

UNFPA Evaluation Quality Assessment Grid

Version: May 2024

REPORT RATING SUMMARY			
Overall Rating		80%	Satisfactory
• • • • •	Excellent	5	
• • • •	Highly Satisfactory	4	
• • • -	Satisfactory	3	The report meets UNFPA/UNEG standards for evaluation reports, but some indicators are inadequately addressed or missing. Decision makers may use the evaluation with some confidence.
• • - -	Fair	2	
• - - -	Unsatisfactory	1	

REPORT DETAILS	
Title of the evaluation report	Government of Timor-Leste and UNFPA: 4th Country Programme Evaluation, 2021-2025
Region	UNFPA Asia and Pacific Regional Office
Country	Timor-Leste
Year of report	2025
Business Unit/programme country (managing evaluation)	Timor-Leste
Date of assessment review (dd/mmm/yyyy)	7-Apr-25
Name of assessment review firm	IOD PARC
CLASSIFICATION OF EVALUATION REPORT	
Primary SDG(s) covered (list provided below)	3, 5, 10, 16, 17
UNFPA Strategic Plan areas covered (lists provided below)	
Three transformative results	All
Six outputs	All
Six accelerators	All
Organizational effectiveness and efficiency	no
Humanitarian evaluation	yes
Evaluation evaluated (e.g. country programme/intervention/policy/thematic area)	Country Programme
Evaluation type (e.g. formative, summative, developmental)	Summative and formative
Geographic scope (e.g. global, regional, national)	National
<p>EQA Summary: The rater will provide top line issues for this evaluation relevant for feedback to senior management (strengths and weaknesses), summarizing how the evaluation report meets or fails to meet all criteria. As relevant, the rater will highlight good practice/added value elements and the level of complexity of the evaluation. The rater should also highlight how cross-cutting issues were addressed in the report. Considerations of significant constraints (e.g. humanitarian crisis or political turmoil) should also be highlighted here.</p> <p>This is a satisfactory evaluation report of the Government of Timor-Leste and UNFPA's Fourth Country Programme, 2021-2025. The key strengths and weaknesses of the evaluation report are provided below, followed by suggestions for future evaluations.</p> <p>Strengths:</p> <ul style="list-style-type: none"> • Section A: Executive Summary – The executive summary is clearly written and can serve as a standalone document that is useful for informing decision-making. • Section B: Background – The evaluation report provides a clear description of the Country Programme and the context of the programme in Time-Leste. A key facts table is also provided in the opening pages of the report, which include data against indicators on maternal mortality, contraceptive prevalence rate, unmet needs for family planning and so on. • Section C: Evaluation Purpose, Objectives and Scope – The purpose and objectives are described and key audiences identified. The scope is clear and further elaborated from what was provided in the ToRs, such as inclusion of humanitarian emergencies. • Section D: Evaluation Design and Methodology – The evaluation questions were amended during the inception phase and clearly explained in the evaluation report, which is good practice. The evaluation matrix is also clearly presented in Annex 1. • Section E: Evaluation Findings – The findings are clearly presented. They are structured according to UNFPA Evaluation Handbook (2024) guidance. The findings follow the logic of the evaluation matrix, making it easy to reference between the two. The evaluation uses credible forms of qualitative and quantitative data, which is presented in the evaluation matrix. • Section F: Evaluation Conclusions – The conclusions are clearly formulated and offer deeper insight and analysis that go beyond restating the findings. • Section G: Recommendations – The recommendations are useful and actionable for primary intended users. They are categorised into strategic and programmatic recommendations. There is a priority level rating for each recommendation. In addition, the start of the chapter describes the process for developing recommendations, which is good practice and helpful to know the stakeholders involved in developing and/or validating the recommendations. <p>Weaknesses:</p> <ul style="list-style-type: none"> • Section A: Executive Summary – The executive summary's methodology section is overly detailed, detracting from its primary focus on evaluation results, while the key findings are too concise and require further elaboration to provide necessary nuance. • Section B: Background – A stakeholder map is provided in Annex 6 and rights holders, though not included in the mapping, are identified in the main text of the report. What is missing are the roles and responsibilities of the duty bearers as well as an analysis of stakeholders to better understand stakeholders' interests, needs, power, influence and potential impact. • Section C: Evaluation Purpose, Objectives and Scope – There are slight changes made to the objectives stated in the ToRs but these changes are not explained or justified. • Section D: Evaluation Design and Methodology – The description of the Theory/ies of Change (ToC) is an area that could be improved to better understand what areas are being tested by the evaluation. This includes the causal relationships. For example, the main report describes the reconstruction of the SRHR ToC and then describes the refinement of the overarching ToC. Furthermore, Annex 5 includes a table titled 'Discussion notes on the above ToC' noting down limitations of the ToC and corresponding action points. However, it is not clear to what extent these actions were addressed. • Ethical considerations could be strengthened by providing further contextualisation. For example, while adherence to principles of independence, impartiality, credibility and utility are mentioned, it is unclear how they were applied in the evaluation. • Section E: Evaluation Findings – While the Evaluation Report describes what worked or did not work, there is less a discussion about the causal pathways and a deeper exploration of why and what led to the results (or otherwise), particularly as this is a theory-based evaluation. • Section F: Evaluation Conclusion – There are several instances where it is difficult to see links between the conclusion statements and sign-posted findings. Please see Q14ii below for examples. • Section I: Cross-cutting issues – While the data collection methods are intended to capture the voices and perspectives of a wide range of stakeholders, including rights holders, the sampling for rights holders/beneficiaries is unclear. The evaluation report also includes some analysis of intersectionality, but it is not systematic. <p>Suggestions for future evaluators: The rater will identify key suggestions to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, examples will be cited to assist evaluation managers in overseeing future evaluations.</p>	

Recommendations:

