UNFPA Evaluation Quality Assessment Grid

Version: August 2025

REPORT RATING SUMMARY			
Overall Rating		93%	Highly Satisfactory
● ● ● ● ●	Excellent	5	
● ● ● ●	Highly Satisfactory	4	The report fully meets all UNFPA/UNEG standards for evaluation reports, with minor shortcomings in certain indicators. Decision-makers may use the evaluation with a high degree of confidence.
● ● ● -	Satisfactory	3	
● ● - -	Fair	2	
● - - -	Unsatisfactory	1	

REPORT DETAILS	
Title of the evaluation report	Independent evaluation of the UNFPA support to the integration of the principles of 'leaving no one behind' and 'reaching the furthest behind.'
Region	Assesses integration of LNOB into UNFPA's programming and operations at global, regional and country levels
Country	
Year of report	2025
Business Unit/programme country *(managing evaluation)*	UNFPA Independent Evaluation Office
Date of assessment review (dd/mmm/yyyy)	8-Sep-25
Name of assessment review firm	IOD PARC
CLASSIFICATION OF EVALUATION REPORT	
Primary SDG(s) covered *(list provided below)*	SDG 3 - Good Health and Well-being; SDG 5 - Gender Equality
UNFPA Strategic Plan areas covered *(lists provided below)*	
Three transformative results	1 - Ending unmet need for family planning; 2 - Ending maternal death; 3 - Ending gender-based violence and all harmful practices
Six outputs	All six output areas
Six accelerators	5. Leaving no one behind and reaching the furthest behind first
Organizational effectiveness and efficiency	Yes
Humanitarian evaluation	Yes
Evaluation evaluand *(e.g. country programme/intervention/policy/thematic area)*	Accelerator
Evaluation type *(e.g. formative, summative, developmental)*	Formative
Geographic scope *(e.g. global, regional, national)*	Assesses integration of LNOB into UNFPA's programming and operations at global, regional and country levels

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 evaluation report of UNFPA support to the integration of the principles of 'leaving no one behind' and 'reaching the furthest behind' is a highly satisfactory. The key strengths and weaknesses of the evaluation report are provided below, followed by recommendations for future evaluations.

Strengths:

• **Section B: Background** – There is a clear description of the evaluand (including geographic coverage, implementation period, implementation status, and partners) as well as the context. Its key strength is the stakeholder mapping that is provided in Annex VII of the main report. The stakeholder map identifies categories and sub-categories of stakeholders and analyses their interests, needs, and potential impact.

• **Section C: Evaluation Purpose, Objectives, and Scope** – The evaluation report clearly distinguishes between the purpose, objectives, and scope, each presented in a dedicated sub-section. It describes why it is needed at this point in time, its intended use, and it identifies the key intended users. The objectives are clearly described and support the purpose of the evaluation. What is included and excluded from the scope of the evaluation is also clearly described.

• **Section D: Evaluation Design and Methodology** – The evaluation questions are appropriate for meeting the purpose and objectives of the evaluation. The sub-questions are framed as assumptions. The Evaluation Matrix is provided in Annex 1 and includes all the required components. The evaluation approach is clearly outlined. It employs a mixed-methods approach with a clear description of the data sources and the sampling strategy. The evaluation report also includes a clear and complete section on limitations and mitigation, as well as 'unanticipated challenges'. In addition, it includes a dedciated section on ethical considerations. A description of the data collection is represented by a visual of the number of sources (Figure 6), which are disaggregated by the types of informants (p. 17). This is good practice that other evaluations can be encouraged to adopt. In addition, the report provides a clear description of data analysis.

• **Section E: Evaluation Findings** – The findings are well written and presented clearly and systematically, broadly following the structure of the OECD-DAC criteria, evaluation questions, and assumptions. The evaluation uses credible forms of qualitative and quantitative data, and triangulation is evident through the data sources provided in the text and also in the footnotes. A key strength of the report is how the findings statements are presented and followed by supporting evidence, which makes the report findings easy to digest. In addition, the findings go beyond reporting on what works or does not work, but discuss and analyse the factors contributing to results, or lack thereof.

• **Section F: Evaluation Conclusions** – Conclusions are clearly formulated, providing an additional layer of analysis that draws on the key findings.

• **Section G: Recommendations** – Recommendations are well formulated, useful, and actionable. They are prioritised, provide key actions for implementation, and identify target audiences to take them forward.

• **Section I: Cross-cutting** – This evaluation's involvement of individuals with an expertise or lived experienced as left-behind in the Evaluation Steering Group, can be highlighted as good practice for other evaluations to consider replicating.

