

UNFPA Evaluation Quality Assessment Grid

Version: May 2024

REPORT RATING SUMMARY			
Overall Rating		59%	Fair
● ● ● ● ●	Excellent	5	
● ● ● ● ●	Highly Satisfactory	4	
● ● ● ● ●	Satisfactory	3	
● ● ● ● ●	Fair	2	
● ● ● ● ●	Unsatisfactory	1	

REPORT DETAILS	
Title of the evaluation report	Country programme evaluation of the United Nations Population Fund (UNFPA) Eswatini 2021-2025:7th Country programme
Region	ESARO
Country	Eswatini
Year of report	2025
Business Unit/programme country (managing evaluation)	Country Office
Date of assessment review (dd/mm/yyyy)	21-Aug-25
Name of assessment review firm	IOD PARC
CLASSIFICATION OF EVALUATION REPORT	
Primary SDG(s) covered (list provided below)	3, 4, 5, 10, 16, and 17
UNFPA Strategic Plan areas covered (lists provided below)	
Three transformative results	Yes
Six outputs	Yes
Six accelerators	Yes
Organizational effectiveness and efficiency	No
Humanitarian evaluation	Yes
Evaluation evalund (e.g. country programme/intervention/policy/thematic area)	Country Programme
Evaluation type (e.g. formative, summative, developmental)	Summative
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 7th Country Programme Evaluation (CPE) of UNFPA Eswatini 2021-2025 is rated satisfactory, and is largely well-planned and designed, with a strong foundational approach to human rights and gender equality. However, it is significantly hampered by weaknesses in the strength and depth of its evidence base, thus limiting overall report quality. The evaluation was inherently complex, given its nature as a CPE covering multiple thematic areas over a multi-year period and across various national and global (COVID) crises. The top-line strengths and weaknesses are:

Strengths

- The evaluation's design and findings integrated GEWE and HR considerations well. This is evident through its use of a good mix of target stakeholders, data disaggregation by sex, age, and disability, and a deliberate inclusive and participatory approach that engaged vulnerable groups and ensured meaningful participation throughout.
- The evaluation methods and design were robust, employing a mixed-methods approach combining qualitative and quantitative data, and explicitly adopting contribution analysis to examine causal linkages, which is appropriate for a complex programme evaluation if applied well.
- The report includes an assessment of the intervention's ToC, noting its logical coherence and areas where future versions could be strengthened, such as specifying risks and assumptions under each thematic area.
- The recommendations are generally well-formulated, actionable, and include responsible actors and operational implications, indicating practical utility for future programming. Key stakeholders, including the Evaluation Reference Group, were involved in their validation.
- The background section provides a clear and detailed socio-economic and political context with disaggregated data on key vulnerabilities. However, it has some inconsistencies in demographic data citation and lacks specific geographic programming details.

Weaknesses

- A major limitation is the inconsistent and often inadequate citation of evidence to support findings. Many statements are presented as unvalidated facts, with little recourse to supporting evidence from collected data, significantly undermining the credibility and analytical rigour of the report.
- While contribution analysis was a stated part of the methodology, the report often lacks detail on the precise causal mechanisms leading to achievements or non-achievements of results, limiting the "why" behind outcomes and primarily remaining at a general output level.
- The executive summary misses some key elements and requires thorough proofreading to ensure it is of optimum quality for all readers.
- The professionalism and readability of the report is impacted by numerous formatting errors, inconsistencies in numbering (findings, conclusions, recommendations), grammar errors, typos, and confusing visuals that detract from readability and navigability.
- Although purposive sampling aimed for diverse stakeholders, the description of the sampling strategy is very brief, lacking specific criteria or a detailed rationale for selecting particular organisations or rights-holders.
- While the report explicitly states adherence to UNEG ethical guidelines and other codes, it does not provide sufficient evidence of how these were contextualized for the evaluation. Crucially, the data collection tools do not explicitly mention obtaining consent, and the report is vague on how it ensured dignity and diversity for vulnerable groups.

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

While the evaluation report demonstrates notable strengths in the background and design elements, it is undermined by shortcomings in its evidence basis, methodological transparency, and overall report quality. For senior management, the top-line recommendations to improve future evaluations are:

- Enhance Evidence Substantiation: Ensure a rigorous and consistent approach to ensuring that every finding/sub-finding and assertion in the analysis is clearly linked to its supporting evidence (e.g., specific reports, KIs, FGDs, or secondary data). Linkages to the evaluation matrix to show where the evidence can be found are also required as per UNFPA evaluation guidance.
- Deepen Causal and Disaggregated Analysis: Provide more granular detail on causal mechanisms leading to results, moving beyond general output levels, and seek to demonstrate how data triangulation from multiple sources (especially diverse rights-holder voices) informs findings.
- Improve Report Quality and Presentation: Conduct more thorough proofreading and editing to eliminate clear and egregious formatting errors, and minimise less serious issues such as inconsistencies in numbering (findings, conclusions, recommendations), grammar, spelling, and confusing visuals throughout the entire report.
- Detail Methodological Transparency: Provide a clearer and more complete description of all methodological elements, including the sampling strategy, specific criteria for selection of all participants (institutional and rights-holders), and elaborate on the specific analytical techniques used for processing qualitative and quantitative data - this may be excerpted to an annex if required.
- More Detailed Ethical Adherence Explanation: While the evaluation explicitly notes adherence to UNEG ethical guidelines (e.g., consent, confidentiality, avoidance of harm, dignity and diversity), further contextualization of their application should have been undertaken to enhance clarity - a table that provides an overview of how ethical considerations were addressed across all the evaluation phases (and not only data collection) should be included in future evaluations.
- Correct Evaluation Question Consistency: Ensure there are zero inconsistencies in foundational evaluation elements (such as the evaluation objectives and questions) between the design report and the final evaluation report and matrix, ensuring they are accurately formulated and consistently numbered.
- Clarify Visual Aids: Ensure all figures and tables are accurately labeled, consistently numbered, and easily decipherable.

SECTION RATINGS	
SECTION A: EXECUTIVE SUMMARY (weight 5%)	67%
Question 1. Can the executive summary inform decision-making?	Comments on Rating
Is a clear, standalone document useful for informing decision making, (a minimum of 5 pages, up to a maximum of 7 pages). Note: YES - the executive summary is within the indicated maximum page limit. PARTIAL - the executive summary exceeds the maximum page limit by 1 to 2 pages. NO - the executive summary exceeds the maximum page limit by more than 2 pages.	The report includes a five-page executive summary, per the UNFPA requirements, and formatted as mandated in the UNFPA evaluation handbook, fully meeting the criterion requirements. It is clearly written although it has a number of formatting/formulation errors that should have been picked up and amended as a priority given that it will be the most widely-read element of the evaluation report.

