UNFPA Evaluation Quality Assessment Grid

Version: May 2024

REPORT RATING SUMMARY			
Overall Rating		90%	Highly Satisfactory
● ● ● ● ● Excellent		5	
● ● ● ● Highly Satisfactory		4	
● ● ● – Satisfactory		3	
● ● – – Fair		2	
● – – – Unsatisfactory		1	

REPORT DETAILS	
Title of the evaluation report	UNFPA Yemen. Independent Country Programme Evaluation 2015-2024
Region	ASR
Country	Yemen
Year of report	2024
Business Unit/programme country *(managing evaluation)*	UNFPA Independent Evaluation Office
Date of assessment review (dd/mmm/yyyy)	February 27, 2025
Name of assessment review firm	IOD PARC
CLASSIFICATION OF EVALUATION REPORT	
Primary SDG(s) covered *(list provided below)*	**3, 5, 10, 16**
UNFPA Strategic Plan areas covered *(lists provided below)*	
Three transformative results	All
Six outputs	All
Six accelerators	All
Organizational effectiveness and efficiency	No
Humanitarian evaluation	Yes
Evaluation evaluand *(e.g. country programme/intervention/policy/thematic area)*	Country Programme
Evaluation type *(e.g. formative, summative, developmental)*	Summative and formative
Geographic scope *(e.g. global, regional, national)*	National

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.*

This is a satisfactory evaluation report of UNFPA Yemen's Independent Country Programme Evaluation, 2025-2024. The key strengths and weaknesses of the Evaluation Report are as follows:

• **Section A: Executive Summary** – The Executive Summary is clearly written and well-presented standalone document that will be useful for decision making. However, it would benefit from the inclusion of a summary of the country context, which is currently missing. Evaluators can also consider expanding on the methodology, which primarily references the Evaluation Matrix and Theory of Change.
• **Section B: Background** – There is clear description of the Country Programme and the context in which it operates. Where it can improve on is its limited description of key stakeholders, aside from their identification as primary and secondary users of the evaluation. However, there is a stakeholder map in the Inception Report (Annex 4a), which is quite comprehensive and evaluators can consider including it in the main report. Similarly, it is unclear who the rights holders are. This may be worth considering including in the report to better understand the sampling strategy for the evaluation.
• **Section C: Evaluation Purpose, Objectives and Scope** – The purpose, objectives and scope of the evaluation are clearly defined. Evaluators have further clarified the temporal scope, which is justified and appropriate for this evaluation.
• **Section D: Evaluation Design and Methodology** – The evaluation report provides a very clear design as seen in the evaluation team's reconstruction of the Theory of Change (ToC) and the clarity of the evaluation matrix, making it easy to trace in the findings section. It is commendable that the evaluation team have further refined the evaluation questions during the design stage. The data collection methods are clearly described. What can be improved is further clarity on the sampling strategy and ethical considerations and processes (please see below under recommendations for more detail).
• **Section E: Evaluation Findings** – The findings are clearly and systematically organised, which each evaluation question addressed in the order outlined in the evaluation matrix (Appendix 5). The evaluators have also developed a 'Strength of Evidence' rubric, with triangulation informing the categorisation of evidence as 'strong,' 'moderate,' or 'weak.' This approach is a good practice that other evaluations could replicate to enhance transparency regarding the strength of evidence. What can be improved is the length of the findings section, which is 70 pages in length (please see below for recommendations).
• **Section F: Evaluation Conclusions** – The evaluation conclusions are clearly formulated with clear referencing provided to the findings for each of the five conclusions. Furthermore, they present a synthesis of key points drawn and also provide a deeper insight and analysis that goes beyond the findings.
• **Section G: Evaluation Recommendations** – The evaluation recommendations are clearly formulated and are clearly and logically derived from the findings and conclusions. Each recommendation is supported by suggested key actions, with each key action given a priority rating and a timeline.