- **Section A: Executive Summary** – While the executive summary is well written and presented in a concise and clear manner, it would be helpful to provide more information on the key findings, particularly as it is well within the page limits. Evaluators can also consider reducing the section on methodology to allow more space for the findings.
- **Section B: Background** – Evaluators are encouraged to include the role and responsibilities of stakeholders to better understand their contributions and interconnections between them. It would also be helpful to include a dedicated analysis of stakeholders to understand their interest, influence and potential impact.
- **Section C: Evaluation Purpose, Objectives and Scope** – Evaluators can be encouraged to provide an explanation and/or justification for the changes made to the objectives stated in the ToR.
- **Section D: Evaluation Design and Methodology** – To improve readability and clarity of the ToC, the evaluation team can consider discussing each ToC separately in the main report, signposting to the appropriate diagram and results framework in the annex. In this way, the reader is not entirely reliant on the diagram to understand what parts of the results chain will be tested by the evaluation. In addition, it would be helpful to state up front what aspects of the ToC, the evaluation would examine through the evaluation and to describe it in a systematic way in the findings section. Beyond merely highlighting what works or does not, evaluators are encouraged to dig deeper to understand why. For example, a thorough exploration of the ToC will enable the team to go further beneath the surface.
- In addition, evaluators are encouraged to describe how ethical considerations, norms and standards, were concretely applied to the evaluation.
- On the sampling, evaluators can consider elaborating on the selection of stakeholders for each data collection method. For instance, a purposive sample was used for field visits and the key criteria included. However, the selection of focus groups is not discussed in the sampling. Some of this information is, instead, provided in a different section where it describes limitations to the selection process. In addition, a list of focus groups is not provided in Annex 3 and so it is unclear how many individuals participated in these groups. The main recommendation is for evaluators to provide a more consistent approach to sampling to ensure the selection of all stakeholders for each data collection method is provided.
- **Section E: Evaluation Findings** – To improve the quality of the findings, evaluators should delve deeper into why the intended results described in the ToC occurred or did not occur, rather than merely describing what results materialized. As this is a theory-based evaluation, it would be important to examine the underlying causal mechanisms more in-depth.
- **Section F: Evaluation Conclusion** – While the conclusions are well written, there are several instances where it is difficult to see the links between the conclusion statements and the sign-posted findings. Q14ii provides some examples. Evaluators are encouraged to check the sign-posts to enhance transparency, as well as improve navigation across a lengthy report.
- **Section I: Cross-cutting issues** – Currently, rights holders are not included in the stakeholder sampling and list of persons met. This would be helpful to document alongside the limitation of access to stakeholders. Evaluators can also be encouraged to reflect on how they apply an intersectional analysis so that it is more systematic.

SECTION RATINGS

SECTION A: EXECUTIVE SUMMARY (weight 5%)		67%	Comments on Rating
Question 1.	Can the executive summary inform decision-making?		
i	Is a clear, standalone document useful for informing decision making, (a minimum of 5 pages, up to a maximum of 7 pages).	Yes	The Executive Summary is clearly written and can serve as a standalone document that is useful for informing decision-making. It is 6 pages in length.
ii	Includes all necessary components of the evaluation report, including: (1) overview of the context and intervention, (2) evaluation purpose, objectives and intended users, 3) scope and evaluation methodology, (4) summary of most significant findings, (5) main conclusions and (6) key recommendations	Partially	The executive summary covers most of the essential components expected in an evaluation report. These include: (1) an overview of the Country Programme (p. 1); however, it does not include the broader context, which would be important to add to fully meet this criterion; (2) the evaluation purpose and intended audience, with a summary of some evaluation objectives (p.1); (3) the scope and methodology (p. 1–2); (4) a summary of key findings (p. 3–4); (5) main conclusions (p.4–5); and (6) key recommendations (p.6).
iii	Includes all significant information in a concise yet clear manner to understand the theme, intervention, programme, project and the evaluation.	Partially	Overall, the main information is presented in a concise and clear manner to understand the programme and the purpose of the evaluation. What would be helpful is for the Executive Summary to ensure that information on intended users is more concise. In addition, evaluators can provide more information on the key findings as it is well within its page limit. Evaluators can also consider reducing the section on methodology as it is not essential for an executive summary to provide information on sampling, for example.
SECTION B: BACKGROUND (weight 5%)		80%	Comments on Rating
Question 2.	Is the evaluand (i.e. intervention/policy/thematic area etc. that is to be evaluated) and context of the evaluation clearly described?		
i	Clear description of the evaluand (e.g. intervention), including: geographic coverage, implementation period, main partners, cost/budget, and implementation status.	Yes	There is a clear description of the Country Programme (CP) in chapter 3 (p.26). The chapter begins by framing the CP by first outlining UNFPA's strategic plans (p.25), which provides a helpful overview for the Country Programme, current and previous. The current CP description begins on p.27 (section 3.2.2), which includes the implementation status. It includes geographical coverage (section 3.2.3, p.30) and budget (see Table 6 and 7) according to outcome areas and also by year. While the main partners are not mentioned in this chapter, they are identified in chapter 1 (p. 10). These include government partners, development partners (such as other UN agencies), and civil society organisations. There is also a stakeholder map, in table format, in Annex 6, which provides more detail. It also includes partners from academia.
ii	Clear description of the context of the evaluand (e.g. economic, social and political context, relevant aspects of UNFPA's institutional, normative and strategic framework, cross cutting issues such as gender equality and human rights, disability and LNOB dimensions) and how the context relates to the evaluand (e.g. key drivers and challenges that affect the implementation of the intervention/policy/thematic area	Partially	There is a clear description of the context of the programme in Timor-Leste in Chapter 2. This includes the socio-economic context (p. 18); development challenges relevant to the Country Programme (p. 18); national strategies such as Timor-Leste's Strategic Development Plan 2011-30, National Youth Policies, and Gender Equality and Empowerment of Women and strategies for managing population dynamics (p. 19-23). The Evaluation Report also touches on the role of external assistance in the country, which indicates percentage of funding received on areas relevant to the programme (p. 23). Cross-cutting issues such as disability (p.21, 22), gender and GBV (p.19-23) are identified. Where it can improve is a more explicit description of LNOB as this principle underpins UNFPA's Strategic Plan 2022-2025. This could include the specific status or situation of vulnerable and marginalised groups such as GBV rates among women and disabilities. For this reason, it is rated as 'partial'.
iii	Linkages drawn between the evaluand and the ICPD benchmarks and SDGs relevant targets and indicators.	Yes	The linkages between the Country Programme and ICPD-related SDGs are outlined in Chapter 3 (p. 27). These are mainly SDGs 3 and 5. Findings 1 and 2 delve into the SDG targets in more detail as EQ asks about country programme alignment with relevant national SDG targets. As this is a Country Programme Evaluation, a key facts table that includes SDG progress is included in the opening pages of the Evaluation Report. This includes indicators on maternal mortality, contraceptive prevalence rate, unmet need for family planning and so on. These are also referenced in the background section of the report (p. 18 onward).
Question 3.	Are key stakeholders clearly identified and analysed?		

