

UNFPA Evaluation Quality Assessment Grid

Version: August 2025 annotated

REPORT RATING SUMMARY		
Overall Rating	71%	Satisfactory
●●●●	Excellent	5
●●●●	Highly Satisfactory	4
●●●●	Satisfactory	3
●●●●	Fair	2
●●●●	Unsatisfactory	1

REPORT DETAILS	
Title of the evaluation report	Country Programme Evaluation of the Government of Zimbabwe /UNFPA 8th Country Programme 2022-2026
Region	ESARO
Country	Zimbabwe
Year of report	2025
Business Unit/programme country (managing evaluation)	Zimbabwe Country Office
Date of assessment review (dd/mm/yyyy)	15-Dec-25
Name of assessment review firm	IOD PARC
CLASSIFICATION OF EVALUATION REPORT	
Primary SDG(s) covered (list provided below)	3, 5, 13, 16, 17
UNFPA Strategic Plan areas covered (lists provided below)	
Three transformative results	Yes
Six outputs	no
Six accelerators	no
Organizational effectiveness and efficiency	no
Humanitarian evaluation	no
Evaluation evaluated (e.g. country programme/intervention/policy/thematic area)	Country Programme
Evaluation type (e.g. formative, summative, developmental)	Summative & Formative
Geographic scope (e.g. global, regional, national)	National

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.

The 8th Country Programme Evaluation (CPE) of UNFPA Zimbabwe 2022-2026 is rated satisfactory, as the evaluation report demonstrates a high level of technical rigour and methodological alignment with UNFPA standards, with robust integration of cross-cutting issues throughout. However, the report is compromised by issues of lack of conciseness and excessive detail in places, leading to poor readability in some sections. A key flaw is in how findings/analysis are presented, with the many numbered "findings" in actuality being simply elements of the analysis rather than summative assessments that are then supported by narrative. This structural issue is then reflected in many such findings being somewhat cursory, and leads to an excessive number of conclusions. Below is a summary of the evaluation's strengths and weaknesses.

- Strengths**
- Comprehensive executive summary which includes all relevant components of the report.
 - A robust stakeholder analysis that includes a comprehensive mapping (Annex 5) and an analysis of the interests, power, and potential impacts of various groups (Table 3).
 - The purpose, objectives and scope of the evaluation are clearly articulated and aligned with the inception report.
 - The methodology is explicitly anchored in a theory-based contribution analysis approach that is appropriate for the Country Programme's structure.
 - Recommendations are clearly formulated, prioritised, and logically derived from the conclusions with specific operational implications and responsible actors.
 - The methodology explicitly targeted vulnerable populations, including youth and people with disabilities, in its approach (included in evaluation scope and data collection).
- Weaknesses**
- While the executive summary is strong, at seven pages, it is too long, with somewhat excessive enumeration of findings.
 - The background and programme description sections are unnecessarily long and repetitive.
 - The report identifies humanitarian coverage and connection between humanitarian and development work as key objectives, yet fails to include dedicated evaluation criteria or mainstream these considerations across the Evaluation Questions.
 - While the evaluation matrix is systematic, its analytical value is limited because the assumptions simply rephrase the evaluation questions. Furthermore, for some evaluation questions, only one assumption is articulated. As the evaluation questions were not broken down into meaningful and testable assumptions, the Evaluation Matrix failed to serve as a framework for rigorous analysis, which negatively affected the quality of the findings.
 - The sampling strategy lacks clear justification for selecting non-programme provinces. In addition, inconsistent details regarding selection criteria and targeted sample sizes across all stakeholder groups undermine the overall rigour of the sampling.
 - There is a section on key limitations and mitigation measures but it omits essential limitations such as systemic data gaps, indicator design flaws, and disaggregation constraints detailed elsewhere in the report.
 - The ethical considerations section relies on generic commitments to standards rather than providing contextualized details on specific safeguarding mechanisms for vulnerable groups.
 - Although the evaluators reconstructed the Theory of Change, successfully integrating previously separate thematic pathways, the diagram is overly complex and spans three pages, which prevents the reader from grasping the logical pathways easily.
 - The Theory of Change includes some assumptions about the context (for example, political stability and sustained donor support) but lacks assumptions about causal mechanisms at play and the capacities, motivations and opportunities of stakeholders, hindering the assessment of how results were achieved.
 - The report is more descriptive than analytical. The findings section frequently resorts to simple activity listing instead of investigating the underlying causal drivers for the achievement or non-achievement of results.
 - An excessive number (110) of findings which function as subjective analysis paragraphs rather than objective assessments supported by clear, rigorous evidence.
 - The 17 conclusions are too numerous and often rephrase the findings rather than presenting the evaluation team's judgement of the Country Programme's performance.
 - Several critical visual aids, including the SDG progress chart and various figures, have poor resolution that makes the details indiscernible.
 - The report lacks deep intersectional analysis and consistent data disaggregation by age, disability, and geography, although it acknowledges that secondary data is insufficiently disaggregated. In addition, unintended consequences of programming are not explored.

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.

- The following are recommendations for improvement:
- The Executive Summary could easily have been shorter (seven pages is a maximum, not a target) to enhance readability by removing excessive detail, repetitive enumeration of findings.
 - To reduce page length and enhance readability, the background and programme description (Chapters 2 and 3) could be truncated by moving detailed information (such as the description of the 7th CP, which is only peripherally relevant), detailed narratives of the 8th CP, and large tables (e.g., Table 11), to an annex.
 - An evaluation of a CP with a humanitarian component should use coverage and connectedness criteria to assess humanitarian preparedness or response or, at minimum, mainstream humanitarian considerations across evaluation questions to ensure comprehensive assessment.
 - The Evaluation Matrix should have further defined the aspects covered by the evaluation questions and translated them into testable assumptions rather than rephrasing the questions as statements. In addition, a robust Evaluation Matrix should include several assumptions per question to ensure dedicated data collection and analysis while ensuring coverage of all the evaluation questions' aspects.
 - The report should explicitly justify the selection of non-programme provinces, even if not intended to be used as a counterfactual, while consistently documenting selection criteria, sample sizes and measures taken to ensure representativeness and mitigate bias across all stakeholder groups, especially if "convenience sampling" is used.
 - The limitations section could have been more comprehensive, to include systemic data constraints identified elsewhere in the report, such as the absence of data for key indicators.
 - More specific, contextualized details on the ethical safeguards and mechanisms used were required, particularly for vulnerable groups, with more comprehensive assurances of safety, confidentiality, and explicit referral plans for sensitive disclosures (e.g., from GBV survivors), rather than relying on generic commitments to UNEG principles.
 - The visual representation of the reconstructed Theory of Change should have been condensed to a single page to enable readers to grasp the logic easily, and avoid categorizing non-specific underlying assumptions (like political stability or donor support) under individual programmatic outcomes.
 - The theory of change should include critical assumptions about contextual factors, causal mechanisms and stakeholder characteristics and clearly label them as such. The assumptions should be clearly linked to those articulated in the Evaluation Matrix.
 - There could have been less simple listing of activities or outputs and more rigorous analysis of the underlying drivers (causal factors) of success or non-achievement for programme results.
 - Presenting numbered analysis paragraphs as "findings" is not standard practice and in this case has led to 110 findings in total, which is excessive. The report could have condensed the analysis and formulate a much more limited number of topline findings that were explicitly and systematically linked to clear, objective evidence.
 - The number of conclusions (17 in this report) should have been fewer – they should have been reformulated into a limited set (e.g., 6–7) of higher-level, strategic judgments, rather than summative findings or operational observations, to better align with accepted good practice for evaluation reports.
 - The resolution of all figures and graphics should have been improved, and a final check on table of contents, in-text references to graphics etc. should have been undertaken.
 - The analysis should move beyond binary gender comparisons and consider intersectional factors of discrimination (e.g. how SRH outcomes of adolescent girls with disabilities in rural areas compare to other adolescent girls), as well as explore unintended effects of programming, especially for vulnerable and marginalized populations.