Suggestions for future evaluators: *The rater will identify key suggestions to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, examples will be cited to assist evaluation managers in overseeing future evaluations.*

Recommendations
• **Section A: Executive Summary** – The Executive Summary would benefit from the inclusion of a summary of the country context. Evaluators can also consider expanding on the methodology as it is currently quite brief and does not provide sufficient information to understand how the evaluation was conducted.
• **Section B: Background** – While the stakeholder map provided in the Annex of the Inception Report (Annex 4a) is quite comprehensive, it would be helpful to include a column that outlines their roles and key areas of responsibility. This will contribute to an understanding of their interest in the intervention and the linkages between the different stakeholders.
• **Section D: Evaluation Design and Methodology** – On the stakeholder sampling, while there is a description provided in Section 2, it would benefit from a clear description of the criteria used for each of the data collection methods. It currently provides an overarching approach but it is unclear what criteria were used and at what point. On ethical considerations, while there is explicit reference to UNEG norms and standards, it would be beneficial to contextualise them for this evaluation. In addition, evaluators can also be encouraged to include, in the Annex, the protocols used to obtain consent as well as safeguarding mechanisms to better understand what was in place.
• **Section E: Evaluation Findings** – This recommendation is linked to Section H as it is linked to the overall length of the report (i.e. 102 pages excluding the Executive Summary). Overall, while the findings are well-supported by an appropriate level of evidence, it has resulted in a findings section that is 70 pages in length. Evaluators may benefit from stepping back to consolidate some findings and/or reconsider the level of evidence or detail required.