Weaknesses:

• **Section A: Executive Summary** – While the Executive Summary is clearly written, it is missing information on the context.

• **Section B: Background** – The background section sets out a helpful description of the evaluand and its context. What would be helpful is some information on the funding mechanisms and budget allocations and the potential impact on stakeholders.

• **Section D: Evaluation Design and Methodology** – Many indicators in the evaluation matrix are worded amiguously ('Evidence of…') without specifying a benchmark for what constitutes sufficient evidence to warrant a positive assessment.

• **Section D: Evaluation Design and Methodology** – While there is a clear description and visual of the theory of change, developed by the evaluators, there are areas that can be improved, particularly how assumptions that are assessed through the evaluation are incorporated into the ToC. Please see below for recommendations.

• **Section H: Report Structure and Presentation** – The opening pages are currently missing the timeframe of the evaluation and the location of the evaluand. In addition, while the table of contents is complete, there are issues with some of the page numbers, such as for figures. The report also exceeds the page limit.

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** – What can strengthen the Executive Summary is its description of the principles of leaving no one behind (LNOB) and reaching the furthest behind (RFB), particularly the significance and the context for it. It is well articulated in the Introduction (Section 1) and Context (Section 2), and evaluators can consider incorporating relevant information, such as global trends. This would be helpful, particularly as the Executive Summary is far below its page limit.

• **Section B: Background** – While the background section is good, what can be strengthened is an analysis of the power and influence that each grouping of stakeholders has that can affect success. Evaluators can also consider providing some information on the funding mechanisms and allocation, which they can then dive into in more depth in the findings section.

• **Section D: Evaluation Design and Methodology** – A suggestion for the Evaluation Matrix is for indicators to be worded as benchmarks or to clarify how judgments will be made. Several of the indicators are worded as 'Evidence of…', but it is not clear if there is a threshold for 'sufficient evidence'. If this links to triangulation of data (e.g. Table 3 on p.20) then evaluators can signpost accordingly.

• **Section D: Evaluation Design and Methodology** – On the theory of change, assumptions about causal relationships and mechanisms are not explicit or explicitly linked to relevant causal pathways. As some of these are being examined through the evaluation questions, as assumptions, evaluators may wish to represent them within the ToC. This can be achieved, for example, by numbering the assumptions to better situate them within the ToC and describing the assumptions about the causal pathways in the narrative. Similarly, external assumptions can be linked to the relevant pathways. In addition, evaluators are encouraged to expand the ToC by incorporating a clear problem statement, a high-level summary of activities and more explicit assumptions concerning causal relationships and stakeholder characteristics.

• **Section H: Report Structure and Presentation** – Evaluators are encouraged to include the timeframe of the evaluation and not just the year of the report in the opening pages. Where possible, evaluators can cross-check the contents page to ensure they are accurate. Lastly, evaluators can consider where content can be removed or moved to the Annex in order to meet the page limit.

SECTION RATINGS			
SECTION A:	**EXECUTIVE SUMMARY (weight 5%)**	50%	**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).	Partially	The Executive Summary is 4 pages in length, which is less than the required length. It can be used as a standalone document useful for 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	The Executive Summary includes nearly all necessary components: 1) Overview of the intervention (under 'Background' on p.i); 2) Evaluation purpose, objectives and intended audience (under 'Purpose and Scope of the Evaluation' on p.i); 3) Scope and evaluation methodology (p.i-ii); 4) Main findings (p.ii-iii); 5) Conclusions (p.iii-iv); and 6) Recommendations (p.iv). As page numbers are not provided in the Executive Summary, it is recommended to number them (i to iv) before publication, to aid navigation of the document. The missing component is the context, which evaluators can consider incorporating in the section titled 'Background'.
iii	Includes all significant information in a concise yet clear manner to understand the evaluand (e.g., theme, intervention, programme, project, or strategy undergoing evaluation) and the evaluation.	Partially	Overall, the Executive Summary provides a clear description to understand the evaluation, including the methodology and evaluation results. What can be strengthened is its description of the principles of leaving no one behind (LNOB) and reaching the furthest behind (RFB), particularly the significance and context for it. This is well articulated in the Introduction (p. 1) and Context (p. 4) sections of the report, and evaluators can consider incorporating relevant information, such as global trends and demographic shifts.
SECTION B:	**BACKGROUND (weight 5%)**	90%	**Comments on Rating**
Question 2.	**Is the evaluand (i.e. intervention/policy/thematic area etc. that is to be evaluated) and context of the evaluation clearly described?**		