SECTION RATINGS			
SECTION A:	EXECUTIVE SUMMARY (weight 5%)	83%	Comments on Rating
Question 1.	Can the executive summary inform decision-making?		
	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 at the maximum permitted length of seven pages. It is generally clear, though the writing could be smoother in places and some formulations could be refined for clarity.
	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	Yes	The summary includes all six specified components of the evaluation report, presenting structured sections for the overview of context and intervention, evaluation purpose and intended users, scope and methodology, findings, conclusions, and recommendations. However, the objectives as articulated in the summary are not especially well-aligned with the actual objectives in the main report text. There is adequate detail provided under each of the elements, for example the scope and methodology detail the temporal and geographic coverage and outline the methodological approach and tools. The structure of the summary reflects that of the overall report, warranting a 'yes' rating.

	iii Includes all significant information in a concise yet clear manner to understand the evaluand (e.g., theme, intervention, programme, project, or strategy undergoing evaluation) and the evaluation.	Partially	The summary includes all significant information necessary to understand the UNFPA 8th CP, the evaluation process itself and the findings, conclusions and recommendations. A good practice is the inclusion of data on people with disabilities. However, at seven pages, it lacks conciseness and thus the readability of the summary is poor. It is compromised by quite detailed enumeration of findings and repetition, with several areas where language and presentation could be more compact without sacrificing meaning. For instance, the findings detailing successes and shortfalls across multiple sub-points (e.g., listing multiple policies the 8th CP was aligned with. The summary of the methodology, while robust, could well have been condensed without losing the key information on methodological rigor. At the same time, some important aspects, notably how data was analyzed, is not included. The writing style (short, staccato sentences), the spelling out of every acronym (e.g. OECD-DAC is fully spelled-out) and enumerating things like the specific SDGs that pertain to the CP could have been avoided. The choice of font and poor formatting (i.e. missing paragraph breaks) does not help in this regard.
SECTION B:	BACKGROUND (weight 5%)	90%	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., theme, intervention, programme, project or strategy undergoing evaluation), including: geographic coverage, implementation period, main partners, cost/budget, and implementation status.	Yes	Chapter 3 provides a detailed, and reasonably clear (but very long) description of the UNFPA programming across thematic areas, geographic scope, implementation period, main partners, financial details, and implementation status. The section describes the development process for the CP (in collaboration with a wide range of stakeholders), and also the full list of partners and the level of engagement with UNFPA (in financial terms) which is positive and useful. A minor missing element in this description is the apparent two-year gap in strategic coverage between the 7th CP (2016-2020) and the 8th CP (2022-2026). While both CPs are extensively described, the reason for this gap - and any provisional strategy such as a CP extension - is not given. The only significant issue is that the section is quite long, and the extensive narrative around the description of the 8th CP plus the very detailed Theory of Change diagram means that there is repetition. The section could have been summarised, with, for example, the description of the 7th CP (one page) being condensed, as well as the very detailed (four pages) narrative description of the 8th CP. Some of this narrative (and numerical data, such as the full-page table 11) could have been moved to an annex. However, this does not impact the rating.
	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 thematic area, intervention, programme, project or strategy).	Partially	Section 2 provides a reasonably clear and comprehensive description of Zimbabwe's context, including its economic, social, and political environment, and explicitly details key drivers and challenges affecting UNFPA programming across its thematic areas. The section is supported with good data (benefiting from relatively recent DHS and MICS reports) on UNFPA's thematic areas, although the issue/prevalence of disability is not noted. Further, the data presented in chapter 2 of the report is not consistently disaggregated, particularly by sex and geographic location. In addition, cross-cutting vulnerability factors are not clearly reflected in the context analysis. While populations furthest behind, including persons with disabilities, are mentioned, this is not systematically supported by disaggregated data, warranting the partial rating. The section also integrates cross-cutting issues, detailing government commitments to upholding international human rights standards and the LNOB principle. Key challenges that affect UNFPA programming are well-covered, specifically SRHR, GBV, adolescents and youth and P&D. There is some repetition in this section (as with others), which could have been eliminated to improve readability and length of the report. For example, data on population structure is provided in three places: sub-sections 2.1.1, 2.1.3 and 2.1.5. One would have sufficed. In addition, the description of ODA flows into Zimbabwe, while important, is quite incomplete - the graphic (figure 6) presents data from 2013-2022 (2022 is the start of the evaluation period), with the narrative noting 2024 data (only the first nine months), but nothing for 2023 or 2025. Given that the report was written in late 2025, more up-to-date data on these metrics should have been available (or the reason for its absence could have been provided).
	iii Linkages between the evaluand and ICPD-related SDGs are outlined. This may include reference to ICPD benchmarks and relevant SDG targets and indicators.	Yes	Section 2 covers the SDGs in Zimbabwe, with the evaluators providing a description of SDG achievement in the general country context sub-section (2.1.1) and in the 'key facts' table. The sub-section refers to the overall status of the SDGs, and particularly notes SDGs 3, 5 and 13 (per the guidance) as well as some additional relevant ones (16 and 17). The evaluators also include a table visually explaining progress on the SDGs. This is minimally adequate for a 'yes' rating. However there are several issues that affect the general quality of this aspect: - The sub-section is somewhat out of place in the 'general country context' section overall - the text speaks firstly to the UNFPA transformative results, rather than the country context, and links them to the SDGs (rather than discussing the country context). - There is little reference to other than the headline goals themselves - no reference to targets/indicators etc. There is, however, discussion of key related indicators later in the text (e.g. SRHR indicators under 2.1.2), so this is mitigated somewhat. - The accompanying graphic (figure 4) is of very poor resolution, and cannot be read.
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 (e.g., stakeholder map).	Yes	Although the Evaluation Handbook does not mandate the inclusion of the stakeholder map in the final report, the report does include such a map in Annex 5. The annex is quite detailed and breaks the stakeholders down across the UNFPA strategic plan outcomes and outputs, and categorises them according to typology (implementing partners, other partners, donors and rights holders. However, rights holders are only identified for the GBV output in Annex 5. They should have been identified for the others, especially those involving service delivery such as SRH for women and adolescents and youth. Further, the main text presents lists of stakeholders in the scope sub-section (1.2) in both narrative and table format (table 3). While the subsection is not quite the appropriate place for this information, the provision of the requisite detail warrants the 'yes' criterion.
	ii Stakeholders are analysed to understand their interests and needs, power and influence, and potential impact on the evaluand.	Yes	Table 3 in section 1.2 provides a succinct and useful analysis of the interests/needs, power/influence and potential impact (on UNFPA programming) of the various stakeholder groups. Along with the breakdown of partners per outcome/output in Annex 5, this provides ample detail to enable the reader to understand the dimensions of UNFPA Zimbabwe's engagement with these external stakeholders. While this is functionally adequate for the purposes of this evaluation quality assessment, it is also a concise and useful summary for readers and could (although not explicitly noted) well have informed the evaluation design and data collection.
SECTION C:	EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)	100%	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	The purpose of the evaluation is clearly described in section 1.0. It is fully in line (and essentially repeats) with the description of the purpose in the ToR, which in turn is aligned with the parameters of CPEs as set out in the Evaluation Handbook. There is no deviation, and it clearly explains how the CPE will feed into the development of the next UNFPA Zimbabwe programme cycle.
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 inception report (if applicable).	Yes	The objectives of the evaluation are presented immediately after the purpose, in section 1.0 (possibly this should have been labelled section 1.1 - the subsection numbering goes from 1.0 to 1.2). The objectives are clear, and are the same as those set out in the evaluation inception report.
	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	The scope is covered under section 1.2. While it does include all of the required detail (on thematic, temporal and geographic scope), there are some issues. Firstly, the section also includes (under the thematic scope) the list of evaluation questions and summary assumptions. This should have been included in the subsequent section (evaluation approach - section 1.3). It also includes summary details of the country programme itself (perhaps better in the introduction), details of the UNFPA stakeholders (possibly better situated under the programme description section), and an outline of the evaluation workplan (section 1.2.5). None of these elements are specific to the scope. Secondly, under the geographic scope, the evaluators note the inclusion of the two "non-focus" provinces of Bulawayo and Harare in the evaluation - the rationale for including these in the scope is not explained. This is partially explained later in the report, i.e. the findings (e.g. finding 3) note that UNFPA did, in actuality, conduct activities in at least one of these locations (Harare - response to a cholera outbreak).
SECTION D:	EVALUATION DESIGN AND METHODOLOGY (weight 20%)	69%	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?		
	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.	Partially	The evaluation questions (unchanged from the inception report) are organised according to the five selected OECD evaluation criteria, in accordance with UNFPA guidance. The specific questions are well aligned with these criteria, with no issues of note. The questions also include important elements related to gender, youth, disability and LNOB. The questions in table 2 (presented incorrectly in the scope section - as noted above), are accompanied by a list of assumptions for verification under each, in the same table. However, the objectives note the following: "To provide an assessment of the geographic and demographic coverage of UNFPA humanitarian assistance and the ability of UNFPA to connect immediate, life-saving support with long-term development objectives." Yet, there are no humanitarian-specific (connectedness and coverage) evaluation criteria, and humanitarian considerations are also not mainstreamed across the EQs.
	ii Evaluation matrix clearly presents the evaluation criteria used as well as the corresponding evaluation questions, assumptions, indicators/ lines of inquiry, and methods for data collection, including relevant data sources.	Partially	The report presents a very detailed (54 pages) Evaluation Matrix in Annex 1 of the report. This well meets the needs of the UNFPA guidance in this regard, in that it systematically addresses each of the EQs (by criterion), the assumptions, and assigns illustrative indicators, sources of information and designates the data collection tools that were to be used to collect this information. This is a systematic and logical approach to the evaluation, with no gaps identified. However, as noted in the main report, (table 2), the assumptions as stated are of little analytical value - they essentially restate the evaluation questions in parsed sentences. For EQs 2-4 and 7, the EQs are rephrased into a single statement, so no parsing of elements of the EQ. There is no apparent breakdown of the components of the questions into underlying elements that can be tested. In the context of the full evaluation matrix (presented in the annexes), this makes more sense, as the individual question components are accompanied by a more comprehensive list of indicators and proposed evidence), but for this table, the evaluators would have been better served by simply listing the EQs organised by criteria. The matrix also includes summaries of the data that was collected to address each of the EQs/assumptions/indicators, which provides a resource for interested readers to cross-reference findings with.
Question 7.	Is the theory of change, results chain, logical framework, or equivalent framework well-articulated?		