	i Clear identification of key stakeholders which should include implementing partner(s), development partners, rights holders, and duty bearers among others; and of linkages between them (e.g., stakeholder map).	Yes	A stakeholder map is included in Annex 6, which was provided by the UNFPA Country Office. The stakeholder map is referenced in Chapter 1 (p. 10), signposting to Annex 6, and provides a list of stakeholders including 'national and local government, UN agencies, donors, and development partners' (p.10). While rights holders are not identified in the stakeholder map, they are identified in the main text of the report and include: 'women of childbearing age, pregnant women and newborn, adolescent and youth, GBV survivors, PLWHIV, PWDs, LGBTQI, student population and community groups' (p.10). What would be helpful is for evaluators also include the role and responsibilities of each of these stakeholders to better understand their contribution and interconnections between them.
	ii Stakeholders are analysed to understand their specific rights, duties, needs, interests, concerns, and potential impact on the evaluand.	Partially	There is no explicit analysis of stakeholders to better understand their interests, needs, power, influence and potential impact. However, some of this information is provided in the context part of the evaluation report. For example, under the section on Development Challenges, the evaluation report describes the rates of teenage pregnancies and rising prevalence of HIV/AIDS and STIs (p.18). It also describes the maternal mortality rates and SGBV prevalence. From the perspective of duty bearers, the background section outlines national policies and strategies relevant to the intervention. For these reasons, this criterion is rated as 'partially' met. Evaluators are encouraged to provide a dedicated analysis to better understand the needs, interests, power, influence and potential impact of the intervention for stakeholder identified, particularly rights holders and duty bearers.
SECTION C:	EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)	83%	Comments on Rating
Question 4.	Is the purpose of the evaluation clearly described?		
	i Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.	Yes	<p>The purpose of the Country Programme Evaluation (CPE) is provided under Sub-section 1.1 (p. 7). It elaborates on the purpose provided in the ToRs (p. 6 of ToRs) to include accountability, lessons to inform the next programme cycle and the implementation of the ICPD. This purpose is aligned with UNFPA's Evaluation Policy 2024.</p> <p>It describes why it is needed at this point in time as the current Country Programme runs from 2021-2025, it is recommended that a final evaluation is conducted in the penultimate year (p.7).</p> <p>The intended use is also provided in the same section that the 'evaluation results are crucial for designing the new country programme document...' (p.7).</p> <p>The key intended users are identified in the last paragraph of the sub-section. It is fairly comprehensive and includes both primary and secondary audiences (p. 7).</p>
Question 5.	Are the objectives and scope of the evaluation clear and realistic?		
	i Clear and complete description of the objectives of the evaluation, including reference to any changes made to the objectives included in the ToR (if applicable).	Partially	While there is no sub-section titles that differentiate between purpose and objectives, it is clear when the Evaluation Report describes the objectives as the starting sentences states: 'The objectives of the evaluation...' (p.7). There are slight changes to the objectives stated in the ToR as the evaluation report objectives no longer refer to UNFPA's contribution to UNSCDF result groups. In addition, the last bullet points specifies recommendations for both UNCT and UNFPA CO, which were not in the ToRs. However, the evaluation report states that no changes were made from the ToRs, and the justification is not provided. For this reason, it is rated as 'partial'.
	ii Clear and relevant description of the scope (e.g. thematic, geographic, and temporal) of the evaluation, covering what will and will not be covered, as well as, if applicable, the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political, humanitarian or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Yes	There is a clear sub-section for scope (see 1.2, p. 8). Within this sub-section, the evaluation report provides the temporal, geographical and thematic scope. These are all clearly identified and described. It also further clarifies from the ToR what component of humanitarian emergencies are included in the evaluation, which were not initially intended. It also identifies cross-cutting aspects including human rights-based approach, integration of LNOB and so on (p.8).
SECTION D:	EVALUATION DESIGN AND METHODOLOGY (weight 20%)	73%	Comments on Rating
Question 6.	Are the selected evaluation questions and evaluation criteria appropriate for the purpose of the evaluation and is there clear justification for their use?		
	<p>Note: UNFPA evaluation standards refer to the OECD/DAC criteria such as: relevance, coherence, effectiveness, efficiency and sustainability (not necessarily applicable to all evaluations) and, for country programmes that include circumscribed and limited humanitarian and/or emergency interventions, the criteria of coverage and connectedness.</p>		