	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	Most of these elements are included in the summary, but the overall context of the location and the intervention is missing - the summary moves from the purpose/scope into the methods and then findings. The section also summarises the purpose and objectives of the evaluation poorly - missing one of the four components of the purpose as articulated in the ToR and Design Report, and also omitting the 'coherence' criterion which is part of objective 1. Although noted in the title, there is no mention of any of the elements (thematic, geographic or temporal) of the scope of the CPE in the summary. Furthermore, the methods section of the summary does not really describe the methods, but rather focuses on the phasing of the evaluation, with the specific methods used (KIs, FGDs, Observations) only briefly referred to.
	iii	Includes all significant information in a concise yet clear manner to understand the theme, intervention, programme, project and the evaluation.	Partially	Notwithstanding the absence of some of the mandated elements noted above, the remainder of the executive summary is well-written, with a reasonable summary of the main findings and the recommendations, although there are some issues with the conclusions, described below. The purpose/scope section includes some details on the main priorities of the UNFPA Eswatini 7th Country Programme, and the methods. The findings themselves synthesise the analysis in the relevant section of the main report reasonably well and concisely, with the recommendations summarised likewise. However, it is advised to check the summary of the conclusions, which are structured according to thematic areas, rather than the entirety of the programme, and have some inconsistencies. For example, the first conclusion statement notes that "The SRHR approach was viewed as overly clinical, with insufficient investment in social behaviour change communication (SBCC)." This conclusion is not reflected in the main report conclusions, which do not refer to 'overly clinical' SRHR work, nor SBCC in the context of SRHR work (beyond noting some male engagement work as being important). Revisiting of the conclusions in particular to ensure they accurately reflect the main report is recommended, hence this 'partially' rating.
SECTION B:		BACKGROUND (weight 5%)	70%	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.	Partially	Overall, chapter 3 of the report provides a reasonably clear description of the background components of the UNFPA Eswatini 7th CP. The section notes fundamental alignment with national plans, the UNSDCF and UNFPA's own strategic plans (2018-2021 and 2022-2025). It also describes UNFPA's modes of engagement, including advocacy and policy dialogue, capacity development, knowledge management, and partnerships and coordination. The CP structure is discussed in detail via the programme theory of change and narrative accompanying this. With respect to geographic coverage, chapter 3 does not provide any specifics on programming locations. The preamble to the report does include a map with programming locations stated as being highlighted, but it is simply a map of the country. The evaluation scope does note the four regions of Eswatini as areas where programming has been implemented, but this is not very specific as it encompasses the entire country. Thus, the geographic component is not clear. The implementation period for the 7th CP is clearly defined as being from 2021 to 2025. The evaluation itself covers the period from 2021 to the first three quarters of 2024, as implementation was ongoing at the time of the evaluation. This is clearly explained. In terms of cost and budget, the chapter does describe the overall budget and gives details of annual and sectoral breakdowns, including implementation status/budget utilisation, showing varied expenditure proportions across thematic areas and years. However, there is an issue with the description of the country programme budget as presented across the narrative on page 24 and the table 7 on page 25. The individual year totals for programme budgets (presented by year in the narrative) do not match with the sum of the sectoral allocations in the table. In some cases the totals in the table exceed the stated budget for the year in the narrative, so the discrepancy is not due to omitted elements such as OEE or management costs. The main partners involved in the programme's implementation are categorised in section 3.1 as the "Government of Eswatini, civil society, bilateral and multilateral development partners, United Nations organisations, the private sector, and academia". Specific partners are noted in the CP theory of change presented on pages 20 and 21.
	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	Chapter 2 of the report provides a clear and quite detailed description of the context of UNFPA's work in Eswatini, encompassing economic, social, educational, gender, health/developmental and political aspects, and explicitly links these to the implementation of the CP. The section provides key data on SRH and GBV dimensions. The report also describes some of the humanitarian or crisis dimensions, e.g. civil unrest and the COVID-19 pandemic, which reversed gains in various sectors and increased vulnerabilities. Furthermore, the report describes the national development policy context (e.g. the NDP and various sectoral strategies), as well as international frameworks such as the UNSDCF and UNFPA's own strategic plans discussed under the UNFPA programming description. A key missing dimension in the country context is the prevalence of specific population vulnerabilities, such as disability and marginalization (e.g., LGBTQ groups). While the section commendably highlights inequities for youth and women, the data is not consistently disaggregated by key factors like disability status and other vulnerabilities. Given the strong focus of the UNFPA SP 2022-2025 on leaving no one behind (LNOB), this section requires a stronger intersectional analysis. There is also a minor issue with the table of key demographic data on page 13 (table 3) - this provides "current (2024) " data per the header, but the source is cited as the 2017 census - if the 2017 provides projections for the future, this should be noted if so. Further, the data in Table 5 does not correlate well with the descriptive text and the final %GNI figure seems incorrect. This should be rechecked.
	iii	Linkages drawn between the evaluand and the ICPD benchmarks and SDGs relevant targets and indicators.	Yes	The report explicitly states in the description of the UNFPA programming (chapter 3) that the 7th CP is aligned with the 2030 Agenda for Sustainable Development, specifically Goals 3, 4, 5, 10, 16, and 17. There is no dedicated SDG dashboard table within chapters 2 and 3. While the Key Facts Table page viii provides data on population and health indicators and their statuses that relate to the SDGs (such as maternal mortality, <5 mortality, adolescent birth rates etc.), it does not explicitly include SDG performance data. The report also discussed the status of gender equality in Eswatini (SDGs) with relevant data on the Gender Inequality Index (GII), GBV incidence and early marriages. While a specific status is not provided for SDG 13 (Climate Action) and SDG 16 (Peace, Justice, and Strong Institutions), the report acknowledges that persistent climate change-related disasters and civil unrest in 2021 have reversed gains and increased vulnerabilities, directly impacting the country's development landscape and UNFPA's work. Thus, while the SDGs are noted, the specific targets and indicators related to the relevant goals are not directly discussed, but the dimensions and indicators that are relevant to the UNFPA mandate, such as the maternal mortality rate, the unmet need for family planning, or the prevalence of gender-based violence and harmful practices, are indeed discussed. This meets the standard for a 'yes' rating.
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	The evaluation Design Report includes a stakeholder map in the annexes to that report (referred to, variously, in the Design Report as "Annex B" or "Annex 5" even though it is actually Annex 7), that was developed by UNFPA Eswatini as part of the evaluation TORs with the intention of being expanded on by the evaluators (as noted in the TOR - the annexes to the TOR were not available to cross-check between the TOR and Design Report versions). This consists of a detailed list of partners by UNFPA CP outcome, with specific names and positions of relevant individuals also included. This map is not referred to or included in the evaluation report, although the ToC for the UNFPA country programme does include fields with information on key stakeholders under each outcome. While the map in the Design Report is extensive (and thus meets the 'yes' criterion), it is somewhat challenging to understand as stakeholder organisations are not presented with any explanatory details, so the specific typology of a given organisation (referred to by their acronyms) is unclear (e.g. UN, Government, NGO etc).