	Clear description of the intervention's intended results, or of the relevant parts of the results chain-being tested by the evaluation, through a ToC, results chain or logical framework. This should be presented in both a visual (graphic or tabular) and accompanying narrative format.	Yes	The report does include a clear description of the UNFPA Zimbabwe intended results through a Theory of Change (ToC) framework (reconstructed, based on the existing - more complex - ToC/Results Framework of the CP), presented in both visual and narrative formats. The evaluation approach is explicitly anchored in a theory-based approach and the initial ToC analysis confirmed that the intervention logic, encompassing inputs, activities, outputs, and outcomes, was clearly articulated. There is a visual representation of the reconstructed ToC/results chain in the findings section (Figure 7, pg. 31), which visually unpacks the pathways of change, linking the inputs and accelerators (such as innovation, partnerships, and data) to the core modes of engagement, the CP outputs, which, in turn, contribute to the high-level UNFPA SP outcomes. While this is a good effort, and comprehensive, it is somewhat excessively detailed - the ToC spans three pages, with considerable detail given to describing the details of the modes of engagement of UNFPA. The text could have been considerably shortened to seek to present the ToC on one page - where the entirety of the logic can be grasped relatively easily by the reader. A detailed, accompanying narrative is integrated throughout Section 3.2.2. The narrative also breaks down the intervention logic for all four thematic outputs of the CP.
	ii Critical assumptions underlying the ToC, results chain or logical framework are clearly identified, including assumptions about causal relationships, contextual factors and stakeholder characteristics.	Partially	The report clearly identifies a wide range of assumptions (both critical and less so - as with the rest of the ToC, the evaluators are exhaustive rather than summative) underlying the ToC in both the narrative and in the diagrammatic ToC (Figure 7). Assumptions are clearly categorized across the four programmatic outcome areas, which is useful, although the logic in categorising them as such is not fully clear. For example, "political stability" is noted under SRH programming, whereas this is surely an underlying assumption for the entirety of the UNFPA country programme. Similarly, under the PD outcome, "Donor...technical and financial support to UNFPA and partners." is noted as a key prerequisite for effective programming - but this is an assumption for ALL programming. Furthermore, the ToC diagram and narrative only partially articulate assumptions linked to the change pathways. While contextual factors such as political stability and donor support are broadly referenced, assumptions related to causal mechanisms and stakeholder characteristics (capacities, incentives and agency of duty bearers and rights holders) are not clearly specified. This weakens the ToC and constrains the evaluation's ability to rigorously assess how and under what conditions UNFPA's interventions contributed to observed results. As noted below, the comprehensive and exhaustive nature of the reconstructed ToC makes it challenging for the reader to quickly or intuitively grasp the logic that is presented therein, undermining the utility of the ToC to some extent.
	iii Comprehensive assessment of the logic, internal coherence and plausibility of the ToC, results chain or logical framework, and if required, it is refined or revised by the evaluators.	Partially	There is initial discussion of the ToC structure and approach in section 1.3.1 and this is further expanded in section 3.2.2 of the findings, which provides a more granular analysis of the ToC. The overall structure of the reconstructed ToC is robust, and builds on identified shortcomings with the original ToC. Specifically, the original framework consisted of three separate thematic ToCs, with the AYP output initially integrated into the broader SRH component. The evaluation integrates this AYP component, addresses gaps in the three original thematic ToCs, and includes risks, assumptions, and explicit modes of engagement. The analysis also identifies weaknesses in the CP ToC around the absence of mitigation planning for the identified risks. However, there is a fundamental flaw in the results logic. First of all, the Country Programme outputs directly contribute to the UNSDCF outcomes, which should be below the UNFPA Strategic Plan 2022-2025 outputs and outcomes. Second, the UNFPA Strategic Plan outputs are obsolete as COs are expected to translate those global outputs into contextualized local outputs. Finally, the impact should be the overarching goal of universal access to SRH from the UNFPA Strategic Plan 2022-2025, rather than the national goal. This is logically at a lower level than the global goal. In addition, the description of the interventions in the diagram is too long, which distracts the focus on results. This is particularly concerning as the report notes that the evaluation team assessed the ToC and updated it. It is not clear if the evaluators introduced this flow or whether it was part of the original ToC. If the latter is the case, the evaluators should have spotted this and rectified the ToC. Further, the evaluation did not explicitly assess the validity of the expected pathways of change. The causal linkages were not critically unpacked (see comments above about missing assumptions). The ToC is used more as a background reference than as an analytical framework.
Question 8. Does the report specify adequate methods for data collection, analysis, and sampling?			
	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.	Yes	Section 1.3 clearly lays out the overall approach of the evaluation: theory based and utilising a contribution analysis-based approach. These are well aligned with the overall structure of the UNFPA CP (which uses its own ToC) and UNFPA evaluation practice. The specifics of the methods are also noted after the (extensive) preamble - a mixed-methods approach was adopted (also appropriate to this type of evaluation), with a summary (followed by progressively greater detail) of the various primary and secondary tools used for both quantitative and qualitative data. This is well aligned with the evaluation purpose and objectives, and is all well-described. One minor criticism is that the first two sub-sections (1.3.1 and 1.3.2) of this section are quite repetitive - they repeat the ToC/theory-based approach and use of contribution analysis. One such sub-section would have sufficed. Finally, the use of AI (in language editing) is acknowledged, with use for no other purpose noted.
	ii Data sources are all clearly described and are relevant and robust; these would normally include qualitative and quantitative sources.	Yes	The methods section clearly and concisely describes the data sources - these are primarily qualitative (KIs, FGDs, Document Review) but with some secondary quantitative data also being collected. The different primary stakeholders are also elucidated in a useful table (table 4), and the specific secondary sources (quantitative) are noted in the main text, with further sources (qualitative) noted as part of the evaluation matrix in Annex 1. This is all well described with no issues.