SECTION RATINGS

SECTION A:	EXECUTIVE SUMMARY (weight 5%)	83%	Comments on Rating
Question 1.	Can the executive summary inform decision-making?		
i	Is a clear, standalone document useful for informing decision making, (a minimum of 5 pages, up to a maximum of 7 pages).	Yes	The Executive Summary is a 5.5 pages in length. It is a clear standalone document that will be useful for informing decision making.
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	It includes nearly all necessary components of the evaluation report. It includes: (1) Overview intervention (p.10); (2) Evaluation Purpose, objectives and intended audience (p.10); (3) Scope and evaluation methodology (p.10-11); (4) Key Findings (p.11-13); (5) Main conclusions (p.13-15); and (6) Key recommendations (p.15). The Executive Summary would benefit from the inclusion of a summary of the country context, which is currently missing. Evaluators can also consider expanding on the methodology, which primarily references the Evaluation Matrix and Theory of Change, to fully meet this criterion.
iii	Includes all significant information in a concise yet clear manner to understand the theme, intervention, programme, project and the evaluation.	Yes	The Executive Summary is written in a clear and concise manner. It includes all significant information needed to understand the findings. While the background section does not provide details about the intervention, the findings include relevant descriptions of its key aspects needed to understand of both the programme and the evaluation.
SECTION B:	BACKGROUND (weight 5%)	80%	Comments on Rating
Question 2.	Is the evaluand (i.e. intervention/policy/thematic area etc. that is to be evaluated) and context of the evaluation clearly described?		
i	Clear description of the evaluand (e.g. intervention), including: geographic coverage, implementation period, main partners, cost/budget, and implementation status.	Yes	There is a clear description of the Country Programme in Section 3 'Background and Context' (p.30-34). In this section, the Country Programme Document (CPD) for the years covered by this evaluation, i.e. 2012-15/22 and 2022-24, is outlined on pages 32-34. As this is a Country Programme Evaluation (CPE), it covers the entirety of the country where UNFPA has implemented its programmatic interventions (p.17). While the background does not describe the programmatic areas of concentration, this is elaborated on in the findings under EQ5 (p.71), which assesses geographic coverage. Similarly, UNFPA's partners in Yemen are not mentioned in the background section aside from identifying them as evaluation users. However, the findings section provides information on partners such as 'UNFPA Yemen partners by category' (p.101), 'by type' (p.102), and by funding contributions (p.101). The budget for 2015-22 is provided on p.33, which includes the distribution of funding across programme areas.
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	Yes	There is a clear description of the context in Yemen needed to understand the country programme. This is provided in Section 3 (p.31 onward). It describes the Yemeni crisis, which began between 2011 and 2012, and the implications this has had on the humanitarian crisis and other related emergencies. The evaluation report also provides the context for the Country Programme in relation to reproductive health and rights and cross-cutting issues including gender equality and gender-based violence. It also discusses the vulnerability of Yemen's adolescents, particularly adolescent girls where a third of girls are married before the age of 18, nearly 20 per cent undergo female genital mutilation and other factors that compound their vulnerability.
iii	Linkages drawn between the evaluand and the ICPD benchmarks and SDGs relevant targets and indicators.	Yes	The linkages to the ICPD benchmarks are described on p.32 and SDGs on p.30-31. The Evaluation Report presents the status of Yemen in relation to SDG 3 and 5 (see Figure 4) using data from the SDG index. As good practice, it would be good for future evaluations to consider including data from SDG performance index in future evaluations.
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).	Partially	There is limited description of key stakeholders in the background section of the report. The implementing partners, development partners and duty bearers are mainly identified as primary and secondary users of the evaluation. However, there is a stakeholder map in Annex 4a of the Inception Report. It would be helpful if this was included in the Annex of the final evaluation report as well. This is quite comprehensive and serves as a good foundation for the sampling strategy. Similarly, rights-holders are only identified as key users but it is unclear who they are. It would be beneficial for evaluators to elaborate who the rights holders are that are the focus of this evaluation. While the stakeholder map is quite comprehensive, it would be helpful to include a column in the Table provided in Annex 4a of the IR that outlines their roles and key areas of responsibility. This will contribute to an understanding of their interest in the intervention and the linkages between the different stakeholders. Alternatively, evaluators can consider presenting the information in the form of a map rather than a table, to make the linkages more explicit.
	ii	Stakeholders are analysed to understand their specific rights, duties, needs, interests, concerns, and potential impact on the evaluand.	Partially	As mentioned in Q3i (above), it would be beneficial if the Table in Annex 4a included a column that outlines the specific duties of the duty bearers. It would also be helpful to elaborate on who the rights holders are, what their needs, interests and concerns are.
SECTION C:		**EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)**	100%	**Comments on Rating**
Question 4.		**Is the purpose of the evaluation clearly described?**		
	i	Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.	Yes	The purpose of the evaluation is clearly defined i.e. to demonstrate accountability, inform decision making, and contribute to lessons learned for progress towards UNFPA commitments to ICPD (p.17). While it does not explicitly describe why it is needed at this point in time, the high-level objectives provide a clear rationale of the time frame of the evaluation (2015-2024) in order to 'inform the design of the next programme cycle' (p.17). The intended use is clear based on the purpose statement and high-level objectives. The primary and secondary users of the evaluation are identified on p.18.
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).	Yes	There is a clear and complete description of the objectives of the evaluation provided in Section 1 (p.17). The objectives largely remain the same as the ones provided in the ToR. This is with the exception of the timeframe, which the report provides an explanation for in the footnote and under the temporal scope (p.18). This is to include 'programmatic work undertaken during 2023 (and from early to mid-2024, when data collection was completed) to avoid imposing an artificial a[and impractical] division on the data collection and analysis' (p.18). It also clarifies that the CPE covers these activities but is not to be considered a formal evaluation of the 2023-24 Sixth Country Programme (p.17).