	ii Stakeholders are analysed to understand their specific rights, duties, needs, interests, concerns, and potential impact on the evaluand.	Partially	While the evaluation report does not include a specific stakeholder analysis, it does provide information across Chapters 2 and 3 from which these elements can be inferred. The report describes the various stakeholders involved and their general roles within the context of the 7th CP), allowing for an implicit understanding of their positions. For example, regarding interests and needs, the report details the country's national development plans, international frameworks (e.g., UNSDCF, SDGs, ICPD Programme of Action), and UNFPA's own strategic plans, thus noting the shared interest among the Government of Eswatini, UNFPA and sister UN agencies in addressing core issues. The report touches upon power and influence and potential impact primarily through descriptions of partnerships and the ToC. The Ministry of Economic Planning and Development's role in "overall programme coordination", UNFPA's own "leadership role in several technical and coordination groups" and its active participation in UNCT coordination mechanisms, describe the dynamics of the UN system and with national partners. The report also notes the CP's engagement with various implementing partners (e.g., government ministries, civil society organisations, academic institutions, other UN agencies). Thus, the evaluation report describes the context and the network of stakeholders involved in the 7th country programme but does not present a formal, systematic stakeholder analysis. The stakeholder map present in the design report does, however, meet the key need for a list from which a sample can be drawn for KIs among institutional stakeholders. That stakeholder map does not include rights-holders, so sampling for primary research at this level cannot be done from the map as presented.
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	The evaluation purpose is clearly defined, and is, in essence, identical to that articulated within the evaluation TORs and the Design Report. Similarly, the evaluation scope is clearly distinguished in its own subsection. These concisely and accurately indicate the need for the evaluation (conclusion of the 7th CP and design of the 8th). The key users, i.e. the audience of the report, are clearly stated in the high-level objectives of the evaluation.
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	There is a clear distinction between the evaluation purpose, objectives and scope - as noted above, the purpose is a direct reflection of that in the TOR and Design Report, with the objectives adjusted to be more coherent. The original TOR version had an overall purpose, two high-level objectives and four more specific objectives, these have been paraphrased/summarised in the evaluation report, mostly accurately, although the last objective (#4) of the evaluation report notes the drawing of "conclusions and lessons-learned" whereas the TOR and Design Report do not mention lessons-learned in the objective. Indeed, there is no specific lessons-learned section in the report, so the point of including this in the objective is not clear - no explanation for the deviation is given. The evaluators could/should simply have copied the objectives from the TOR, as was the case from the Design 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	Similarly to the purpose, the evaluation scope is clearly defined, with clear geographical, programmatic and temporal aspects noted in the relevant subsection, although the specific geographic locations are not noted in the report as they are in the TORs (the report simply notes "interventions implemented at national, regional and community levels". This is reasonably well aligned with the TORs, although it should be noted that the Design Report is missing a description of the scope in the relevant section (1.2 in the Design Report).
SECTION D: EVALUATION DESIGN AND METHODOLOGY (weight 20%)		64%	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>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.</i>		
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 report questions are articulated in two places: a list in the main report in section 1.3, and again via the evaluation matrix which is presented in Annex 1. The questions themselves are linked explicitly to the OECD DAC evaluation criteria, including two criteria related to humanitarian action. This is standard and acceptable practice. There are several issues with the list of questions as presented in Section 1.3: 1. An area of inconsistency is between the list of evaluation criteria specified in the evaluation objective 1 (relevance, coherence, effectiveness, efficiency and sustainability), and the wider list of criteria specified under the evaluation questions preamble, which includes coordination, coverage and connectedness. The latter two are present in the list of EQs (1.3) and evaluation matrix, whereas the 'coordination' criterion is not addressed anywhere else (other than in the findings - discussed below). Coordination as an evaluation criterion was removed from the 2024 Evaluation Handbook and should no longer be used in UNFPA evaluations. 2. There are clear inconsistencies between the list of questions in the TOR and the evaluation report. For example, EQ 7 has been incoherently transposed - it misses elements of UNFPA programming that were described in the TOR, copies the text of EQ 8 into the question body. Further, EQ6 is a fragment of EQ5 which has been separated - thus throwing out the overall numbering scheme. EQ8 (Efficiency) is missing - the text is a fragment of EQ7, rather than the original TOR question. There are various other formulation errors in the list. These seem to be editing errors that originated in the Design Report, were not caught and addressed therein, and the list has been simply copied across without further editing. 3. There have been some changes to the text of questions (combining of EQ10 and 11 in the original TOR), as well as changes to the coordination bodies for GBV in EQ 5. The amendment of EQs has been noted in the Design Report, but not in this report. 4. Finally, the TOR has 13 EQs, the Design Report states that it has 12 EQs, but in fact has 11, and the evaluation report reflects the list in the Design Report. These editing and writing errors aside, the EQs themselves, including their sub-components (not formally broken into sub-questions) are well aligned with the evaluation criteria and are appropriate for meeting the evaluation purpose and objectives.
ii	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.	Partially	With respect to the issues noted under the previous criterion in relation to the evaluation questions, some of these errors have been reflected in the questions as articulated in the evaluation matrix (e.g. the incorrect copy/pasting for EQ6), but this has not been propagated across the analytical aspect of the matrix itself (i.e. in the assumptions), further indicating that this is an editing error that is easily rectified, rather than a fundamental error in the evaluation design. There is one exception to this, however - EQ 10 (Coherence), which has two questions in the TOR, but was combined into one in the Design Report (the geographic component integrated into the question) - this has not been reflected in the Evaluation Matrix, which simply omits the original EQ10, and keeps EQ11 as per the TOR. Beyond this clear omission of an evaluation question between the Design Report and Evaluation Matrix (which justifies a 'partial' rating for the criterion), the general structure of the analytical approach - the DAC criteria, the EQs, the assumptions, indicators, sources of information - are all present in the matrix and are generally well-articulated and appropriate to answer the EQs.
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 parts of the results chain that are applicable to, or are being tested by, the evaluation.	Yes	The evaluation report provides a clear description of the intervention's intended results and the relevant parts of the results chain being tested via the country programme ToC, presented in both visual and accompanying narrative formats in section/chapter 3.2.2 (pages 19-24). The ToC complemented by details on the 7th CP components, outcomes and related UNSDCF outcomes on table 6 on page 22, which systematically lists the four interlinked outcome areas.
	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).	Partially	The ToC and accompanying narrative, while clearly articulating the four thematic outcomes, which are aligned with the UNFPA Strategic Plan, national development priorities and the UNSDCF outcomes, provides insufficient information on critical assumptions related to causality, context and stakeholders. While it can be argued that many of the 'problem root causes' in the ToC diagram could be assumptions about stakeholder characteristics, especially their needs, and the context, they are not labelled as such. In addition, the contextual assumptions included in the dedicated box in the ToC diagram are too generic to be useful (e.g. peace and security will be maintained). Furthermore, the ToC diagram is overly focused on problems and activities, rather than the results and causal relationships between them. Although the ToC design is logically linked to the results framework of the country programme, including the output and outcome indicators, there is limited information on assumptions about causal relationships (explaining what is going on behind the arrows in the diagram), contextual factors (explaining the conditions under which the interventions will work), and the stakeholder characteristics (explaining the needs, capacities and motivations/interests of stakeholders to support the interventions). It is also important to note that the evaluators themselves criticized that more information on the assumptions is required.
	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.	Partially	The ToC diagram visually depicts the logical flow of programme results, with the accompanying narrative in the report providing a critical assessment of the ToC's logic and internal coherence, affirming its comprehensiveness and links to the four country programme sectoral outcomes, and due alignment with the UNFPA Strategic Plan, national development priorities and the UNSDCF. The evaluation report critiques the design logic and links to outputs and intended outcomes, identifying several areas where the ToC's plausibility and completeness could be strengthened, suggesting revisions for future iterations (for example specifying more clearly risks and assumptions under each thematic area). Further, the evaluators note the absence of a documented risk mitigation plan and a clearly stated timeframe for achieving the outcomes. While it is commendable that the evaluators provide an assessment identifying where refinements are required for improved effectiveness and accountability, they fell short of revising the ToC for the CPE. A critical review and refinement of the ToC is clearly required in the UNFPA Evaluation Handbook and would have been more useful than giving recommendations (p. 24) on how to improve the ToC in the future.
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	The overall evaluation design and methods are briefly described. This can be a strength in terms of limiting the overall length of the report, but generally, additional methodological details are annexed - this is not the case with this CPE. The evaluation approach is explicitly stated as a mixed-methods approach in section 1.3.2. It also applied a contribution analysis approach, which relies on an explicit ToC to examine causal linkages and reduce uncertainty about attributing observed results to UNFPA's interventions. The evaluation is described as "highly participatory". The evaluation questions are well-aligned with the stated purpose and objectives of the CPE and the evaluation questions, categorised under the chosen DAC criteria (discussed above), directly address these areas. A small area of improvement would be to explicitly note in the methodology that a theory-based approach was adopted, which is not done. However, it is clear from the heading "contribution analysis and theory of change" that such an approach was adopted. The data collection methods are explicitly mentioned in the report as mixed methods, i.e. Key Informant Interviews, Focus Group Discussions, and Direct Observation for primary data, alongside document review for secondary data. These methods are generally held good practice for evaluations and thus suited to the development and humanitarian context and are designed to capture the perspectives of diverse stakeholders, including vulnerable and marginalised groups, as the evaluation was guided by the "Leaving No One Behind" principle and made a "conscious effort... to include PWD and people from vulnerable groups" during participant mobilisation (although the term 'conscious effort' could be defined more specifically). Additionally, KIIs were conducted with a broad range of internal and external stakeholders at national and sub-national levels, covering various partners crucial for comprehensive understanding of the context and UNFPA's programming. The report does not explicitly state whether or not AI was used in the evaluation.
	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).	Yes	The evaluation includes both primary and secondary, quantitative and qualitative data sources, with a good diversity of typologies. For primary data, the report clearly lists KIIs, FGDs, and Direct Observation. Secondary data collection involved "extensive review of relevant documents". The methodology section (1.3.2) notes the sources, with the annexes providing additional source details. Secondary quantitative data sources include the 2017 Census Preliminary Results, MICS 2022, Labour survey 2021 and SHIMS 3 in 2021. The data collection methods mentioned are designed to capture the perspectives of diverse stakeholders, including vulnerable and marginalised group and a wide range of institutional stakeholders both internal and external to UNFPA.
	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, while noted (section 1.3.3 - a single paragraph), lacks important details that would normally accompany a report of this nature, nor is any further information on sampling provided in the annexes. Positively, the report explicitly states that a purposive sampling technique was used for the selection of informants, FGD participants, and health facilities for direct observation. However, there is no further detail beyond this simple statement. The actual final sample size of research targets is also noted in section 1.3 - the report indicates that 46 KIIs were conducted. While the report states that KIIs covered a "full range of partners" and the sampling followed UNFPA guidelines on stakeholder selection and sampling from the Evaluation Handbook, the specific criteria or detailed rationale for selecting particular external organisations or government officials to ensure representativeness or minimise specific biases are not detailed. This is also the case with the selection of geographic locations (6) and participants in FGDs (6 - 50 participants in total). While the report states that participants were disaggregated by age and sex and that groups included "adolescents, young mothers, midwives and peer educators" (section 1.3.2) as well as making a "conscious effort" to include PWDs and people from vulnerable groups, the specific criteria for their selection or how they were chosen to represent a broader context (e.g., urban/rural balance, facility size/type beyond "public health facilities" for geographic locations) are not explicitly articulated in these sections, beyond a general reference to "UNFPA guidelines".
	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).	Yes	Although the report does not specify that the evaluation is theory-based, it can be inferred that it is indeed so, given that it elucidates the CP theory of change clearly and applies a contribution analysis approach to examine causal linkages and reduce uncertainty about the extent to which observed results could be attributed to UNFPA's interventions particularly at the output and outcome levels (section 1.3.1). While the methods described in section 1 provide a broad outline rather than a detailed explanation of how causal connections between outputs and expected outcomes are rigorously tested, the mixed methods approach combining primary and secondary qualitative and (some) quantitative data allows adequate triangulation to test programme outcomes and explore the reasons for success or failure.