	i Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Yes	<p>The evaluation questions and sub-questions are appropriate for meeting the objectives (as outlined in the ToR and the evaluation report) and the purpose of the evaluation. Initially, the extent to which the third objective in the ToR (i.e., UNCT coordination mechanisms) was addressed is unclear. However, assumption 29 (under EQ8) specifically looks at UNFPA CO's contribution to UNCT (Annex 1) and is further explored in assumption 30 (Annex 1).</p> <p>The evaluation questions differ slightly from those in the ToRs. However, the changes are explained in Chapter 1 (p. 9). This includes changes to reflect humanitarian emergencies that occurred during the Country Programme (CP4). These changes are also elaborated on in Table 4 of the Final Design Report (p. 28-30).</p> <p>OECD-DAC criteria are specified and evaluation questions are organised according to the criteria. These include relevance, effectiveness, efficiency, sustainability, coherence, coverage and connectedness.</p>
	iii Evaluation matrix clearly presents the evaluation criteria used as well as the corresponding evaluation questions, indicators, lines of inquiry, benchmarks, assumptions, source of data, methods for data collection and analysis, and/or other processes from which the analysis can be based, and conclusions drawn.	Yes	<p>The Evaluation Matrix is clearly presented in Annex 1. It includes all six elements for each of the evaluation questions such as (i) evaluation criteria; (ii) evaluation questions; (iii) assumptions associated with each evaluation question; (iv) assessment indicators; (v) data collection methods and (vi) data sources.</p> <p>On the latter two elements, the data collection methods are noted as 'sources of information' and the types of sources are provided below each source. For example, for Assumption 25 (EQ5), one of the sources identified is 'secondary data' and the list of the types of documents are provided alongside, such as 'UNFPA CPD, CPDAP, Annual Report, Reports of response to COVID and floods'. This is comprehensive and consistent across the whole matrix.</p> <p>Additionally, the evaluation team has included notes on the key findings against each assumption. While this information is helpful, it makes navigating the Evaluation Matrix (currently spanning 89 pages in length) more challenging. Evaluators can consider providing the matrix separately without the findings, for ease of reference. However, this is a suggestion for consideration and does not affect the rating.</p>
Question 7.	Is the theory of change, results chain, logical framework, or equivalent framework well-articulated?		
	i Clear description of the intervention's intended results, or of the parts of the results chain that are applicable to, or are being tested by, the evaluation.	Partially	<p>There is a clear description of the Country Programme (CP) in the background section of the evaluation report (see Table 5 on p. 28).</p> <p>The evaluation team employed a theory-based approach, which outlines the process of reconstructing the outcome-level ToCs. This section can be better described as it jumps between the ToCs, for example, explaining the reconstruction of the SRHR ToC and then describing the refinement of all ToCs. It then discusses a singular ToC, but it is unclear to which one it is referring.</p> <p>Annex 5 includes diagrams of four ToCs, according to the four CP outcomes. However, they are not legible. The evaluation team recognises this issue (see p.3 of Annex 5) and state that the PPT version was attached but this was not provided to the reviewer. Furthermore, Annex 5 includes a table titled 'Discussion notes on the above ToC' noting down limitations with the ToC and corresponding action points. However, it is not clear the extent to which these actions were addressed.</p> <p>To improve readability and clarity, the evaluation team can consider discussing each ToC separately in the main report, signposting to the appropriate diagram and results framework in the Annex. This way, the reader is not entirely reliant on the diagram to understand what parts of the results chain will be tested by the evaluation. The output and outcome tables and graphs in Annex 5 (titled '<i>Additional information</i>') would also benefit from an introduction to orient the reader to the information that is being provided.</p>

	<p>ii Causal relationships between the various elements (e.g. outcomes, including the three or relevant Transformative Results, outputs) of the theory of change, results chain or logical framework are presented in narrative and/or graphic form).</p>	Partially	<p>There is limited description of the causal relationships between the different elements of the ToC in the main body of the report (p. 11-12). It primarily focuses on describing the links to the Transformative Results and the process of refining the ToC with the country teams. While it mentions that the 'ToC helped in understanding the causal connections...' (p.12), it does not describe them. In addition, the ToCs also lack problem statements, including the 'root causes', which the intervention aims to address.</p> <p>As mentioned above, the outcome-level ToCs are presented in Annex 5 but they are not legible. Nevertheless, it is possible to see that assumptions are identified in each of the ToC, which provides clear areas for assessing. However, the text within these assumptions is difficult to read.</p> <p>Additionally, there are two ToCs for Sexual and Reproductive Health and Rights (SRHR). Since the main report outlines the refinement process, it is assumed that the first diagram represents the original ToC, while the second is the reconstructed version.</p> <p>To enhance clarity, it would be beneficial to provide a clearer explanation of the presented diagrams, including the causal relationships, particularly as the evaluation team recognises the issue of legibility. This need not be comprehensive but including one or two examples will help illustrate the evaluation team's approach to assessing the causal relationships.</p>
	<p>iii Comprehensive analysis and assessment of the theory of change, results chain or logical framework, and if requested in the ToR, it is retrofitted/reconstructed by the evaluators.</p>	Partially	<p>There is some analysis and assessment of the ToC in the section on Evaluation Approach (Chapter 1, p.11). However, this is not comprehensive and is not helped by the illegibility of the ToC diagrams in Annex 5 (see Q7i and ii above). There are four ToCs, based on programme outcomes in the Annex, it may be worth incorporating an overarching ToC that then links to the 'nested' theories of change. This would help with coherence. As mentioned above, although difficult to read the text, the figure includes assumptions that are identified in the outcome level ToC. Evaluators can be encouraged to link them to the assumptions in the Evaluation Matrix, to enhance clarity on the links with the evaluation questions. This can be done using numbering labels, for example. Nevertheless, it is clear that there was a process for assessing the Theory of Change, documented in Annex 5 includes a table titled 'Discussion notes on the above ToC', which notes down limitations with the ToC and corresponding action points. However, it is not clear the extent to which these actions were addressed.</p>
Question 8.	Does the report specify adequate methods for data collection, analysis, and sampling?		
	<p>i Evaluation design and set of methods are clearly described, and are relevant and robust for the evaluation's purpose, objectives and scope, including the use of AI in the evaluation process if applicable.</p>	Yes	<p>The evaluation questions were refined during inception and are aligned with the purpose and objectives of the evaluation. This is elaborated on under Q6i above.</p> <p>Data collection methods are explicitly mentioned in Section 1.3.2 (p. 12) and also in the Evaluation Matrix (Annex 1). They include primary and secondary data. Primary data includes individual interviews, group interviews, focus group discussion, and observations made during field visits. A stakeholder survey undertaken by UN Women was also incorporated as a source. The link between the Spotlight Initiative (SI) case study, commissioned by UNFPA and UN Women, is provided on p.12.</p> <p>AI was not used in the evaluation aside from translation, from Tetum to English, and also for the use of paraphrasing in some instances (p. 13).</p>
	<p>ii Data sources are all clearly described and are relevant and robust; these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR).</p>	Yes	<p>The evaluation employs a mixed-methods approach, including both qualitative and quantitative sources. The data sources are clearly described and sufficiently robust. Quantitative data is sourced mainly from secondary sources (a full list of document consulted is provided in Annex 2) and this includes desk reviews, online databases and surveys to inform financial data and key indicators at output and outcome level (p. 12). Qualitative data is collected through individual interviews, groups interviews and focus group discussions (p. 12). More on the sampling strategy is covered below under Q8iii.</p>

	iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or if not, provide reasons for this).	Partially	<p>A description of the sampling strategy is provided in Section 1.3.3 (p. 14). It includes the sampling of stakeholders for interviews as well as sampling of field sites. The stakeholder map (Annex 6), provided by the country office, was used as the sampling frame. They were then 'refined based on interventions, geographic location, and roles in CP4, ensuring representation from all participating UN agencies, ministries, donors and beneficiaries' (p.14). The use of snowball sampling technique was also mentioned but it was unclear to what extent this was used. It would be helpful for evaluators to include an explanation for how and when a snowball sampling strategy was used, what its benefits were as well as potential limitations. A full list of stakeholders met is provided in Annex 3.</p> <p>On field sites, a purposive sample was used, based on key criteria (presented on p. 14-15). The evaluation report also provides information on adaptations made to some initial selected sites due to privacy and logistical concerns (p. 15). The justification was provided.</p> <p>The selection of focus groups is not discussed in the section on sampling. However, under Evaluation Approach (section 1.3, p. 10), the evaluation report mentions that 'the selection of beneficiaries was done both [by] ET as well as CO and in some cases CSO' (p.10). It then added that due to time limitations, the evaluation team met with beneficiaries selected by the implementing partners (IPs), which the report acknowledges as leading to potential bias. In addition, adult men could not be interviewed except for male students. It is unclear why this is. A list of focus groups is also not provided in Annex 3 so it is unclear how many were interviewed, in what locations and who (e.g. disaggregated by gender, age and so on). This would be good to include to increase transparency.</p> <p>While the evaluation report provides a sampling strategy and some criteria for the selection of stakeholders and site visits, it would be helpful for all sampling to be covered in one section rather than in different parts of the report. For example, information on focus groups is provided under Evaluation approach rather than sampling. This may be a result of multiple authors but worth looking for when editing to ensure coherence.</p>
	iv	Methods allow for rigorous testing of the theory of change, results chain or logical framework (e.g. methods help to understand the causal connections, if any, between outputs and expected outcomes (3TRs).	Partially	<p>The use of contribution analysis would allow for rigorous testing of the theory of change, including causal linkages. However, its description of how contribution analysis is used and applied in the evaluation report and the design report is limited. It is briefly mentioned on page 12 and again on page 13. Evaluators are encouraged to briefly describe what it is and the steps involved in applying the approach.</p>
	v	Clear and complete description of the methods of analysis, including explainability and full disclosure of the use of AI in the evaluation process, if applicable.	Yes	<p>Data analysis is described in Section 1.3.2 (p.13). The Evaluation Matrix served as the guiding structure for data analysis (full Evaluation Matrix is included in Annex 1). The evaluation report describes the analysis process where common these and patterns were identified according to evaluation questions. It also describes the process of triangulation across different sources and methods. Evaluators are also encouraged to apply a strength of evidence rubrics to assess data quality. However, this is only a suggestion and does not affect the score.</p> <p>The evaluation report specifies that it did not use AI in the evaluation, aside for translation purposes.</p>
	vi	Clear and complete description of limitations and constraints faced by the evaluation in its data collection and analysis, including gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators (as feasible).	Yes	<p>There is a clear and comprehensive set of limitations and mitigation measures provided in Section 1.3.4 (p. 17). This is very helpful to understand the constraints and some of the limitations of the evidence gathered. Evaluators can consider presenting the limitations in a table format with a RAG rating for each risk/limitation. Evaluators can then present the corresponding mitigating steps that includes a RAG rating for residual risks or limitations. This is helpful for understanding the extent to which the mitigating steps address the limitations. However, this is a suggested good practice and does not affect the rating.</p>
Question 9.		Are ethical issues and considerations described? The evaluation should be guided by the UNEG ethical standards for evaluation. As such, the evaluation report should include:		
	i	Explicit and contextualized reference to the UNEG obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) and/or UNEG Ethical Principles.	Partially	<p>The Evaluation Report makes explicit reference to UNEG Norms and Standards for Evaluation and to UNEG Ethical guidelines (p. 9). It also mentions adherence to principles of independence, impartiality, credibility and utility. There is a sub-section in 1.3.2 that provides contextualisation (p. 13-14). However, they are mainly related to ethical considerations and safeguards rather than principles of independence, impartiality, credibility, conflicts of interest, accountability. Evaluators are encouraged to reflect on how these principles are considered in the evaluation design and process, specific to this CPE, in order to fully meet this criterion.</p>
	ii	Clear description of ethical issues and considerations (e.g. respect for dignity and diversity, fair representation, confidentiality, and avoidance of harm) that may arise in the evaluation, safeguard mechanisms for respondents (e.g. parental consent forms for adolescents, compliance with codes for vulnerable groups; WHO standards of safe data collection on GBV) and ethical considerations in the use of AI as applicable (e.g., transparency of use, explainability, privacy, data protection, accuracy, human rights). If AI is used in the evaluation, there should be transparency and disclosure on the ethical and responsible use of AI in the report.	Partially	<p>There is a description of ethical considerations provided on p. 13-14. These include steps taken to obtain consent, ensuring confidentiality, maintaining anonymity, respect and adherence to 'do no harm'. There is mention of a shared storage space with access only for the evaluation team but it does not discuss data retention or disposal. The section refers to more information in Annex 3. However, the information in Annex 3 is limited (p. 9 of Annex 3) where it outlines consent forms for interviewees and for those under 18 years'. There are also protocols for confidentiality and consent in the protocols of the data collection instruments. In addition, Annex 4 includes General guidance (i.e. protocols); Semi-structured interviews; Facility visit; Focus Group Discussions; and Consent Form. In order to fully meet this criterion, evaluators can consider expanding on steps taken in all phases of the evaluation to ensure informed consent, protection of confidentiality, data storage and management. While some description is provided, it would be helpful to expand on what it means. For instance, it is unclear how the evaluation team applied the 'do no harm policy'. Similarly, how data was managed and steps to retain and dispose data are also not discussed.</p>
Question 10.		Does the evaluation incorporate innovative practice that adds value to the evaluation process?		