	ii	Clear and relevant description of the scope (e.g. thematic, geographic, and temporal) of the evaluation, covering what will and will not be covered, as well as, if applicable, the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political, humanitarian or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Yes	There is a clear and relevant description of the scope, which includes geographic, thematic and temporal scope (p.17-18). There are no changes made to the geographic scope as the evaluation will continue to focus on governorates and districts where UNFPA has implemented its programmes. The thematic scope also remains broadly the same with the Evaluation Report further clarifying that the evaluation will look at interventions in both development and humanitarian settings. The main change is the temporal scope, which is mentioned above under Q5i. This is to include activities undertaken during 2023 and part of 2024 (p.18). The rationale for this is provided. It is also mentioned under scope of the Inception/Design Report, which indicates that this modification was agreed prior to the start of the evaluation.
SECTION D:		**EVALUATION DESIGN AND METHODOLOGY (weight 20%)**	86%	**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?** *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	Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Yes	The evaluation questions are well-aligned with the objectives and purpose of the evaluation. During the design phase, evaluators refined the evaluation questions and criteria in consultation with the Evaluation Reference Group (ERG) to ensure adherence to best humanitarian evaluation practices and to avoid duplication in data collection (p.20). This approach is considered good practice, as it ensures that the evaluation questions are clear and concise, allowing for a more focused evaluation. In this case, the number of evaluation questions (EQs) was reduced from 17 (including combined questions) in the Terms of Reference (ToR) to 9 EQs. Appendix 3a provides a comparison of the original and revised EQs, this is helpful as it allows users to track the changes made. In addition, the evaluation matrix (Appendix 5) shows how some original EQs have been reframed as assessment criteria, such as indicators under coverage and connectedness. This is appropriate, as it ensures that the original questions are still addressed. OECD DAC and additional humanitarian criteria are specified and are aligned with the evaluation questions. For this evaluation, they include: relevance, effectiveness, coverage, coherence, efficiency, and connectedness.
	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.	Yes	The Evaluation Matrix is provided in Appendix 5. It clearly presents the evaluation questions under the corresponding criteria. It also includes assumptions about each question that will be tested in the evaluation. In addition, the matrix includes assessment criteria/indicators against each assumption, along with interview/discussion questions and methods for data collection.
Question 7.		**Is the theory of change, results chain, logical framework, or equivalent framework well-articulated?**		
	i	Clear description of the intervention's intended results, or of the parts of the results chain that are applicable to, or are being tested by, the evaluation.	Yes	There is a clear description of the Country Programme's intended results that are being tested by this evaluation. This is summarised in Section 2 of the Evaluation Report (p.19) and elaborated on in Appendix 4 with the help of a reconstruction Theory of Change (ToC). It includes the activities, outputs, outcomes and impact of the intervention.
	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).	Yes	The causal relationships between the activities, outputs, outcomes and impacts (including the Transformative Results) are clearly demonstrated in the Theory of Change (ToC) figure provided in Appendix 3 (p.15). The narrative of the linkages between the various elements are provided on p.13-14 of Appendix 3. A summary of the relationships is also provided in the main body of the Evaluation Report, which sign-posts to the Appendix (p.19-20).
	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.	Yes	The Evaluation Report provides a fairly comprehensive analysis of the theory of change (ToC) in Section 2 (p.19), which signposts to Appendix 3 for more information. The ToC was reconstructed by the evaluation team drawing from the original Country Programme Document and relevant documents such as the country programme's extension and relevant strategies such as the UNFPA Strategic Plan for 2018-2021 and 2022-2025 (p.13 of Appendix 4).
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 evaluation design and methods used are clearly described in Section 2. It includes both qualitative and quantitative data, the data sources and sampling are assessed separately below. The evaluation matrix is the basis of the analytical framework for the evaluation and the report shows clear linkages between the evaluation criteria, questions, and Theory of Change (ToC). Overall, the design and methods are relevant and robust for meeting the evaluation's purpose, objectives and scope. The use of AI is not mentioned in the evaluation report.
	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 data sources are clearly described in Section 2 (p.22), these include: document review; key informant interviews; focus-group discussions; and site visits/observations (p.22). It includes qualitative and quantitative data. The primary sources are mainly qualitative. The full list of secondary sources is provided in Appendix 9.
	iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or if not, provide reasons for this).	Partially	There is a description of the sampling strategy provided in Section 2 (p.23-24). It describes a 'multi-stage sampling approach' for primary data collection. While it is unclear how the stages were employed, it provides key criteria for sampling including regional representation, security considerations to sites, balance between urban and rural representation and so on. It also mentions purposive sampling and snowball sampling. It also mentions that a stakeholder mapping framework was created to identify the organisational stakeholders (detailed in Appendix 3b) with the full list of key informants provided in Appendix 8. What can be improved is the sampling strategy for each of the data collection methods to enhance clarity and transparency on how diverse perspectives were captured. For example, it is unclear how participants for FGDs were selected.
	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	The methods allow for a rigorous testing of the ToC, particularly as the ToC is clearly laid out and described, including the linkages between the different components of the ToC. The evaluation framework also includes some of the underpinning assumptions linked to the evaluation questions and, to some extent, the ToC.