	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.	Partially	As with other aspects of the methods, the approach to analysis of data is present, but very general and lacking specifics - a single paragraph. The report notes (in section 1.3.2), as part of the mixed-methods approach, a combination of secondary quantitative analysis with primary qualitative data collection. While the report states that data was analyzed "along the evaluation questions and triangulated with different sources to ensure robustness and validity" and refers to joint analysis and validation of preliminary findings, it does not provide any more details about the specific analytical techniques or methodologies employed for processing the qualitative and quantitative data beyond broad statements. For example, no information on how qualitative data was coded in line with the evaluation matrix assumptions or questions is provided, as is good practice and is specified in the ToR. The report notes the development of the matrix within the methodology, but does not clearly and completely explain how it was used for data processing, analysis, and interpretation. It is stated in the overall evaluation approach (section 1.3) that the evaluation questions were "unpacked and linked to corresponding assumptions, indicators, data sources and data collection methods and tools, which are indicated in the Evaluation Matrix" but does not elaborate on the specific operational steps or processes by which the evaluation matrix served as a tool for actual data processing, analysis, or interpretation; instead, referring to it primarily as a repository of methodological details found in an annex.
	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	There is a table of the limitations and mitigation measures in section 1.3.4, outlining specific limitations encountered during the evaluation process and the strategies employed to address them. However, more information on some mitigation measures would have been useful to strengthen their credibility. For example, the mitigation measure for outdated MICS data is not fully clear. In addition, the report may need to acknowledge that even the health facility assessment data may not be sufficiently geographically disaggregated to understand and attribute any changes to UNFPA interventions. Similarly, the mitigation measure for an overly qualitative evaluation approach should have stressed that triangulation across qualitative sources is the primary mechanism for corroborating data.
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	The evaluation report makes explicit reference to UNEG obligations and ethical principles as guiding frameworks for the evaluation in section 1.3.2. The report explicitly states that the evaluation was conducted in accordance with the "UNFPA Evaluation Policy, United Nations Evaluation Group Ethical Guidelines, Code of Conduct for Evaluation in the United Nations evaluation group (UNEG), and the United Nations Norms and Standards for evaluation in the United Nations System". It elaborates on adherence to accepted codes of conduct, specifically mentioning the seeking of verbal and written consent from respondents, maintaining confidentiality, avoiding bias, avoidance of harm and dignity and diversity. It also specifies that no data collection took place among those under 18, thus avoiding the need for more rigorous ethical approaches. While the report provides some degree of contextualisation for these ethical principles, it is very limited. For example, regarding the principle of "dignity and diversity" and "leaving no one behind," it mentions that "participants were selected by age, sex, PWDs and service function" and a "conscious effort was made to include PWD and people from vulnerable groups" during the mobilisation of FGD participants. While this demonstrates an intention towards the principle of inclusivity during sampling, the actual application of this in practice is not explained (a 'conscious effort' is highly subjective and could mean success or failure). There are few other details on how the principles informed the evaluation design or reporting decision. As a good practice it is recommended that a table with ethical considerations per evaluation phase is presented, showing how they were applied throughout the evaluation process. Information for each UNEG ethical principle in each evaluation phase may be presented in this table.
	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 noted above, the (brief) description of the ethical approaches is minimally adequate with respect to the various elements of the criterion. It does not go much further than explicitly stating that the evaluation was conducted in accordance with the UNFPA Evaluation Policy, UNEG Ethical Guidelines, UNEG Code of Conduct and the UN Norms and Standards for evaluation. Some explicit mention of the standards/codes/practices therein is made, e.g. • Seeking verbal and written consent from respondents. • Maintaining confidentiality. • Avoiding bias. • Avoidance of harm. • Dignity and diversity. While the report is explicit about these general principles and some specific safeguards, it does not detail contextualisation with the evaluation process nor compliance with specific codes for vulnerable groups beyond the age restriction for interviews or mention WHO standards of safe data collection on GBV. Importantly (and justifying the 'partial' rating), the data collection tools presented in Annex 5 make no reference in the preamble text to obtaining verbal or written consent - there is a brief reference to ensuring confidentiality in the preamble which should have also explicitly noted that consent was obtained.
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.	No	The evaluation TpR does explicitly note that "The use of innovative and context-adapted evaluation tools (including ICT) is encouraged" and the Design Report copies this text, stating in the general approach section (section 4.1.1) that "The use of innovative and context-adapted evaluation tools (including ICT) will be adopted" but there is no indication of their use in the evaluation itself, hence the rating is 'no'.
SECTION E:	EVALUATION FINDINGS (weight 25%)	42%	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	While the findings of the report (presented in Chapter 4), are reasonably well-structured and organised, directly corresponding to the evaluation questions and collectively responding to all of them, it is not fully and consistently organised per the mandated structure in the UNFPA Evaluation Handbook. For example EQ7 and EQ12 do not include the preamble summary. Further, linked to the feedback above under 6i, the evaluation report includes an additional evaluation question not present in the evaluation matrix - EQ13 - related to the coordination criterion. This criterion is no longer in use as of the 2024 version of UNFPA evaluation Handbook. The analytical approach and findings should have fully and consistently reflected this (i.e. in the evaluation matrix). Within each of the evaluation question subsections, individual findings are clearly numbered for ease of reference. However, each EQ restarts the finding numbering (so there are, as a result, multiple "Finding 1's", "Finding 2's" etc.). While there is a logical flow throughout the report, specific findings cannot be easily referred to. Further, the analysis is structured according to the evaluation questions and criteria, with the relevant assumptions in the evaluation matrix in Annex 1 are not linked or referred to, limiting transparency of the evidence base and requiring the reader to manually connect findings to the detailed methodological framework, thereby weakening the direct traceability of causal relationships and the analytical rigor within the main findings narrative. Finally, many of the findings are weakly linked to specific evidence/have quite limited levels of evidence against them. For example, the analysis under EQ12 is a single paragraph, noting that collaboration with or support to (both unspecified) various partners has led to improved capacity. This is not adequate evidence. Similar limitations on the quantity and quality of evidence can be found under EQs 5, 8, 10 which are all very short and present weak support to the findings.