	i Innovation practice is used to improve the quality of evaluation process. This could include efforts to optimize the evaluation process (e.g., use of AI or new technology for data gathering, content analysis, outcome harvesting among others), or components introduced to enhance inclusion and participation in the evaluation processes (e.g. a youth steering committee), or ways of sharing of evaluation results.	Not Rated	The ToRs do not explicitly mention innovation or innovative practices for this evaluation.
SECTION E:	EVALUATION FINDINGS (weight 25%)	90%	Comments on Rating
Question 11.	Do the findings clearly and adequately address all evaluation questions and sub-questions?		
	i Findings are presented clearly and provide sufficient levels of evidence to systematically address all the evaluation's questions	Yes	The findings are clearly presented in Chapter 4 (p. 33 onward). They are structured according to UNFPA Evaluation Handbook (2024) guidance where each finding corresponds to an evaluation question. Each question opens with a summary, presented in a box. In some instances, it includes a section introduction to help orient the reader. Following the summary box, the key findings are highlighted in bold and numbered. Each finding statement is substantiated with evidence. The findings section follows the logic of the Evaluation Matrix, making it easy to reference between the two. The findings clearly respond to the evaluation matrix and, often though not consistent, also respond to the assumptions presented in the Evaluation Matrix.
	ii Explicit use of the evaluand's theory of change, results chain, logical framework in the formulation of the findings.	Yes	There is explicit use of the Country Programme ToC and results framework in the formulation of the findings. This is particularly evident in the sub-section on effectiveness, which presents the evidence according to the four Country Programme outcomes and outputs as well as UNFPA Strategic Plan outcomes, as depicted on Table 5 (p. 28-29). What would be helpful is a better understanding of how contribution analysis was applied for the broader and nested ToCs. As mentioned above, the narrative description of the nested ToCs are limited and the diagrams are unclear. While the evaluation report discusses contribution, it is challenging to understand the causal pathways with these limitations. One suggestion is for evaluators to clearly set out at the onset the areas of the ToC that will be assessed by the evaluation and then to present and analysis output and outcome level data in a clearer and more systematic way. However, this suggestion does not affect the score as the evaluation makes explicit reference to the ToC.
Question 12.	Are evaluation findings derived from credible data sources as well as a rigorous data analysis?		
	i Evaluation uses credible forms of qualitative and quantitative data. It presents both output and outcome-level data as relevant to the evaluation framework. Triangulation is evident using multiple data sources.	Yes	The evaluation uses credible forms of qualitative and quantitative data, which is presented in the Evaluation Matrix. In addition, it includes references for data presented in the footnotes of the findings section. This includes both qualitative and quantitative data. The list of documents consulted is provided in Annex 2 and the list of people met is presented in Annex 3. Output and outcome level data is presented. It is evident from the sources provided in the footnotes that triangulation has taken place e.g. evidence that participation of LGBTQI and those with disability had not been 'adequately reported or visible in the programme implementation' is supported by a footnote that shows triangulation from workshop report, CO staff interview and interview with community association (p. 36). However, this is not consistent throughout the findings section. To strengthen the quality of evidence, evaluators can consider the use of 'quality/strength of evidence' rubric and rate each finding accordingly. This would increase transparency and consistency regarding the quality and/or strength of evidence (e.g. level of triangulation) behind each finding. Please note that it does not mean that only findings that cross a certain threshold would be reported. While this is a suggestion to strengthen transparency when reporting on findings, it does not affect the score.
	ii Findings are clearly supported by the evidence presented, both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Yes	The findings are supported by the evidence collected. They include both positive and negative information, and some are more nuanced. For example, Finding 5 presents a positive picture of progress being made in capacity building but also highlights the gaps that remain in integration of SRH services (p. 40). Overall, findings are based on assessment indicators and/or criteria, which are outlined in the Evaluation Matrix.
	iii Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings analyse the logical chain (progression -or not- from outputs to high level results).	Partially	Overall, the Evaluation Report provides an analysis of what worked or did not work across all the evaluation questions. This is particularly evident in the section under 'effectiveness'. What is less evident is a discussion of the causal pathways, as this is a theory-based evaluation. As mentioned in earlier section, it would be helpful to set up front what aspects of the ToC, the evaluation would examine through the evaluation and to describe it in a systematic way in the findings section. Beyond merely highlighting what works or does not, evaluators should dig deeper to understand why. For example, a thorough exploration of the ToC will enable the team to go further beneath the surface.
Question 13.	Does the evaluation assess and use the intervention's Results Based Management elements?		