			Rating	Comments on Rating
	v	Clear and complete description of the methods of analysis, including explainability and full disclosure of the use of AI in the evaluation process, if applicable.	Yes	There is a limited description of the method of analysis in the Evaluation Report. It mainly describes the analytical process to include the use of the OECD DAC criteria, evaluation questions, assumption, and indicators, as depicted in Figure 2 (p.20). This is also elaborated on in the Evaluation Matrix (Appendix 5). It also provides a description of triangulation and its strength of evidence rating (p.22-23). However, a more comprehensive description of the methods of analysis is provided in the Inception Report (p.29). There is no mention of the use of AI in the evaluation.
	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 clear and complete description of the limitations and constraints faced by the evaluation provided in Table 5 (p.27-29). The table includes a column describing the risk and/or limitation alongside the risk outcome and mitigation strategy. Each one is then provided with a rating on its impact on the evaluation. The main areas cited were related to the security situation in Yemen and other external risks. There were a few based on stakeholder engagement and the challenges of the temporal scope (spanning 8 to 10 years), which had implications on data availability and access to relevant informants. The issue on mitigation of bias is not described in the Table but this is partly covered by the section on sampling, triangulation and strength of evidence rubric.
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	While there is explicit reference to UNEG norms and standards for evaluations and UNEG Ethical Guidelines for Evaluation (p.25), it would be helpful if they were contextualised for this evaluation.
	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	There is a description of ethical considerations in the evaluation report (p.25). It describes the processes in place to ensure data security, confidentiality and conflicts of interest. There is also a table provided that list the ethical considerations, risks and safeguards to mitigate risks such as contact with vulnerable women and children (Table 4, p.25). It describes the use of explicit informed consent from participants and that no involvement from participants under the age of 18 years. Evaluators can also be encouraged to include in the Annex, the protocols used to obtain consent as well as safeguard mechanisms, to better understand what was in place.
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.	Partially	In the Inception Report, evaluators consider the use of innovative tools in their contingency strategy in case of a security situation in the country. These tools include SMS web-based text interviews and telephone interviews. While telephone interviews are a more conventional approach, the use of an internet-enabled mobile application, described by the evaluators as having 'gained popularity among researchers' since the COVID-19 pandemic, represents an innovative method for engaging stakeholders (p.25 of the Inception Report). It would be interesting to understand how such an application would be implemented and what controls would be necessary. However, although there were restrictions in data collection, the use of this tool was not mentioned in the final Evaluation Report. Nonetheless, this criterion is rated as partially met as it demonstrated some innovative planning in its design.
SECTION E:		**EVALUATION FINDINGS (weight 25%)**	92%	**Comments on Rating**
Question 11.		**Do the findings clearly and adequately address all evaluation questions and sub-questions?**		
	i	Findings are presented clearly and provide sufficient levels of evidence to systematically address all the evaluation's questions	Partially	The findings are clearly and systematically organised, with each evaluation question (EQ) addressed in the order outlined in the Evaluation Matrix (Appendix 5). In total, there are 39 findings corresponding to the 9 evaluation questions. The links between some findings and the assumptions documented for each EQ in the Evaluation Matrix are also evident. For instance, for EQ2, each finding is clearly linked to the assumptions identified in the Evaluation Matrix, which this evaluation aims to test. The findings are well-supported by an appropriate level of evidence. While this thorough approach is commendable, it has resulted in a findings section that is 70 pages in length (discussed further in Q18ii below). Evaluators may benefit from stepping back to consolidate some findings and/or reconsider the level of evidence or detail required. In addition, the UNFPA Evaluation Handbook 2024 asks that each section that corresponds to an evaluation question is opened with a summary. This is an area for improvement in order to fully meet this criterion.