	iii Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or if not, provide reasons for this).	Partially	The sampling strategy is described in both outline (section 1.2) and in detail (section 1.3), specifically noting the use of purposive and convenience sampling - which is an appropriate and standard practice approach to these types of evaluations. The detailed sampling description also covers the selection of individuals (described in Table 4) and locations (Table 5), with full details provided on the rationale. One issue is that the selection of non-programme provinces as part of the sample is not clearly explained/justified. It appears that the non-programme provinces selected are done so in an attempt to provide some kind of counterfactual analysis - but this is not explained or noted in the text. Indeed, the use of counterfactuals is challenging in contexts where they were not built into programme design at the beginning - there are many potential confounding factors that might influence development trajectories that cannot be controlled for, so such a measure needs to be carefully considered (and explained) - this is lacking in the report. Further, the evaluation describes the categories of stakeholders consulted and the data collection methods used. However, the sampling strategy is not fully detailed for all relevant aspects. Particularly, the rationale, criteria, sample sizes and measures to ensure representativeness are not consistently specified.
	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).	Yes	The evaluators are fully clear in the text of sub-section 1.3.1 and the preamble of 1.3.2 that the evaluation approach is clearly linked to the reconstructed ToC and that it will be explicitly tested and analysed during the findings. The methods themselves are commensurate with this aspiration, enabling (if applied correctly) good coverage of the main provisions of the ToC. Given the level of detail of the ToC as presented in Figure 7 it is unlikely that every element can be explored to the required level of detail. However, a higher-level testing of the ToC and associated intended results should be possible.
	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	Sub-section 1.3.2 explains the specifics of the data analysis for the evaluation. The description is relatively clear and comprehensive, linking explicitly to the assumptions and indicators in the evaluation matrix. The descriptive statistical analysis for quantitative data (e.g., from census, ZDHS, MICS) and content, context, and contribution analysis for qualitative data (from KIs, FGDs, observations) is appropriate to this type of evaluation. Further, the content analysis description notes the categorization and coding of data specifically "in relation to assumptions and indicators identified in the evaluation matrix", which is good practice for such evaluations. The report also includes a disclosure regarding AI use, i.e. that "no use was made of Artificial Intelligence in terms of data gathering and analysis," (but that it was used for language editing), in full compliance with UNFPA guidance and policies.
	vi Clear and complete description of limitations and constraints faced by the evaluation in its data collection and analysis, along with the mitigation measures implemented by the evaluation to address these limitations, where feasible.	Partially	Table 6 (p. 17-18) provides a concise description of several limitations and constraints faced by the evaluation team during data collection and analysis. It also includes, the mitigation strategies adopted. The section is short, but does accurately capture key operational challenges, such as logistical difficulties related to scheduling interviews, the emphasis on qualitative data, risking various biases and lack of statistically representativeness. The inclusion of mitigation strategies for each of the limitations/constraints is useful, and the evaluators note as a postscript to the table that the limitations were "adequately resolved and did not significantly affect the validity or credibility of the evaluation findings" - a good practice that is useful to include. While the section is minimally adequate, it is short, and the table does not explicitly include some systemic data limitations detailed elsewhere, such as the "absence of data for certain key indicators" (e.g., the Client Satisfaction Index and the Vital Medicine Availability and Health Services (VMHAS) survey), an inconsistency in the comprehensiveness of the limitations documented in this section versus the broader evaluation findings. Other examples of missed limitations are related to indicator design and outcome-level measurement (e.g. activity-based indicators for AYP), data gaps and disaggregation constraints affecting equity and LNOB analysis (e.g. DHS2 and district-level data gaps) and constraints affecting interpretation of effectiveness and sustainability findings discussed in the findings and conclusions chapters (Chapters 4 and 5) are not consistently reflected in Section 1.3.4 or accompanied by corresponding mitigation measures
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	There is explicit, but not contextualised reference in sub-section 1.3.5 to the UNEG ethical standards. The section explicitly states that the evaluation team members committed to adhering to internationally accepted ethical standards and the UNEG Norms and Standards for evaluation, referencing UNEG's documents for credibility and accountability. The section provides some minimal context, outlining that the evaluators obtained informed consent, ensured respondent anonymity and data confidentiality, allowed non-response to sensitive questions, avoided bias or misinformation, and respected dignity, diversity, and anti-discrimination principles. The report also notes application of the "do-no-harm" principle by noting "safeguarding measures", particularly for vulnerable groups like PWs, but does not specify what these measures were. It also cites other normative guidance, specifically the WHO guidelines for dealing with GBV survivors (but does not provide any further details on these guidelines or how they were applied. Therefore, the section only partially addresses the contextualized application of ethical principles, relying heavily on a generic commitment to UNEG principles without explicitly detailing the specific mechanisms used to ensure the independence of the final written report or the absence of conflicts beyond general commitment clauses.
	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	As described above, the ethical approaches section only minimally and partially addresses the requirement to provide contextualised and detailed description of how the ethical principles and commitments were ensured during the evaluation, notably that the report does not discuss ethical considerations across the entire evaluation process (incl. design and reporting phases), focusing on data collection only. The annexes (annex 4) which contain the data collection tools include some basic scripts for some of the tools, but not others. Specifically, there is no script regarding confidentiality, consent etc. for the KI tools, and while the FGD tool does include instructions for requesting consent and (then) assuring confidentiality, it is limited to this, and indeed the assurance speaks about "information exchange for the purposes of analysis" which is vague (i.e. information exchange with what actors?, what type of analysis?) and inadequate. The evaluators should have provided more details on the specifics of the approaches in the main report - including more comprehensive assurances of safety, confidentiality etc., discussed contingencies/referral plans for sensitive disclosures (such as from survivors of GBV) and noted ethical processes for any interactions with youth/children (or indeed explicitly noted an intention not to engage with this group).