	ii) Explicit use of the evaluator's theory of change, results chain, logical framework in the formulation of the findings.	Partially	This criterion is 'not rated' per the guidance. However, the theory of change is not used explicitly in the formulation of the findings of the evaluation - the specific subsection in the background sections of the report (section 3.2.2) provides a robust critique of the theory of change but does not reference this elsewhere in the findings/analysis. The findings section does, under the effectiveness criterion, provide an accounting of UNFPA achievements under both broader changes (outcomes) and specific UNFPA-supported programmatic activities (outputs), that are part of the country programme, and thus specifically uses the UNFPA country programme planned results (outputs and outcomes) across all of the programmatic areas. However, despite a commitment by the evaluation to use "evidence on observed results [to] confirm the chain of expected results in the theory of change" there is no substantive discussion, analysis or demonstration of explicit or implicit linkages between these outputs and the outcome - even from qualitative sources that can attest to the importance or value (and thus linking) of UNFPA support to the outcomes. Thus, the rating should be "partially".
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.	Partially	The evaluators have clearly undertaken data collection from a variety of sources, and have built their analyses and derived findings from a mix of qualitative and quantitative primary and secondary data. Notwithstanding issues of citing sources of data (discussed below), the findings do present data at both output and outcome levels, in line with UNFPA's results chain and ToC, particularly evident in the assessment of the effectiveness criterion (4.3, pages 34-42). For example, under "Ending Preventable Maternal Deaths," the report provides outcome-level data on the reduction of the Maternal Mortality Ratio (MMR) and institutional Maternal Mortality Ratio (IMMR). This is then directly linked to output-level interventions such as the procurement of tools like the Safe Delivery Application, training for midwives on emergency obstetric and newborn care, and maternal death audits. However, it is difficult to link some of these analyses with UNFPA sources - for example, for Adolescents and Youth, the report provides output achievements, such as the number of beneficiaries trained using the national out-of-school Life Skills Education manual - but the number reported in the analysis (4,598) does not match that cited in the UNFPA results framework in Annex 2 (4,337). Finally, while the report's methodology explicitly states that triangulation of data sources and methods enhanced the accuracy of findings, the degree to which this is transparently demonstrated for the findings varies. Many individual findings rely primarily on a single cited source for a particular data point, whether it's a survey statistic, a policy document review, or an interview excerpt. However, instances of <i>implicit</i> triangulation exist, where qualitative insights (e.g., from CSO partners) provide some context. For example, a quote from a Civil Society KII (page 46) directly challenging the effectiveness of GBV interventions. While the findings generally reflect diverse data types, a more explicit and consistent presentation of how multiple sources collectively corroborate or challenge specific findings would further strengthen the analytical rigor and direct evidence of triangulation.
	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.	No	As noted above, while the evaluation report presents analyses based on data that can be inferred as coming from a variety of qualitative and quantitative primary and secondary sources, the overall picture in terms of quality of the evidence itself is of poor performance. A significant issue is that the report only infrequently cites the sources of the evidence that is presented. In many, if not most, of the analyses under the various EQ findings, the evidence is simply presented as a list of statements without any discussion or reference to the source of the evidence (such as reports/documents, specific key informants or from rights-holders. One specific example (among many) is under EQ5, finding 1 (page 34, where the evaluators note that "<i>The implementation of the ending violence strategy has greatly improved than [sic] the previous strategy and has attracted interest from lawmakers and development partners.</i>" There is no evidence or data provided to support this assertion. There is an occasional recourse to a footnote that identifies the source of the evidence - e.g. on page 40, where a sub finding notes the source of the evidence as "KII and Documentation...". Another example, under EQ2, Finding 2 (page 31) notes that "Interviews...indicated that" - followed by an inactive in-line footnote/reference that suggests the statement was copied and pasted from another source. The report does present discussion on performance indicators specifically related to UNFPA's mandate, notably under effectiveness where changes in the maternal mortality rate and unmet need for FP are noted (with some data drawn from the MICS, but other referenced data that cited by UNFPA in annual reporting, with no indication of the underlying source of this). However, these references to actual evidence and sources are very much the exception. For example, under the Efficiency criterion (EQ8), finding 2 simply lists a number of statements ("The CO was..."; "The CO has...") that are not evidence, but unvalidated findings. There is no evidence presented that can demonstrate how or why the evaluators conclude that (for example), the CO has "strong financial systems and controls", as the "analysis" contends. Most of the findings are presented, at least in part, like this. This is a central flaw of the majority of the analysis - statements (which are a mix of both positive performance and areas which can be improved - but without underlying evidence it is impossible to determine how balanced the mix is) are presented as facts, with little recourse to supporting evidence that has been collected, analysed and referenced. Thus, as findings are largely unsupported by clearly documented evidence and thus the credibility of the analysis cannot be
	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	The evaluators clearly aim to identify causal factors through use of a contribution analysis approach noted in the methods. The report does make some attempts to explain why interventions performed well or faced challenges. For example, under EQ6a, finding 1 (page 35) the decline in the MMR/IMMR is attributed to output-level interventions such as the procurement of medical equipment ("gadgets") and training for midwives. Conversely, the report identifies resource constraints as a reason why UNFPA could not adequately meet "some" [unspecified] demands during humanitarian responses (EQ2, Finding 2 (page 31)). Similarly, delays in disbursement of non-core funds to implementing partners are cited as affecting the timeliness of implementation. The "why" is also addressed in places for negative findings, such as under EQ6a, Finding 1, where the increase in unmet need for family planning was linked to COVID-19 lockdowns affecting commodity availability. These positive examples are primarily within the analysis under the effectiveness criterion (section 4.3), which links activities/outputs to their intended contributions to broader outcomes, aligning with the ToC. However, evidence that goes beyond mere association or alignment of UNFPA outputs with documented changes in outcomes is lacking - such as qualitative testimony from key informants or rights-holders that activity X (supported by UNFPA) was a contributory factor in outcome Y - such an approach is generally-held practice among mixed-methods evaluations such as this. Further, the depth of the existing causal analysis and logical chain progression is limited to a few findings. Much analysis remain at a general level rather than providing granular detail on the precise causal mechanisms for the results cited. For instance, the decrease in correct HIV prevention knowledge among young people (Finding 2, EQ6b) is presented without a discussion of the causal factors contributing to this decline, beyond general implications. As noted above, other findings simply state what UNFPA has done, or present summary statements of achievement or judgements without recourse to any evidence that can serve to underpin them and thus ascribe causality.
Question 13.	Does the evaluation assess and use the intervention's Results Based Management elements?		