	ii	Explicit use of the evaluand's theory of change, results chain, logical framework in the formulation of the findings.	Yes	There is explicit use of the reconstructed theory of change (ToC) in the formulation of the findings, which is evident throughout the findings section. While this evaluation is not described as 'theory based', assumptions are identified for each evaluation question in the Evaluation Matrix, which are tested through the evaluation and reported on in the findings section. The evaluators have also indicated where each evaluation question is positioned within the ToC (see red circles in Appendix 5), enhancing the understanding of both the causal links between different ToC components and the underlying assumptions. While the evaluators describe the use of these assumptions as part of their analytical process, they can also be encouraged to discuss them more explicitly in relation to the ToC.
Question 12.		**Are evaluation findings derived from credible data sources as well as a rigorous data analysis?**		
	i	Evaluation uses credible forms of qualitative and quantitative data. It presents both output and outcome-level data as relevant to the evaluation framework. Triangulation is evident using multiple data sources.	Yes	The evaluation uses credible forms of qualitative and quantitative data, incorporating both output and outcome level data as presented in the theory of change, and relevant to the Evaluation Framework. The references are provided in the footnotes, which demonstrate the use of multiple data sources. Additionally, the evaluators have developed a 'Strength of Evidence' rubric (p.23), with triangulation informing the categorisation of evidence as 'strong,' 'moderate,' or 'weak.' This approach is a good practice that other evaluations could replicate to enhance transparency regarding the strength of evidence. In Appendix 7, each of the 34 findings are assigned a 'strength of evidence' rating based on primary data from various stakeholder groups, triangulation levels, secondary data sources, and weighting. One suggestion is for evaluators to include an overall rating for the findings in the main report, with references to the Appendix for further details. They can also consider using the three rating categories (e.g., through colour coding), as the overall strength is currently indicated in different shades of green, yellow, and orange, which is confusing without a closer look.
	ii	Findings are clearly supported by the evidence presented, both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Yes	Findings are clearly supported by the evidence presented. They include both positive and negative findings, and many that are more nuanced. For example, see Findings 3, 4, 6, and 11 as examples. The findings are based on clear indicators, which are provided in the Evaluation Matrix (Appendix 5).
	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).	Yes	The causal factors leading to achievement or non-achievement of results are identified throughout the findings section. This is particularly evident in relation to EQ3 and EQ4, on effectiveness, which discusses contextual factors, resourcing, data and other factors that contribute or constrain the achievement of results.
Question 13.		**Does the evaluation assess and use the intervention's Results Based Management elements?**		
	i	Assessment of the adequacy of the intervention's planning, monitoring, and reporting system (including completeness and appropriateness of results/performance framework - including vertical and horizontal logic, M&E tools and their usage) to support decision-making.	Yes	There is a specific finding on the data and management system of UNFPA Yemen (see Findings 31, p.96-97). It outlines some of the main issues related to reporting and management as well as utilisation for decision making. Earlier in the report, the evaluation also assesses some of the monitoring and reporting mechanisms, not all UNFPA initiative, such as third-party monitoring (p.40) and Rapid Response Mechanism (p.68).
SECTION F:		**EVALUATION CONCLUSIONS (weight 10%)**	100%	Comments on Rating
Question 14.		**Do the conclusions clearly present an unbiased overall assessment of the evaluand?**		
	i	Conclusions are clearly formulated and present unbiased summative statements that respond to the evaluation questions.	Yes	The conclusions are clearly formulated and present an unbiased summary of the findings. They respond to the evaluation questions and there is clear referencing provided to the findings for each of the five conclusions. A minor area for improvement is for the evaluation to note down the links to the evaluation questions as well as the associated recommendations, as per the guidance in the UNFPA Evaluation Handbook 2024. The conclusions may also be structured into strategic and programmatic conclusions.
	ii	Conclusions are well substantiated and derived from findings and add deeper insight and analysis beyond the findings.	Yes	The conclusions are well written and well substantiated. It is clear that they are derived from findings and sign-posting to specific findings are also provided to increase transparency. The conclusions present a synthesis of key points drawn and also provide a deeper insight and analysis that goes beyond the findings.
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	Please note that this criterion has not been rated as lessons learned was not required in the ToR. However, feedback is provided for learning purposes. The lessons learned are derived from findings and are well described and substantiated. They are also practical with illustrative examples provided, where relevant. It would be helpful to include sign-posts to the relevant findings for audiences who are interested.