	i Assessment of the adequacy of the intervention's planning, monitoring, and reporting system (including completeness and appropriateness of results/performance framework - including vertical and horizontal logic, M&E tools and their usage) to support decision-making.	Not Rated	The assessment of the Country Programme's planning, monitoring and reporting system is not explicitly mentioned in the ToRs. As a result, this criterion is 'not rated.' However, EQ2 under relevance, asks about the extent to which needs of targeted populations are taken into account in the planning, implementation and monitoring of UNFPA supported interventions. This is discussed under Finding 4 (p. 36-38). There are other parts of the findings section that discuss data gaps, measurement and monitoring issues such as Findings 10, 11, 12, 15 and more.
SECTION F:	EVALUATION CONCLUSIONS (weight 10%)	75%	Comments on Rating
Question 14.	Do the conclusions clearly present an unbiased overall assessment of the evaluand?		
	i Conclusions are clearly formulated and present unbiased summative statements that respond to the evaluation questions.	Yes	The conclusions are clearly formulated and present unbiased summative statements that synthesise the key findings. They are presented as strategic and programmatic conclusions. Each conclusion is concise and is supported by a more detailed description that synthesises the relevant findings. Sign-posting to relevant findings is provided at the end of each conclusion using numbering. However, it would be helpful if the relevant EQs and their numbers are added.
	ii Conclusions are well substantiated and derived from findings and add deeper insight and analysis beyond the findings.	Partially	The conclusions are well written and nearly all conclusion statements are followed by a description, with the exception of conclusion 5 that appears to provide a summary of the conclusion statement only. However, there are instances where it is difficult to see the links between the conclusion statements and the sign-posted findings. For example, conclusion 6 discusses issues relating to indicators and causal linkages between outputs and outcomes. It also signposts to findings 1, 2, and 3. However, these findings do not directly discuss issues of indicator quality of the results framework. Similarly, conclusion 3 on access to ASRH services is discussed in several areas across the findings section, although there is limited discussion in the findings sign-posted (Finding 18 may be more relevant than Finding 5, 9, 10). There are other issues with sign-posting, such as finding 14 stating that UNFPA support drive census data utilization, while the conclusion notes that this is a critical area for improvement. Evaluators are encouraged to check the sign-posts to enhance transparency, as well as improve navigation across a lengthy report.
Question 15.	Are lessons learned identified? [N/A if lessons are not referenced or requested in ToR]		
	i Lessons learned are derived from the findings and are well substantiated with practical, illustrative examples.	Not Rated	Lessons learned are not identified in the report. However, they were not stated as a requirement in the ToR.
	ii Lessons learned are clearly presented and provide actionable insights on the positive aspects of the evaluand as well as any areas of improvement.	Not Rated	Lessons learned are not identified in the report. However, they were not stated as a requirement in the ToR.
SECTION G:	EVALUATION RECOMMENDATIONS (weight 15%)	88%	Comments on Rating
Question 16.	Are recommendations well-grounded and articulated?		
	i Recommendations are clearly formulated and logically derived from the findings and/or conclusions.	Yes	The recommendations are clearly formulated with a heading followed by an articulation of the full recommendations. This is then followed by operational steps/suggested action. The recommendations are logically derived from the conclusions. Each recommendation includes a clear reference to the corresponding conclusion.
	ii Recommendations are useful and actionable for primary intended users. Specific guidance is provided for its implementation (e.g. actions, deadlines, responsible actors), as appropriate.	Partially	The recommendations are useful and actionable for primary intended users. They are categorised into strategic and programmatic recommendations. For each recommendation, there is a list of operational implications that outlines specific steps that those identified as holding responsibility for the recommendation can take. However, the report does not identify specific teams or units within the CO who are responsible for implementing recommendations. These need to be included to fully meet the criterion.
	iii Process for developing the recommendations is described, and includes the Involvement of key stakeholders (e.g. evaluation reference group members), including those who will be affected by the recommendations.	Yes	The start of the chapter on recommendations (Chapter 6, p. 85) describes the process for developing the recommendations. The report says that the recommendation were discussed and validated by the Evaluation Reference Group (ERG), Country Office programme staff and UNFPA Asia and Pacific Regional Office (APRO).
	iv Recommendations are clearly articulated and prioritized based on their importance, urgency, and potential impact.	Yes	Each recommendation has a priority rating as high, medium or low. While there is no ranking according to level of importance, the most pressing recommendation is placed at the top.
SECTION H:	REPORT STRUCTURE AND PRESENTATION (weight 5%)	75%	Comments on Rating
Question 17.	Does the evaluation report include all required information?		
	i Opening pages include: Name of evaluation and/title of evaluation, timeframe of the evaluation, date of report, location of evaluand, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents (including, as relevant, tables, graphs, figures, annexes);- list of acronyms/abbreviations.	Yes	The opening pages include the name of the evaluation; timeframe of the evaluation; date of the report; location of the country programme; names and titles of the evaluators; list of the evaluation reference group members; acknowledgements; table of contents (including tables, figures and charts); list of acronyms and abbreviations; and the key facts table. What is missing is a table of contents for the Annex - this is also not provided in the Annex. This would be helpful in navigating the Annex, particularly as it is 253 pages in length and the page numbers are not in order - they appear to be a consolidating of different documents into one. This has required the reviewer to create a table of content of the Annex in order to review related content between both documents.