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 report includes a short (two bullets) subsection on innovative practices (1.3.6). While the efforts of the evaluators to complete the section are laudable, the examples cited (Textual and visual analysis; and "Immediate sharing of draft recommendations with the stakeholders interviewed") are not innovative in any discernible manner. Per UNFPA guidance, evaluations should not be negatively rated if they attempted innovation that failed or implemented non-innovative elements, hence this is categorized as 'not rated'. While the ToR for the evaluation does 'encourage' use of innovative tools (such as ICT), the inception report takes precedence in the evaluation design, and indeed the use of AI by the evaluators (even in the limited fashion that was employed) could have been used as a genuine example, but the evaluators did not do so in the section, instead describing some useful, but hardly innovative, approaches.
SECTION E:	EVALUATION FINDINGS (weight 25%)	50%	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.	Partially	The format of the findings is largely, but not completely in line with the UNFPA guidance. The findings are organised under the evaluation criteria first, and then the evaluation questions - these are clearly noted in the text, with the latter categorization being required by UNFPA guidance. Each sub-section for the EQs is accompanied by a summary box, also noting clearly that the specifics of evidence can be found in the Evaluation Matrix in the annexes. Findings are also numbered, although there are many of them (110 in total), and they do address the evaluation questions and present some evidentiary sources within the discussion of the findings. However, the analysis is not fully systematic. Specifically, under EQ1, the findings section systematically addresses assumptions 1.1 and 1.2 (findings 1-9), but omits assumption 1.3 (related to emerging issues/climate change). Under EQ6 (Efficiency), the subheadings, while very much relevant to the efficiency criterion, are not organised according to the assumptions in the evaluation matrix. There is a similar issue with EQ7 (sustainability). Further, the reference to evidence sources is quite cursory in many cases, with simple statements that a given finding is (for example) "supported by interviews and document review" - this is not fully adequately rigorous as it does not present an argument for a given finding or an analysis that is supported by clear evidence. Thus, many, if not most, of the findings are more like subjective judgements than objective assessments as required. In other cases, the analysis is not well-written - the excessive detail is confusing and unclear - see Findings 27/28 (p. 50), which are somewhat incomprehensible. This also points to a fundamental issue around the findings themselves - they are not really 'findings' as is standard practice in evaluations, essentially they are numbered paragraphs of analysis. There are way too findings to meaningfully distill the information and provide a user-friendly assessment. A total of 110 is excessive and it would have been useful to further summarize those into a smaller number that can provide clearer foundation for the conclusions. The evaluators should have focused on condensing the analysis (discussed further below) and presenting a much more limited number of topline findings based on this.
	ii Explicit use of the evaluand's theory of change, results chain, logical framework in the formulation of the findings.	Not Rated	This question is 'not rated' as per the annotated Evaluation Quality Assessment grid of UNFPA. As discussed above, the evaluation goes into some depth of analysis of the programmatic and reconstructed ToC (sub-section 3.2.2, prior to the findings) as well as referring to the overall CP structure throughout the findings, notably under the effectiveness criterion (for each of the four programme areas). For example, the report highlights inconsistencies directly related to the ToC performance measurement, and notes under at least one finding whether UNFPA programmatic performance was in line with the ToC or not (e.g. for PD under finding 67). While there is not extensive reference to the ToC in the findings, the analysis is covered in section 3 and the EQs/Evaluation Matrix are used more heavily in the findings, meeting the standard for the criterion.
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 findings chapter (4), supported by the Evaluation Matrix in Annex 1, uses credible forms of qualitative and quantitative data, and consistently refer the analysis to the diverse sources. For instance, quantitative data from national surveys (ZDHS 2023/24, Census 2022) is used to track progress against key outcomes such as the Maternal Mortality Ratio (MMR) decline and family planning demand met. This is presented alongside qualitative data sourced from KIIs and FGDs that provide explanatory context and stakeholder perspectives on impact, including quotations to illustrate UNFPA's positioning, activities and technical assistance. The systematic integration of different sources is somewhat perfunctory, and there are quality issues on detailed scrutiny (discussed under the next criterion), but the sources themselves are those set out in the evaluation plan and the evaluators have clearly made efforts to triangulate between them, fulfilling the requirement for clear identification of both qualitative sources (KIIs, FGDs) and quantitative monitoring and survey data. Similarly, the evaluators seek to provide a combination of data at both output and outcome levels, particularly under the effectiveness criterion (Section 4.3), which directly assesses the viability of the reconstructed ToC. For example, under the SRHR programme component, the analysis details output indicators (e.g., facility readiness for EmONC - Table 13) while concurrently assessing the resultant outcome trends (MMR and unmet FP need) (Finding 24). This covers both positive achievements (e.g. the MMR/FP outcomes) and weaknesses (e.g. inconsistencies around the adolescents and youth output indicators or the absence of data for the Client Satisfaction Index (Finding 26)).