	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	Please note that this criterion has not been rated as lessons learned was not required in the ToR. However, feedback is provided for learning purposes. The lessons learned are clearly presented and they provide both positive insights on the programme as well as areas for improvement. Some of the lessons are nuanced and include both positive and negative insights. As there are a total of 17 lessons, evaluators can consider organising them according to sub-categories such as 'strategic', 'programmatic', 'resourcing', 'partnership' and so on.
SECTION G:		**EVALUATION RECOMMENDATIONS (weight 15%)**	**100%**	**Comments on Rating**
Question 16.		**Are recommendations well-grounded and articulated?**		
	i	Recommendations are clearly formulated and logically derived from the findings and/or conclusions.	Yes	The recommendations are clearly formulated and are logically derived from findings and conclusions. Each recommendation has clear signposting to the relevant conclusions, which in turn sign-post to relevant findings.
	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.	Yes	The recommendations are useful and actionable for primary intended audience, which are primarily the UNFPA Yemen country office (p.112). Each recommendation is supported by suggested key actions. Each key action is given a priority rating and a timeline. The definitions for the priority rating and timeline is provided in a table on p.112.
	iii	Process for developing the recommendations is described, and includes the Involvement of key stakeholders (e.g. evaluation reference group members), including those who will be affected by the recommendations.	Yes	The process for developing recommendations is described at the start of Section 7 (p.112). It describes a first draft was prepared by the evaluation team, which were discussed with the Evaluation Reference Group (ERG) and then in a dedicated workshop with the UNFPA country office.
	iv	Recommendations are clearly articulated and prioritized based on their importance, urgency, and potential impact.	Yes	The recommendations are clearly articulated. While each key action has a priority rating, the evaluation report does not say if the recommendations are organised in order of priority. In addition, even though the recommendations are directed to UNFPA Yemen country office, which is a broad group, it specifies that the *'delegation of individual actions or tasks will be the responsibility of management.'*
SECTION H:		**REPORT STRUCTURE AND PRESENTATION (weight 5%)**	**75%**	**Comments on Rating**
Question 17.		**Does the evaluation report include all required information?**		
	i	Opening pages include: Name of evaluation and/title of evaluation, timeframe of the evaluation, date of report, location of evaluand, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents (including, as relevant, tables, graphs, figures, annexes)-; list of acronyms/abbreviations.	Yes	The opening pages include all the required information expected of a UNFPA evaluation. This includes the name of the evaluation, the timeframe of the evaluation, the date of the report, and the location of the country programme. It also includes the names of the evaluation team, the names of the evaluation managers and the evaluation reference group. There is a clear table of contents that includes relevant tables, graphs, figures and annexes. There is also a list of Acronyms provided.
	ii	Annexes include, if not in body of report: terms of reference, evaluation matrix, list of respondents, results chain/ToC/logical framework, list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology (e.g. inception report), case study reports.	Yes	The Annexes are provided as a separate document. They include the Terms of Reference (Appendix 1); Evaluation Matrix (Appendix 5); List of respondents (Appendix 8), Theory of Change (Appendix 4), List of sites (Appendix 3c), data collection instruments (Appendix 3d) - although these were provided as links and were not accessible, and list of documentary evidence (Appendix 9). There were other documents added to the Annex such as Evaluation Management (Appendix 2), Additional methodological detail such as revised vs original evaluation questions (Appendix 3), Additional Analysis (Appendix 6) and Strength of Evidence (Appendix 7).
Question 18.		**Is the report logically structured and of reasonable length?**		
	i	The report has a logical structure that is easy to identify and navigate (for instance, with numbered sections, clear titles, well formatted).	Yes	The report has a logical structure that is easy to identify and navigate. It follows the guidance in the UNFPA Evaluation Handbook 2024. It includes clear numbered sections with clear titles. It is well formatted.
	ii	Structure and length accords to UNFPA guidelines for evaluation reports; it does not exceed number of pages that may be specified in ToR. *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)*	No	The structure of the Evaluation Report adheres to UNFPA guidelines. However, it is 102 pages in length, excluding the Executive Summary, which exceeds the maximum number of pages for Country Programme Evaluations by nearly a third. As mentioned above, the findings section alone is 70 pages.
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 well written and easy to understand. It is generally free from grammar, spelling and punctuation errors.
	ii	Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labelled, and referenced in text.	Partially	The evaluation includes frequent use of visual aids including graphs, maps, and tables. This has helped to break up an otherwise dense report. They are clearly presented and labelled. However, they are not always referenced in the text. For example, there is no referencing of Tables 1, 2, and 4 in the narrative of the report.
SECTION I:		**CROSS-CUTTING ISSUES (weight 10%)**	**80%**	**Comments on Rating**
Question 20.		**Are cross cutting issues - in particular, human rights-based approach, gender equality, disability inclusion, LNOB - integrated in the core elements of the evaluation (e.g. evaluation design, methodology, findings, conclusions and recommendations)?**		
	i	Evaluation's data collection methods designed to capture the voices/perspectives of a wide range of stakeholders including right holders, marginalized and vulnerable persons, young people, people with disabilities, migrants or refugee populations, indigenous communities, and other persons that are often left behind.	Partially	The evaluation's data collection methods are designed to capture the perspectives of a wide range of stakeholders, particularly, stakeholders at the regional and country level. The evaluation has also involved the participation of 175 rightsholders. However, it could make clearer who the rightsholders are and how they are identified.