	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.	Partially	The Design Report does incorporate a commitment to assess the M&E system of the CP, specifically EQ8, Assumption 4 of the Evaluation Matrix, which commits the evaluators to validate the extent to which there was a "Robust M&E system in place and efficiently utilised". The report provides some assessment of the adequacy of the UNFPA Eswatini planning, monitoring, and reporting system, specifically highlighting areas of strength and notable weaknesses concerning its ability to support decision-making. In terms of planning and the results framework, the ToC is assessed (in Section 3), with the evaluators noting the presence of "measurable indicators at both outcome and output levels" with "realistic" baselines and targets, demonstrating a sound vertical logic linking activities to intended outcomes. The analysis refers to the CP's M&E framework, but the language on page 23 (paragraphs 3 & 4) is contradictory: it both claims the ToC has a "clearly stated timeframe for achieving the outcome indicators" and identifies the "lack of a clearly stated timeframe" as a gap. Regarding CO monitoring systems, the evaluators do provide some details of the M&E infrastructure and some related activities under EQ8, but this is quite cursory. There is reasonable documentation of UNFPA support to population data (e.g. support to improving real-time demographic data and data accessibility for policymakers to strengthen evidence-based planning), but these are programmatic, and not related to the M&E systems. The report does identify a reliance on outdated national surveys, specifically the 2021 MICS at the start of the CP, as a challenge to the accuracy of reported outcomes beyond 2021 and 2022 and thus undermined the completeness and appropriateness of the results framework. However, this is again less about the M&E systems and more about programme results. While the report does make a strategic recommendation (#5) for a "more robust M&E system" to ensure "continuous data availability", thus acknowledging deficiencies in the CO's monitoring and reporting capabilities, this is not supported by adequate analysis and evidence in the findings sections.
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 summative statements that respond to the evaluation questions.	Partially	The conclusions presented in Chapter 5 of the report are reasonably clearly-formulated and strive to present unbiased judgments of the CP performance, aligning with the various findings. The conclusions are also categorised into strategic and programmatic levels and indicate their associated evaluation criterion, EQs and the resulting recommendation(s). The strategic conclusions are structured with a concise judgment followed by a more detailed narrative, although the majority of the programme conclusions are presented as 1-2 short sentences only, which is not good practice (conclusions should include a more comprehensive narrative/supporting paragraph that explains and justifies the judgement). In addition, some conclusions simply restate the findings (strategic conclusions 3 and 4) in bullet point format, which is also not good practice. While some finding-specific information can be provided for clarity, it should not be at this level of detail. The conclusions also do not identify the specific finding numbers, making it more challenging to trace the direct evidence for each. Further, while the organisation of the conclusions under programme areas is acceptable, the numbering convention of the programmatic conclusions is poor - they restart for each programme area, thus there are 5x Conclusion 1s, 5x Conclusion 2s, 2x Conclusion 3s. This is confusing - all conclusions (and indeed findings, as noted above) should be uniquely numbered, to assist cross-referencing by reviewers and those tasked with acting on the conclusions and related recommendations.
	ii Conclusions are well substantiated and derived from findings and add deeper insight and analysis beyond the findings.	Partially	The conclusions present a mix in terms of the quality of the substantiation of their judgments of CP performance. While the conclusions do not introduce any new information, and link clearly back to specific EQs from which the evidence is derived, there are also some instances of inadequate/inaccurate attribution of origins. An example of positive conclusions include Strategic Conclusion 3, which states that the CP "partially achieved" expected results, with the absence of updated outcome-level data a major constraint on more accurate attribution. This is substantiated by Finding 1 under EQ6a, which presents evidence to corroborate this conclusion. Conversely, conclusions that could be better corroborated or substantiated include Strategic Conclusion 7, which highlights UNFPA's high value and leadership among partners, attributes its origin to EQ9 (Sustainability), EQ11 (Connectedness), and EQ12 (Connectedness). While UNFPA's role is indeed discussed throughout, more direct evidence for its comparative strengths and influence is found under EQ13 (UNCT Coordination), but is not included. Similarly, Programmatic Conclusion 3 under GEWE discusses the CO's "humanitarian positioning" and its value, but cites EQ 10 (noted as "Sustainability" in the conclusions, but in the main report is under "Coverage" - a clear inconsistency) as its origin. EQ10 specifically covers coverage of humanitarian interventions, whereas the discussion on the value and niche of humanitarian response is also reflected in the findings related to EQ11 (Connectedness) and EQ2 (Relevance) on page 31. These inconsistencies and limited reference to the entirety of the analysis undermine the traceability of the conclusions back to their specific evidence base.
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 have not been explicitly requested in either the ToR or committed to in the Design Report. Thus, there is no specific section related to these to be rated.
	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 have not been explicitly requested in either the ToR or committed to in the Design Report. Thus, there is no specific section related to these to be rated.
SECTION G:	EVALUATION RECOMMENDATIONS (weight 15%)	63%	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.	Partially	The recommendations are well-formulated and outline specific, actionable steps for the future CP. They are clearly articulated and categorised into strategic and programmatic groupings, adding to clarity and usability, although the restarting of the numbering between the two groupings (as with the findings and conclusions) is not good practice. Many recommendations demonstrate a clear and logical connection to the conclusions, building upon identified strengths, weaknesses or gaps. For example, Programmatic Recommendation 5 in relation to improving disability inclusion directly responds to programmatic conclusion 2 under GEWE which notes that PWDs are underserved. However, while the recommendations are often sound in substance, there are some inconsistencies and gaps in explicitly tracing some back to their cited conclusions. For example, programme recommendation 2, relating to SRHR refers to conclusion 5 which is related to efficiency and resources. Further, the section has inconsistent numerical referencing for programmatic conclusions (e.g., citing "C8, C9, C10" for GEWE recommendations (page 61) while the conclusions for that section are numbered 1, 2, and 3, which is a barrier to following the precise logical progression from conclusion to recommendation. These issues, while not entirely undermining the utility of the recommendations, weaken the report's analytical rigor and compliance with the assessment criterion.
	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 largely useful and actionable for primary intended users due to their clear formulation and the inclusion of specific guidance for implementation as well as the separation into strategic and programmatic groupings. Most recommendations are written in clear, unambiguous language and outline concrete actions to be taken, with each recommendation typically including designated recipient(s) (responsible actors) and operational implications that detail the suggested actions and technical or financial considerations, which is a very useful practice. This structured approach enhances their practical utility for UNFPA and related stakeholders in shaping the next programme cycle. Another weakness of the conclusions is that the intended implementers of the recommendations are not clear. They only broadly note UNFPA CO or IP, rather than identifying the specific teams or units within the CO. This also does not facilitate the drafting of the management response.
	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	In the preamble to the recommendations chapter, the evaluators note that "feedback from individuals and groups validated the recommendations", while the methods section (1.3.2) notes that the ERG (which was composed of governmental and NGO representatives) provided validation of the draft recommendations, which process also included "key informants and participants of FGDs, CO programme staff, IPs, Evaluation Manager and ERG" (page 10). Although the process for developing the recommendations is explained, it would be better to include all of this information at the beginning of the recommendations section, rather than scattering the information across the report.