	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.	Partially	The evaluators have clearly sought to link the analysis and findings to the various sources of primary and secondary evidence, noting either in the text or via footnotes in many places the sources from which the analysis is derived. However, for some findings, an overabundance of detail presented, which is often unnecessary and serves more to obscure the validity of the finding and hence is not a balanced viewpoint. For example, under Finding 2, the evaluators note a number of national policies/strategies that UNFPA contributed to, which is good, but then provide summaries of the content of some of these strategies, which is not relevant to the finding. See also an entirely superfluous paragraph on EmONC under finding 35. For other findings, the evidence links are somewhat tenuous, or even absent. For example, finding 8, around inclusion of stakeholders in CP design, the evaluators note that the UNFPA CP was "shaped through extensive consultations with key [external] stakeholders", but no sources for this are provided - it is crucial that the evaluators triangulate and balance such findings particularly with external stakeholders, as self-reported achievements in this regard are easily questionable as potentially overly-positive and not counterbalanced with external experiences that might be less positive. Another significant example is in effectiveness, where the outcome measure of MMR reduction (finding 24, p. 49) is poorly presented - the text is incomplete/incorrect, appearing to claim a reduction of MMR from 364 to 362 but should say 364/362 to 212. There is poor corroboration of this in the Evaluation Matrix (p. 110), which notes that "UNFPA has done a fantastic job and given the reduction of MMR from 364 to 212...one can assume that UNFPA has played a key role in the reduction of MMR as result of its work on maternal health." - this is a very subjective (and highly unbalanced) statement relying on a very sweeping assumption that lacks any evidence. Finally, there is some mismatch of the extent to which evidence is presented. Some EQs (e.g. under relevance and effectiveness), provide extensive evidence from varying sources, which is well-referenced, whereas for others (e.g. under coherence), the analysis is much thinner, with little or no reference to sources of the evidence, and the analysis presented simply as a straightforward finding or statement of fact.
	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 lack thereof - from outputs to high level results).	Partially	The evaluation does a partial job of moving beyond simple description of achievements to explore underlying factors or drivers. Positively, for some aspects, the evaluators seek to unpack findings to determine these factors, such as the results of UNFPA advocacy (e.g. on government co-funding for RH commodities - finding 33) and findings related to insufficient funding (for fistula management - finding 38). There are also useful 'challenges' and 'unintended effects' sub-sections for each of the programme areas under the effectiveness EQ, which present underlying factors or determinants of how impactful (or not) UNFPA's work has been. Something similar is undertaken under the sustainability criterion, i.e. the 'non-sustainable elements' presented. However in others, as noted above, there is basic recounting of activities with little in the way of analysis of what these imply for UNFPA programming or even the extent to which they are reasonably attributable to the UNFPA programme. In some cases, the gap is conspicuously absent, such as in the effectiveness of UNFPA-supported family planning services (finding 24) where the evaluators note that variations in provincial data "has got some implications for programming which need to be thought through by UNFPA and partners", without noting what these implications might be. There is somewhat of an over-emphasis on simple listing of activities or outputs across many of the findings. This is not reserved to the analysis under the effectiveness criterion (where it might be expected) - for example, much of the latter half of finding 8 (relevance) is a recounting of activities to enhance disability inclusion rather than assessing if the needs of people with disabilities were indeed met. This contributes to one of the core issues with this report - superfluous or misplaced detail - which could/should have been pared back, and a corresponding greater focus placed on the underlying drivers of success or failure, which is missing in many of the findings.
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.	No	The report provides some, but quite cursory, review of the UNFPA M&E system under the efficiency criterion (findings 86-88), but it is quite limited. Across some of the findings, the report identifies limitations around the completeness and appropriateness of the results framework and M&E tools, for example that performance indicators were sometimes inappropriate for measuring transformation (e.g. that the AYP indicators were focused on output rather than behaviour change - finding 44). However, a more macro-level review of the overall M&E system is limited. Essentially, finding 88 confirms that UNFPA utilised the Quantum system effectively for M&E tracking, and finding 97 provides some administrative analysis of the M&E function (focusing on the resources available to fulfil the M&E function), but no more in-depth analysis is provided here. The analysis about indicator weaknesses should also have been part of findings 86-88, ensuring coherence in the assessment, but was not done so. This description does demonstrate that the vertical logic of the framework—linking national priorities through the UNSDCF to UNFPA outcomes—was actively monitored and supported through the reporting system, which is positive, but a higher rating requires a more systematic approach to the overall M&E system, in particular linking the analysis of the programme intervention logic/ToC (in section 3.2.2) to the CO's M&E system.
SECTION F:	EVALUATION CONCLUSIONS (weight 10%)	50%	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 judgements of the evaluand.	Partially	The report presents 17 conclusions, organised by 'strategic' and 'programmatic' levels. They are indeed clear and unbiased judgements of the UNFPA programme performance over the programme period. Further, they are reasonably balanced, presenting the strengths/achievements of the CP with contrasting areas where UNFPA has not achieved to the extent it might, or areas for future work or growth. The structure of the conclusions is somewhat compact and not fully in line with the UNFPA evaluation guidance, in that they are not in the format of a single judgement followed by a narrative description - they are presented as 2-3 sentences of narrative, which is somewhat short for what is required. They are, however, clearly linked to the relevant EQ/criterion, findings and recommendations. The core issue with the conclusions - evidenced by the fact that there are 17 of them - is that they are not sufficiently high-level, i.e. are more summative of findings. They are actually more akin to findings themselves. Given that the report has 110 findings, which is excessive (and the findings are not really findings, rather numbered analysis paragraphs - see 11i above), shifting the conclusions to the findings section, and relabelling as such, would have been a practical revision of the report. The evaluators could then have produced a short (6-7) list of higher-level conclusions more in line with accepted good practice. A final issue is around categorisation of some of the conclusions - specifically conclusions 5 and 6, which are operational/programmatic, not strategic.