			Rating	Comments on Rating
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	While the evaluation questions do not explicitly mention specific cross-cutting issues, they ask about the implications on or considerations for 'vulnerable and marginalised groups'. Specific cross-cutting components on gender equality, people with disabilities, refugee populations, LNOB and intersectionality are specified in the assumptions or indicators supporting each evaluation question, as seen in the Evaluation Matrix (Appendix 5).
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	Data disaggregation is evident in the findings. However, differential results across groups are not always systematically assessed, and this is likely due to data availability that is discussed in relation to finding 16 (p.63) and finding 31 (p.96). In addition, although the total number of stakeholders consulted is mentioned in the report and a list of stakeholders met are included in the Annex, there is no gender disaggregation provided. Therefore, it is not entirely clear how balanced the sample was.
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	An intersectional lens is applied to some extent in the data analysis, particularly in discussions focused on vulnerable populations. For example, young people, particularly adolescents girls, with disabilities or caring responsibilities together with the Mahram requirement face more challenges in accessing reproductive health services (p.61). However, such analysis is limited, and this may be partly due to data availability so it would have been helpful to hear some of these discussions from the primary data carried out by the evaluators.
v	Findings, conclusions and recommendations, address cross-cutting issues such as equality and vulnerability, disability inclusion, leave no-one behind, social and environmental as relevant.		Yes	The findings, conclusion and recommendations address cross-cutting issues such as equality and vulnerability, disability inclusion, gender considerations, leave-no-one-behind, and environmental considerations.
vi	Inclusion of young people in the evaluation team and/or Reference Group [N/A if not requested in ToR]		Yes	The evaluation team includes a young and emerging evaluator on the team.
Question 21.	**Does the evaluation meet UN SWAP evaluation performance indicators?** *Note: this question will be rated according to UN SWAP standards with detail provided below*		**8**	**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	GEEW is integrated into the evaluation scope of analysis as the key outcomes of the country programme are highly relevant, such as reproductive rights and access to services that GBV. While there is no standalone criteria on GEEW, there is one evaluation question on coverage that explicitly mentions women as well indicators and assessment criteria. Together it ensures that GEEW-related data will be collected.
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.		Satisfactorily integrated	As the focus of the country programme is very much relevant to gender issues, the need to specify how gender issues are being addressed is less relevant than for other evaluation. Nevertheless, data is disaggregated by gender, where available. The evaluation methodology employs a mixed-methods approach and the analysis involves clear triangulation processes to ensure accuracy and credibility for findings. Where the evaluation report could improve is to provide more detail on the rightsholders participants. Currently, it only provides the location and gender and it is unclear whose voices were represented.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.		Fully integrated	The evaluation findings, conclusions and recommendations reflect a gender analysis. It includes a background section that discusses some intersectional analysis, particularly relevant to adolescent girls. The findings include data analysis that explicitly triangulate the voices for different groups. The report also includes specific recommendation on GEEW issues (see for example, recommendation 5 on p. 115).

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.

a. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?

b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?

c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?

d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?

ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.

a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex?

b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations?

c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?

d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?

e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality?

iii The evaluation findings, conclusions and recommendations reflect a gender analysis.

a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality?

b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable?

c. Are unanticipated effects of the intervention on human rights and gender equality described?

d. Does the evaluation report provide specific recommendations addressing GEWE issues, and priorities for action to improve GEWE or the intervention or future initiatives in this area?

List of SDGs

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

Three transformative results

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

Six outputs

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

Six accelerators

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