	ii Annexes include, if not in body of report: terms of reference, evaluation matrix, list of respondents, results chain/ToC/logical framework, list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology (e.g. inception report), case study reports.	Yes	The Annex includes the Terms of Reference; Evaluation Matrix; Stakeholder Map; List of participants; List of documents consulted; Facilities visited; ToC; and data collection instruments. It also includes the Spotlight Initiative Timor Leste – Case Study Report, What it does not include are the informants who participated in focus group discussions.
Question 18.	Is the report logically structured and of reasonable length?		
	i The report has a logical structure that is easy to identify and navigate (for instance, with numbered sections, clear titles, well formatted).	Yes	The evaluation follows the standard structure for evaluation reports as outlined in the UNFPA Handbook. The main report is easy to identify with clear titles, sub-titles and numbered chapters and sections. It is the Annex that was difficult to navigate and this feedback is provided above under Q17i. However, this does not affect the scoring for meeting this criterion.
	ii Structure and length accords to UNFPA guidelines for evaluation reports; it does not exceed number of pages that may be specified in ToR. <i>Note: Maximum pages for the main report, excluding executive summary and annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations and 50 for other types of evaluations)</i>	Partially	The Evaluation Report is 85 pages in length (excluding the Executive Summary and Annexes), exceeding the recommended page length for country programme evaluations.
Question 19.	Is the report well presented?		
	i Report is easy to understand (written in an accessible way for the intended audience) and generally free from grammar, spelling and punctuation errors.	Partially	Overall, the report is easy to understand. However, there are sections that make it difficult to follow. For example, the description of the ToC underlying the Country Programme level and the outcome-level ToCs is unclear. It is also not clear which ones were refined by the evaluators (this point is covered above). This is an important section as it frames what the evaluation sets out to do. Otherwise, it is generally free from grammar, spelling and punctuation errors.
	ii Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labelled, and referenced in text.	Partially	The report is densely written, and evaluators are encouraged to consider using visual aids, particularly in the findings section, to support the communication of key information. Where visual aids are included, they are clearly presented, appropriately labelled, and referenced within the text.
SECTION I:	CROSS-CUTTING ISSUES (weight 10%)	73%	Comments on Rating
Question 20.	Are cross cutting issues - in particular, human rights-based approach, gender equality, disability inclusion, LNOB - integrated in the core elements of the evaluation (e.g. evaluation design, methodology, findings, conclusions and recommendations)?		
	i Evaluation's data collection methods designed to capture the voices/perspectives of a wide range of stakeholders including right holders, marginalized and vulnerable persons, young people, people with disabilities, migrants or refugee populations, indigenous communities, and other persons that are often left behind.	Partially	The evaluation data collection methods are intended to capture the voices and perspectives of a wide range of stakeholders including rights holders. However, the sampling for rights holders beneficiaries is not clear. As discussed above, there were limitations to access but the evaluation report states that a 'convenient sample of beneficiaries participated in informal group discussions and focus group discussions to provide insights on service quality, accessibility and utility' (p. 12). Evaluators are encouraged to describe who was included alongside the limitations. Rights holders are not included in the stakeholder sampling (Annex 3). The list of person met in Annex 3 does not include FGDs and so it is unclear who were interviewed and whether the targets were achieved.
	ii Evaluation questions address cross cutting issues, such as human rights-based approach, gender equality, disability inclusion, LNOB, social and environmental standards as appropriate.	Yes	At least half of the evaluation questions included explicit references to LNOB considerations, gender equality and human rights issues. Some of the remaining question, while not explicit, touch on issues such as access. For this reason, this criterion is rated as being met.
	iii Data is disaggregated by population groups (e.g. persons with disability, age, gender, etc.) where there are implications related to UNFPA's portfolio/interventions for these population groups; differential results are assessed (distribution of results across different groups).	Partially	The evaluation disaggregates by age, gender, disability, and other characteristics, where data is available. The issue of data availability is raised in the recommendations, particularly operational steps. The list of respondents in the Annex (3) are disaggregated by gender. However, evaluators are also encouraged to include and disaggregate the focus group discussions to fully meet this criterion.
	iv Intersectional lens is applied in the data analysis, looking at various and multiple forms of exclusion and discrimination (and how they overlap with each other) and how this may impact the performance or results of the evaluand.	Partially	There is some analysis of intersectionality in the evaluation report but it is not systematic. For example, looking at young girls (age and gender) and access depending on whether they are based in urban or rural areas. The evaluation report also discusses the limited analysis on intersectionality within the programme, which it attributes to data utilisation as well as data availability (p. 82).
	v Findings, conclusions and recommendations, address cross-cutting issues such as equality and vulnerability, disability inclusion, leave no-one behind, social and environmental as relevant.	Yes	Cross-cutting issues including equality and vulnerability, disability inclusion, and leave no one behind are discussed throughout the findings section, conclusion and recommendations.
	vi Inclusion of young people in the evaluation team and/or Reference Group [N/A if not requested in ToR]	Not Rated	While the profile of the evaluation team (Section 12) of the ToRs do not specify the inclusion of young people in the evaluation team, the evaluation team has a young and emerging evaluator (see CPE Evaluation Team in opening pages).
Question 21.	Does the evaluation meet UN SWAP evaluation performance indicators?	7	Comments on Rating
	i GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Satisfactorily integrated	While the evaluation objectives do not explicitly mention GEEW considerations, it is made explicit in the thematic scope. The evaluation also assesses that the level of information collected and analysed during the implementation period is related to GEEW. There is no stand-alone criterion on gender and/or human rights. However, it is effectively mainstreamed into other evaluation questions as half of the evaluation questions explicitly reference gender specific considerations.

ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated	The evaluation methodology employs a mixed-methods approach, appropriate to evaluating human rights and gender considerations. A diverse range of data sources and processes were employed. With the exception of focus group discussions, all stakeholders interviewed are disaggregated by gender. The sampling frame addresses the diversity of stakeholders affected but the description could be improved. As mentioned in comments above, the sampling map does not include rights holders. Furthermore, ethical standards were considered but could be further elaborated on to better understand how they were operationalised. Protocols for confidentiality is provided in the data collection guides (Annex 4).
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Fully integrated	The evaluation findings, conclusions and recommendations reflect a gender analysis. This is seen in the findings, which includes an intersectional analysis. This is also evident in the conclusion and recommendations. The data analysis transparently triangulates the voices of different groups and disaggregates, based on available information. The evaluation report also provides specific recommendations addressing GEEW issues and priorities for action to improve the Country Programme in the next phase.

List of SDGs

1. No Poverty
2. Zero Hunger
3. Good Health and Well-being
4. Quality Education
5. Gender Equality
6. Clean Water and Sanitation
7. Affordable and Clean Energy
8. Decent Work and Economic Growth
9. Industry, Innovation and Infrastructure
10. Reduced Inequality
11. Sustainable Cities and Communities
12. Responsible Consumption and Production
13. Climate Action
14. Life Below Water
15. Life on Land
16. Peace, Justice and Strong Institutions
17. Partnerships for the Goals

Three transformative results

1. Ending unmet need for family planning
2. Ending preventable maternal deaths
3. Ending gender-based violence and harmful practices

Six outputs

1. Policy and accountability
2. Quality of care and services
3. Gender and social norms
4. Population change and data
5. Humanitarian action
6. Adolescents and youth

Six accelerators

1. Human rights-based and gender-transformative approaches
2. Innovation and digitalization
3. Partnerships, South-South and triangular cooperation, and financing
4. Data and evidence
5. Leaving no one behind and reaching the furthest behind first
6. Resilience and adaptation, and complementarity among development, humanitarian and peace-responsive efforts