	ii Conclusions are well substantiated and derived from findings.	Partially	The logical link between the conclusions and the preceding findings and analysis is clear - both in terms of the conclusion statements themselves, which are well-derived from the findings section, and the itemised linkages between the conclusions and the relevant EQs/evaluation criteria and findings, presented under each conclusion. While most conclusions are based on multiple findings, not all are - specifically, conclusions 3, 9, 11, 13 and 14 are based on a single numbered finding. For this evaluation, given that the convention of numbering of findings is more a straightforward paragraph numbering, this is questionable. For example, conclusion 9 is based on finding 38, which is a single short paragraph, similarly conclusion 11 (finding 10) and conclusion 13 (finding 56). This undermines the validity of the conclusions.
Question 15.	Are lessons learned identified?		
	i Lessons learned are derived from the findings and are well substantiated with practical, illustrative examples.	Not Rated	The original design report for the evaluation does not specify the need for a lessons learned section, so this criterion is 'not rated.' However, the report does provide a short list of lessons learned. These are quite generic lessons that are really just examples of good-practice programming (multi-sectoral partnerships, integrated programming, continual advocacy), rather than original or context-specific strategies or approaches that overcame unanticipated challenges.
	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	As above, the criterion is 'not rated' due to the absence of the requirement for the lessons in the design report. The lessons themselves are clearly presented, although short (a summary statement in bold followed by a single explanatory sentence) and with limited examples to illustrate their application or value. The section gives the impression as having been included in a rather token manner, rather than a genuine effort to present examples of original and useful lessons or practices that were noted.
SECTION G:	EVALUATION RECOMMENDATIONS (weight 15%)	100%	Comments on Rating
Question 16.	Are recommendations well-grounded and articulated?		
	i Recommendations are clearly formulated and logically derived from the conclusions.	Yes	The recommendations in Chapter 6 are quite clearly formulated and logically derived from the preceding conclusions, so the logic is clear. The recommendations are largely written in clear and unambiguous language and clearly outline detailed operational implications. Each is explicitly linked to one or more conclusions. For example, Recommendation 7, which calls for actively engaging the MoHCC to set operational standards for the MoONC, is logically derived from Conclusion 8, which notes that low-quality obstetric care and inadequate equipment results in complications and deaths. The section also assigns a priority level (High/Medium) and defines the target/responsibility for each recommendation.
	ii Recommendations are useful and actionable for primary intended users. Specific guidance is provided for its implementation (e.g. strategic or programmatic level, suggested actions/operational implications, and; responsible actors), as appropriate.	Yes	The recommendations are structured to be adequately actionable for the specified primary intended users. They are systematically presented with specific guidance for implementation, outlining the relevant priority level (all are assigned either a high or medium priority), and the specific responsible actor (UNFPA CO units, MoHCC, ZIMSTAT, etc.). However, UNFPA evaluations cannot and should not indicate any other entities than UNFPA units as responsible parties for implementing recommendations. There should be no reference to government bodies and implementing partners. The recommendations should only be addressed to units within the UNFPA Country Office (SRH, M&E, Comms, etc.) and may, at a maximum, include support from UNFPA regional offices. Further, 15 recommendations with several action points is excessive. The UNFPA Country Office does not have the capacity to absorb all of these and implement them in a relatively short timeframe, since all of them are either high or medium priority, especially considering the current environment where it will not be possible to secure more human and financial resources. This limits the utility of these recommendations as they are not sufficiently prioritized. The recommended actions are quite specific for the responsible parties ensures that the actions target genuine systemic issues identified throughout the report. A minor issue, is that while the majority are strong, some "high" priority recommendations, such as Recommendation 3 regarding sustainability plans for SRHR/FP, rely heavily on the non-mandatory operational implication of UNFPA needing to invest "time and energy to sustain buy-in" from government leadership, rather than purely technical action points that UNFPA has direct capacity to address, which is a slightly weaker premise for ensuring long-term fiscal commitment. This is, however, likely an issue that the management response to the recommendations can address.
	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 preamble to the recommendations chapter clearly notes the process whereby they were developed, including co-creation/validation with the UNFPA CO, the ERG and contributions from ESARO. This is in line with UNFPA guidance.
	iv Recommendations are prioritized based on their importance, urgency, and potential impact.	Yes	As noted above, the recommendations are prioritised according to either high or medium. This is accompanied by a rubric explaining the priority levels in the preamble text (although this also includes description of 'low' priority - which no recommendations were so designated, so is unnecessary). The recommendations are not organised in order of importance, but this is not a mandated practice so it does not affect the rating.
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 the evaluation manager, name of organization commissioning the evaluation, list of the evaluation reference group members, acknowledgements, table of contents (including, as relevant, tables, graphs, figures, annexes), and list of acronyms/abbreviations.	Yes	All of these elements are present in the title/opening pages.
	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), additional details on ethical considerations, (country) case study reports, thematic papers or "deep dives."	Yes	All of the mandated annexes are included, plus an additional annex (7) on the country programme reported performance against the results framework by 2024 and another annex (5) that is not required, the stakeholder map, is also included. The list of stakeholders (in the 2.0 version of the report) the names of the stakeholders consulted is redacted, as per UNFPA policy.
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).	Partially	The report is adequately formatted, with some minor errors in style/spelling/formulation - these are covered below. The report is well-structured, in line with the UNFPA guidance on CPEs. There are some inconsistencies in layout/formatting, however, specifically: - The table numbering between the table of contents and the main text is out - it appears that the TOC goes from table 11 to table 13, which corresponds to table 12 in the main text (and subsequent numbers are out by one as a result). - Some page numbers are missing (e.g. p. 1, p. 35) - Some incorrect formulations - perhaps an artifact of use of AI for editing (e.g. use of "seven point six million" instead of "7.6 million" on p 21, para. 3) - Annexes are not always correctly numbered (Annex 7 on p. 17: the text should be labelled Annex 6) - The report notes that AI was used for language editing but it still has some copy-editing issues.
	ii Structure and length accords to UNFPA guidelines for evaluation reports; it does not exceed number of pages that may be specified in ToR.	Yes	The report (excluding cover pages and annexes) comes to exactly 80 pages, which is in line with the mandated guidance from UNFPA. This said, there are various areas of excessive detail that could have been truncated to shorten it further; although, to balance this, some findings are quite thinly presented, so could have been improved with more content.
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 reasonably well-written and fully comprehensible. The main criticism is that the writing is long-winded and repetitive in many cases, and the report is unnecessarily long in some sections as a result. Further, some of the text is misplaced in various sections/subsections. For example the sub-section on SDGs (2.1.1) has information about UNFPA's transformative results, which should be in section 3 only. Similarly, sub-section 2.1.5, on PD, has (at the end) data on general national strategies in relation to the SDGs and UNSDCF - this should be in the following section.
	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 makes reasonable use of visual aids, primarily in the sections preceding the findings, with some tables in the findings sections to illustrate programme performance. No charts or figures are presented in the findings section, and no other types of visuals (beyond tables/figures) are used. All are captioned, but there is some inconsistency between the text references to tables after #11 (noted above). Finally, the resolution of some figures (Bs 4-6) is very poor and the details cannot be discerned.
SECTION I:	CROSS-CUTTING ISSUES (weight 10%)	69%	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 rights holders, marginalized and vulnerable persons, young people, people with disabilities, migrants or refugee populations, indigenous communities, and other persons that are often left behind.	Yes	The evaluation's approach and methods demonstrate an explicit commitment to capturing the voices of a wide range of rights holders, including members of marginalized groups, driven by the principles of human rights, gender equality, and LNOB. The evaluation methodology is variously described as "gender-sensitive" (in the executive summary, p. 1), "gender transformative" (scope subsection 1.2.1, p. 10), "gender-responsive" (subsection 1.3.2, p. 12) (although each of these terms has a different meaning) and prioritised inclusion of marginalized groups. The design of the data collection explicitly specified targeting key beneficiary groups, notably women and adolescents and youth, and outlined mechanisms to gather input from vulnerable populations. For instance, the evaluation team specifically sampled people with disabilities and "diverse participants, including vulnerable groups" in the FGDs. The data collection tools do reflect this commitment, with all primary data tools containing various specific or general question related to gender, disability and/or human rights and other vulnerable groups (although none are specifically itemised). This warrants a 'yes' rating, although the report could be more specific in the data collection tools about marginalized and vulnerable groups and specific gender issues in the guides for KIIs and FGDs.