	iv Recommendations are clearly articulated and prioritized based on their importance, urgency, and potential impact.	Partially	Most of the recommendations are indeed prioritised according to high, medium or low priority, as per the guidance in the UNFPA Evaluation Handbook - but not all. Some recommendations (i.e. programmatic recommendations 1, 3 and 7) lack this prioritisation. Further some have a time-horizon associated with them (contrary to guidance) whereas others do not. The evaluators should review the list of recommendations carefully and ensure complete consistency internally and compliance with the UNFPA Evaluation Handbook in this regard.
SECTION H:	REPORT STRUCTURE AND PRESENTATION (weight 5%)	83%	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 evaluation, 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	Most, but not all, of these components are present and correct. An omission is the absence of the name of the Evaluation Manager and there is also an issue with the table of contents - it is present, but omits the list of annexes - it skips from the recommendations pages to two of the subheadings of Annex #6 (the ToR). This is a clear error in the formulation of the table of contents that requires correction through updating of the automatically-generated TOC or manual revision. As a minor editing issue, however, it does not affect the 'yes' rating.
	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	All of the required annexes are present (although not listed in the Table of Contents, there is a list of annexes at the start of the section, page 64; also note that this includes the programme ToC as Annex 7, but this is not present - the ToC is included in the main report).
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	While the structure of the report is good, with logical flow of chapters/sections, subsections etc. (per the UNFPA Evaluation Handbook), some of the formatting could be improved, specifically: • Some subheadings start at the end of pages (e.g. Acronyms/Abbreviations section, Operational Recommendations) • Some acronyms are incorrect (e.g. GBV-MIS should be GBVIMS), explained multiple times (e.g. CP on pages 1 and 4, SRHR is explained multiple times across the document) and the full list of acronyms in the preamble should be formatted to fit on a single page for ease of reference/readability • There is a mix of fonts/styles in places (e.g. footnotes) and inconsistent line/paragraph spacing (e.g. Conclusion 3) • Some of the writing or formulation of language is poor, e.g. incomplete sentence at the end of Conclusion 3 (pg. 53 para 1). • A wide range of language errors/misspellings and formatting errors that detract from the professionalism of the report • A random, unattributed, footnote at the top of page 40. • Use of inappropriate terms (e.g. "gadgets" procured for safe deliveries cited on page 35). • Annex 7 (in the list of annexes on page 64) is not present.
	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>	Yes	The main report (including the executive summary) is 64 pages - this is within the 70-page limit mandated by the criterion.
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.	Yes	The report is generally easy to understand, with acceptable writing/language quality and no major issues noted. However, with respect to grammar, spelling and punctuation (i.e. the quality of the writing), there are a range of mistakes in formulations that detract from key elements of the report. Some examples are provided in 18i above, and another significant issue is the mistake with the list of evaluation questions in Table 1, which has been promulgated across different sections (and originates in the Design Report). Other errors in the numbering or referencing of recommendations or conclusions, or the missing priority levels in some recommendations, are significant given the importance of these elements of the report (although do not specifically relate to the overall understanding of the report, so do not affect the rating). Overall, the report requires a thorough review to catch these errors to be addressed for a final version.
	ii Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labeled, and referenced in text.	Partially	The report has multiple types of visual aids, i.e. figures (maps, process diagrams, graphs) and tables, to present information within the main body of the report (excluding annexes), enhancing readability and data presentation. These visuals are generally labelled with titles/captions and numbers, and where applicable, include legends and are typically well-formatted with appropriate size and resolution for readability. However, there are some inconsistencies and errors that undermine their utility and the overall professionalism of the report. For example, there are no data visualizations in the findings section, even though this could have been possible (for example a table on indicator status to show progress or lack thereof). The visual aids are focused on the background and methodology sections only, although visualizations would be helpful to easily present complex evaluation findings. Further, two entirely different visuals are both labelled as "Figure 3" (pages 15 and 24). This creates confusion when referencing these figures within the text (The fig. 3 on page 15 is also incorrectly captioned as "Total Fertility Rare ", instead of "Rate"). Furthermore, the table labelled "Table 1: Evaluation criteria and evaluation questions" on page 7 is referred to as "table 2" in the preceding text. The ToC diagram on page 21 is unlabelled as a figure, only being introduced generically as "the figure below". Regarding readability and integration, while most tables are clear, the unlabelled ToC diagram on page 21 is overcrowded with very small text and numerous overlapping lines, rendering it exceptionally difficult to read and decipher. Moreover, the map on page ii (Figure 1), while titled to include "UNFPA Programme areas," lacks a corresponding legend to identify these areas, limiting its utility. Finally, some of the charts would have benefited from either improved formatting, or being presented in an alternative way. For example, figure 3 (page 24) presents UNFPA 7th CP financial/expenditure data, but the information presented is unclear and not well-linked to the financial data in the narrative.
SECTION I:	CROSS-CUTTING ISSUES (weight 10%)	61%	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)?		

	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's data collection methods were designed with an emphasis on capturing the voices and perspectives of a wide range of stakeholders, particularly rights-holders and marginalised groups. The overall evaluation approach was explicitly framed as "inclusive, transparent and participatory," involving a broad spectrum of partners and stakeholders, including "rights-holders (notably women, adolescents and youth)," with attention paid to vulnerable and marginalised groups such as young people and women with disabilities (noted in the Evaluation Approach, section 1.3). As discussed above, the evaluators specify that a "conscious effort" was made to include PWD and people from vulnerable groups for FGDs, though the precise definition of this (somewhat ambiguous) phrase is unclear. The evaluation questions, matrix and data collection tools based on these reflected inquiry on the needs of vulnerable and marginalised groups (e.g. young people and women with disabilities). The data collection methods reflected the inclusive design: KIs were conducted with individuals from diverse organisations, including various government ministries, CSOs, UN agencies, and academic institutions, ensuring a broad institutional perspective. Importantly, the FGDs directly engaged 50 beneficiaries, disaggregated by age and sex, and explicitly included adolescents, young mothers, midwives, and peer educators to ensure direct input from those impacted by UNFPA supported activities. The ERG itself had wide-ranging representatives from governmental and non-governmental organisations, including those representing persons with disabilities, with only a single UNFPA member, which is an outlier among similar evaluations, which normally have high UNFPA representation. However, none of the data collection tools (Annex 5) include any information that would indicate that they were designed to collect disaggregated data, despite claims in the methodology of a focus on this. Further, not all the elements of the evaluation matrix address cross-cutting issues systematically, for example the indicators of the effectiveness questions should have integrated cross-cutting issues more deeply, notably around disability inclusion (part of EQ7) but they are not referenced in the evaluation matrix assumptions or indicators.
	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, as detailed in Table 1 (pages 7 & 8) and the Evaluation Matrix, demonstrate a comprehensive and systematic inclusion and addressing of cross-cutting issues such as human rights-based approaches, gender equality, disability inclusion, and the LNOB principle. As noted above, there are copy/pasting errors in the list of EQs in Table 1 and also the Evaluation Matrix (which also omits EQ13), so reviewers should refer to the questions in the main report. Looking at these, out of the 13 evaluation questions listed, 8 directly reference the integration of one or more of these cross-cutting issues (1-4, 6b, 6c, 7, 9, 10, with three specific questions (3, 7, 10) specifically focused on these dimensions (EQ7 notes environmental sustainability)).
	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	Firstly, while the design and intent for broad inclusion of different stakeholders (including rights-holders via FGDs) are commendable, the explicit reflection of all these diverse voices in the findings is limited, with minimal detail on how these voices or unique needs of these particular groups were captured or directly influenced the conclusions. Further, quotes or narratives from rights-holders are not well represented in the findings (quotes from duty-bearers via KIs are used only). Several findings present disaggregated secondary quantitative data, for example the analysis of unmet need for family planning (page 36); % of young women/men with correct HIV knowledge (page 38); youth reached with Life Skills Education efforts (page 38). For PWDs, the evaluation provides some specific data disaggregation (page 44). Although the evaluation design is quite robust in referencing specific focus on different population groups, the findings are quite general and focused on UNFPA outputs/activities and lack more specific disaggregated analysis - for example, while the FGDs (with 50 rights-holders) were reported to be disaggregated by age and sex (Noted in section 1.3.2, page 10), neither the main body of the report nor Annex 4 (List of Stakeholders Consulted) provides explicit overall gender disaggregation for the 46 KIs or the collective group of persons listed in the annex. This was exacerbated by the weak UNFPA M&E system (which did not produce sufficient disaggregated data - even though the evaluation claims that IPs were requested to report disaggregated data), but the evaluation did not compensate for this by including disaggregated data during primary data collection, as per the commitment in the design. Aside from this issue, findings do present some disaggregation - both in quantitative terms and (more commonly) in discussion of activities, outputs or results. The report frequently discusses the CP's alignment with the needs of diverse populations, including vulnerable and marginalised groups (e.g. young people and women with disabilities, etc.). For example, the evaluation notes that during the CP design, national data on vulnerability was used to prioritise interventions and target beneficiaries and geographical locations, and UNFPA guided IPs to deliver interventions to reach both sexes, target young people and to leave no one behind (EQ7, page 42). Further, this EQ presents specific analysis on integration of these groups, for example the requirement for IPs to report disaggregated data and that overall UNFPA reporting aimed for data to be disaggregated by gender and age "to some extent where relevant" (page 43).
	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 findings, in conjunction with its evaluation questions (Table 1), do indicate some, but inconsistent application of an intersectional lens in data analysis. More than half of the evaluation questions for which the integration of cross-cutting issues is feasible explicitly reference LNOB considerations, and the findings disaggregate both primary and secondary data by population groups and assess differential results. This is especially notable in the stakeholder analysis preceding the findings, for example, in Chapter 3 (page 14), the evaluators present disaggregated data (from the MICS) to highlight greater needs among unmarried women, adolescents and the poorest, rural, and least educated young women. However, this is in the preamble to the findings section, and those related to this (EQ6a, page 36) have much more limited analysis of primary (or secondary) data limiting the ability of UNFPA to see how its work has impacted the most vulnerable groups. Narratively (i.e. not quantitatively) the evaluation does consistently discuss UNFPA reach and impact on vulnerable groups such as PWDs, for example comparing their access to services and amenities against national averages (on page 43), but this is very specific and not related to UNFPA mandate areas. A positive example is under the Relevance criterion (EQ1) which notes that an autism baseline survey revealed specific vulnerabilities of this group to GBV and accessing SRHR information (page 29), although it does not specify by whom or when this report was prepared, nor explore the details any further. Overall, there are gaps in the quantity and quality of the analysis (and the evaluators do note that data limitations (incomplete or out-of-date data) were a factor in the depth of the analysis) evidence from the evaluation questions and findings supports the 'partially' rating.
	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	As discussed above, the analysis/findings make some efforts to meet the commitments set out in the evaluation questions around addressing LNOB and equality/vulnerability considerations, but it is limited, and reflects the overall criticism that the quality and depth of the analysis is not robust. The findings do consistently discuss interventions and outcomes related to women, adolescents, youth, and PWDs, possibly implying that their input shaped the analysis. For example, the report details specific policy briefs and initiatives for PWDs, and extensive youth empowerment programmes focusing on life skills, SRHR, and entrepreneurship, implicitly reflecting their needs and participation. However, as noted above, the findings section has no 'voices' from rights-holders, and provides minimal specific detail on how the voices or unique needs of these particular groups were captured or directly influenced the conclusions. The report only generally notes that "certain populations remained underserved, such as PWDs and national minorities" and that the reach for PWDs was "not optimal due to limited catchment". While the discussion under EQ10 notes that the SRH needs of "displaced people" were addressed through partners such as IOM, specific details on direct input from these (or other) populations are not prominently reflected in the findings. The conclusions and recommendations do include some consideration of these cross cutting issues, with specific strategic and programmatic recommendations covering PWDs, marginalised groups, and vulnerabilities, for example, highlighting that PWDs remain an underserved population, leading to recommendations (Strategic Rec 1; Programmatic Rec 1) to prioritise the inclusion of marginalised populations in the design of the next CP. Overall, however, the conclusions, in particular, while referencing disability in several places, do not discuss other dimensions to any extent, if at all (e.g. 'underserved populations' are mentioned only in passing as part of strategic conclusion 7, as is environmental sustainability, which is merely mentioned in the report (EQ7) and equally briefly noted in programmatic conclusion 1).

