including disability, receive limited attention in the ERM Policy and related processes. In addition, while conclusions and recommendations do not explicitly mention disability inclusion, the report clarifies that references to cross-cutting issues encompass disability.

Post Hoc Quality Assessment – Rating scale and definitions at overall report and criteria levels	
Highly Satisfactory	*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.
	Definition at criterion level: The criterion is addressed without any gaps or limitations.
Satisfactory	*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.
	Definition at criterion level: There are no significant gaps or limitations in addressing the criterion.
Partly Satisfactory	*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.
	Definition at criterion level: There are some significant gaps or limitations in addressing the criterion.
Unsatisfactory	*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.
	Definition at criterion level: There are critical gaps or limitations in addressing the criterion. Most of the required parameters are not met.