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	The evaluation questions (and the assumptions that the evaluators have derived from them) clearly integrate issues of gender, vulnerability, disability and inclusion, as this is a stated priority of the CP under evaluation. Of the seven EQs, four (EQs 1, 2, 4, 5) explicitly articulate a focus on these groups, meeting the criterion. These dimensions are then reflected in the assumptions (although merely noting "ensured that no one is left behind"), and on through the data collection tools.
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), as feasible.	Partially	The findings of the report (and supporting sections) demonstrate an effort to disaggregate data by population groups where possible and relevant to UNFPA's portfolio, primarily focusing on age, gender, and geographic location to assess differential results and identify disparities. Age disaggregation is frequently provided, for example, noting the unmet need for family planning and HIV incidence is higher among younger people aged 15-19 years (Finding 48). Geographic disparities are also included, for example the MMR being higher in rural areas than in urban areas (Finding 24), which is then related to differences in access based on education and rural/urban status. The analysis also discusses weaknesses around data disaggregation, such as gaps in Ministry of Health and Child Care data tools (DHIS2), which are not yet configured to capture data on PWDs (Finding 83, Conclusion 16). However, while the findings present some disaggregation were (mainly secondary) data is available, gender (or other) disaggregation of primary data is limited to that of the description of respondents who were included in the evaluation (59% women and 41% men, with 2% of respondents PWDs). Although the evaluation does seek to focus on gender and LNOB considerations (forming a key part of the fundamental design of the evaluation), the actual disaggregation when it comes to primary data presentation is lacking - for example the evaluators do not quote the 'voices' of women, PWDs or other vulnerable groups specifically. Thus a 'partial' rating is warranted.
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	The evaluation speaks to a certain extent to intersectional analysis, particularly through use of some disaggregated (secondary) data to show how overlapping identities compound exclusion or disadvantage, and this analysis is reflected in the conclusions and recommendations. The evaluation design emphasises LNOB as a lens, with some of the analysis exploring how various vulnerabilities intersect: for example the analysis of FP needs met or MMR by geographic location (urban/rural) and age (Findings 24, 27, 48). Similarly, the assessment of secondary data on GBV is analysed by geographic location and economic status (noted in the background in section 2.1.3). While this secondary data analysis is positive, it is not especially extensive (i.e. not across more than half of the EQs). The evaluators do note the challenges faced in accessing some specifically vulnerable groups that might allow a deeper analysis (specifically the challenges in identifying and working with key populations (LGBTQI+) due to legal restrictions, which limits the UNFPA's ability to provide extensive support, an important finding.
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.	Partially	The evaluation report integrates the relevant cross-cutting issues well throughout the findings, conclusions, and recommendations. Findings routinely (if not very deeply, i.e. not "transformative" as is claimed in the methodology, but more "sensitive") assess differential results where data is available to highlight disparities through the lens of gender, LNOB and inclusivity. This is reflected across the conclusions and recommendations: for example Conclusion 2, which affirms that the CP addressed the needs of vulnerable populations, including considering "disability, human rights, and gender", and Conclusion 16, which notes weaknesses in capturing data on PWDs. Following on from these, Recommendation 6 calls for ensuring LNOB is fully integrate in UNFPA strategy and programmes to be responsive to intersecting vulnerabilities. However, a significant gap is around analysis related to humanitarian issues and climate change-induced natural disasters, the latter of which, despite being explicitly noted in EQ1, is not present at all in the findings (there are two very cursory references), with the only related UNFPA activity being the short-term humanitarian responses to an El Niño drought anticipatory action, rather than explicit integration. Further, despite the explicit focus on vulnerability and disability inclusion, only 5 of the 268 individuals engaged in the field visits were identified as PWDs, indicating that the methods used did not fully overcome barriers to achieving representative participation for this specific vulnerable group, regardless of the intent. Furthermore, the use of different terms on the gender continuum (gender sensitive, responsive, transformative) in the methods suggests that the evaluators were not fully conversant with this approach.
vi	Inclusion of young people in the evaluation team and/or Reference Group.	Yes	The evaluation team (as specified in the ToR) included a young and emerging evaluator as a core team member (Ms. Exhibit Matumbi), noted as part of the team in the cover pages.
Question 21. Does the evaluation meet UN SWAP evaluation performance indicators?			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.	5	
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated	The evaluation objectives and scope explicitly and comprehensively include analysis of HR&GE as a core element of the evaluation, thereby meeting the first and second criteria. The thematic scope explicitly includes human rights and gender equality as cross-cutting issues, and EQs 1 and 5 are dedicated to assessing the relevance of UNFPA programming to rights-related and gendered needs and the achievement of HR&GE-related results, specifically the "advancement of gender equality and the empowerment of all women and girls" (EQ5iii). The evaluation partially, but incompletely, meets the third and fourth criteria by noting that data gathering explicitly aimed to obtain disaggregated data and focused on intersectionality to identify inequalities. But the indicators noted in the Evaluation Matrix are not transformative, rather sensitive. For example, assumption 1.3 has vague indicators to operationalize how UNFPA may have responded to differential and evolving needs. In addition, it did not examine how UNFPA may have prioritized the needs of some vulnerable groups over others, given resource constraints that are increasingly growing. Furthermore, a "yes" rating requires that the "evaluation includes a gender and power analysis, an intersectional analysis, and/or references the relevant normative instruments or policies related to HR&GE." A power analysis of stakeholders is done in Table 3 but it remains superficial. The background section does include references to international normative frameworks for GEWE, thus meeting this criterion partially, but overall not fully.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Partially integrated	The evaluation report does reflect a reasonably robust and comprehensive gender analysis, although does not meet all four criteria, and not all to the same extent. Specifically, (as discussed above) the findings integrate some gender disaggregation of data across key indicators where available (using secondary sources), although explicit and transparent triangulation of analysis/findings across different groups is quite limited and thus not very visible. The evaluation conclusions explicitly discuss human rights and gender equality implications of programming, acknowledging successful adaptation to emergent issues while highlighting shortcomings, such as limited support for LGBTQI+ populations and the unintended negative impacts of policy changes that restricted adolescent access to SRH services. Finally, the recommendations clearly address gender equality and women's empowerment issues with specific priorities for future action (e.g. Recommendations 1 and 10), including noting a need to reinforce LNOB considerations in programming (Recommendation 6), the mandate to strengthen disaggregation by sex, age, and disability in data collection tools (Recommendation 14), and the emphasis on scaling up economic empowerment for GBV survivors (Recommendation 11). However, in the recommendations, while gender equality is clearly addressed, specific priorities for advancing women's empowerment beyond GBV (e.g. structural gender inequalities or economic empowerment) are not consistently articulated. UNFPA gender work is not limited to GBV prevention and response, but should underlie all the work it does across its mandate. For the highest rating to be assigned, findings should also be transparently triangulated and validated to credibly represent the diverse groups. However, as discussed above in relation to the findings, this is not the case. It is also expected that "the conclusions discuss human rights and gender equality implications, including the interventions "unintended or unanticipated impacts on diverse groups." However, no unintended consequences are discussed in the conclusions. Therefore, the evaluation only fully meets one of the four criteria set in the annotated EQA grid (August 2025). This warrants a "partially integrated" rating only.

- 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