	vi) Inclusion of young people in the evaluation team and/or Reference Group [N/A if not requested in ToR]	Partially	While the recruitment of a young and emerging evaluator was specified in the ToR, and noted in the Design Report as being a priority for the evaluation team, there is no indication in the evaluation report itself that such an evaluator was recruited - the list of (three) evaluators in the cover pages has no such designation. The ERG did, however, include a representation from Eswatini National Youth Council, which partially satisfies the criterion.
Question 21.	Does the evaluation meet UN SWAP evaluation performance indicators? <i>Note: this question will be rated according to UN SWAP standards with detail provided below</i>	6	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.	Fully integrated	The design of the evaluation report demonstrates a comprehensive and explicit integration of human rights and gender equality, along with related cross-cutting issues such as vulnerability, disability inclusion, and LNOB, across its objectives, scope, evaluation framework, and specific questions. This integration is evident in both dedicated aspects and through mainstreaming within other evaluation criteria. As noted under 20ii above, out of the 13 evaluation questions in the report, 8 directly reference the integration of one or more of these cross-cutting issues (1-4, 6b, 6c, 7, 9, 10, with three specific questions (3, 7, 10) specifically focused on GEEW dimensions. EQ7 is specifically focused on the integration of gender equality and human rights.
ii)	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated	The evaluation design also has an explicit integration of GEEW considerations, for example with data collection designed to be disaggregated by sex, age, disability (although the "conscious effort" to include PWDs and individuals from vulnerable groups in FGDs was not fully clarified). The evaluators noted the intention of the research to be "highly participatory" from its inception with a well-rounded ERG with representation from gender and SRH specialists. FGDs specifically included separate discussions for male and female participants to ensure a comfortable environment for sharing experiences and perspectives. However, none of the data collection tools (Annex 5) include any information that would indicate that they were designed to collect disaggregated data. While the indicators in the evaluation matrix do a good job in addressing GEWE issues, they could have been more comprehensively integrated across all the EQs, notably under the relevance and effectiveness assumptions, where GEWE considerations are of greatest importance. Further, the sampling frame (although not extensively described) was designed to address the diversity of stakeholders affected by the intervention, with a particular focus on the most vulnerable. However, the level of detail on the sampling process is so limited that it is impossible to determine the full extent to which these considerations were integrated. Overall, while there is clear reference to the integration of GEWE in the methods, the detailed description of the evaluation design do not fully display this consideration.
iii)	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Partially integrated	Notwithstanding the limitations of the overall analysis (poor recourse to documented evidence, limited depth and granularity of analysis) as discussed and rated under question 12 above, the evaluation report demonstrates a satisfactorily integrated approach to GEWE considerations across the findings, conclusions and recommendations. The findings do include some analyses/sub findings on specific GEWE considerations, but (as with much of the analysis) it is cursory and does not present much in the way of verifiable evidence against which the findings can be compared. This is particularly important for UNFPA where progress has not been made, so the absence of more sophisticated analysis of causal aspects is a gap for future programme design. For example, the report acknowledges that despite efforts in GBV response, GBV cases had increased - this piece of evidence (reported by an IP key informant) should have been significantly unpacked and triangulated with any other data, and some analysis of the causes presented, to assist UNFPA in tailoring its work to address the issue (if validated). The conclusions and recommendations do extensively, but superficially, address the GEWE dimension. For the conclusions, there are several references to gender (e.g. strategic conclusions 2, 3, 4), but no specific conclusion related to gender equality. Further, Conclusion 4 notes that the CO "implemented gender transformative approaches", but the finding (EQ6a, finding 2) has no evidence or even discussion of the basis for this (the analysis under the finding is focused entirely on disability). The recommendations include some positive gender-related elements, for example under Strategic Recommendation 2 which advocates for "human rights, gender and disability inclusion mainstreaming in interventions" and targeting "underserved populations and hard-to-reach communities". Programmatic Recommendation 3 specifically calls for the "roll out transformative approaches in GBV" and "advocat[ing] for gender transformative approaches for gender equality and ending GBV". Similarly, Programmatic Recommendation 5 mandates the systematic incorporation of "disability inclusion in all areas of operation SRH, AY and GBV". While these are laudable, the absence of more granular analysis means that their value might be of limited significance to UNFPA in terms of operationalising them.

SWAP Rating Guidance

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 digitalisation
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