i	Clear description of the evaluand (e.g., theme, intervention, programme, project or strategy undergoing evaluation), including: geographic coverage, implementation period, main partners, cost/budget, and implementation status.	Yes	There is a clear description of LNOB and RFB in the introduction section of the evaluation report (p. 1-3) i.e. 'the integration of LNOB and RFB approaches and principles across UNFPA programmatic and operational work' (the evaluand) (p.1). The geographical scope is outlined on p. 2, covering all six UNFPA regions, with analysis conducted at global, regional, and country levels. Further detail on the sampling approach is provided in the methodology section. The temporal scope (implementation period) spans two strategic plans: 2018–2021 and 2022–2025 (p. 2). At the time of the evaluation, the UNFPA Strategic Plan 2022–2025 was under implementation (implementation status). Main partners are not explicitly identified in the main report, except in reference to "evaluation stakeholders and audience" (p. 3). However, the section signposts to the stakeholder mapping in Annex VII, which sets out the primary stakeholder groups and sub-groups, along with their responsibilities, interests, and concerns. It should also be noted that the 2024 Evaluation Handbook does not require a description of the main partners. The evaluation report does not include a budget or cost, explaining that, as the evaluand is not a specific programme, it does not have specific budget lines that can be "accurately tracked" (p. 1). Nevertheless, given that EQ5 focuses on the efficient allocation of resources, it would be useful for the report to provide a summary of funding mechanisms and allocations, which it can then dive into in more depth in the findings section.
ii	Clear description of the context of the evaluand (e.g. economic, social and political context, relevant aspects of UNFPA's institutional, normative and strategic framework, cross-cutting issues such as gender equality and human rights, disability and LNOB dimensions) and how the context relates to the evaluand (e.g. key drivers and challenges that affect the implementation of the thematic area, intervention, programme, project or strategy).	Yes	Section 2 provides a clear description of the context for integrating LNOB and RFB approaches and principles. It summarises global trends, demographic shifts, gender equality, and the political and socio-economic factors influencing access to sexual and reproductive healthcare (pp. 5–6). The link between this context and the evaluand is clearly articulated. The section also outlines UNFPA's strategic frameworks, including its Strategic Plans and the LNOB Operational Plan (p. 7), and describes key drivers and challenges affecting LNOB implementation (p. 8). In addition, the evaluators note the "significant financial and geopolitical shifts that began in early 2025" to capture the evolving context as the evaluation was concluding (p. 4). This is a welcome addition for primary audiences as they take forward the evaluation results. While the section provides a good outline of the global context of LNOB, it would be beneficial if the 'global trends' paragraph could provide quantitative data about global trends on the development status or outcomes of left-behind groups, such as people with disabilities and others. For example, the paragraph could report on certain SDG indicators (these could be limited to the UNFPA mandate area) for which disaggregated data is available to demonstrate that certain groups are still left behind.
iii	Linkages between the evaluand and ICPD-related SDGs are outlined. This may include reference to ICPD benchmarks and relevant SDG targets and indicators.	Yes	Section 2.2 provides the linkages between the evaluand, SDGs and ICPD (p.6-7). It states that the SDGs of particular relevance for this evaluation are SDG 3, SDG 5, SDG 10, and SDG 16. The relevant target indicators are provided in footnote 33. An SDG dashboard data or table is not provided but this is not a current requirement for a centralized evaluation to meet this criterion.
Question 3.	**Are key stakeholders clearly identified and analysed?**		
i	Clear identification of key stakeholders which should include implementing partner(s), development partners, rights holders, and duty bearers among others (e.g., stakeholder map).	Yes	There is a brief mention of key stakeholders in the main report (p. 3) that sign-posts to a more detailed Stakeholder Mapping in Annex VII of the main report. The stakeholder map identifies categories of stakeholder groups (i.e. primary group) and breaks them down to sub-groups. This stakeholder map builds on the stakeholder map provided in the Annex of the Inception Report.
ii	Stakeholders are analysed to understand their interests and needs, power and influence, and potential impact on the evaluand.	Partially	The stakeholders are analysed to understand their interests and needs and potential impact. The stakeholder mapping presented in Annex VII provides, for each 'grouping', the rights, duties, needs, interests, concerns, and potential impact on the evaluand are described. What can be strengthened, in order to fully meet this criterion, is an analysis of the power and influence that each grouping of stakeholders has that could affect success.
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 evaluation report clearly distinguishes between purpose, objectives, and scope (pp. 2–3), with each presented in a dedicated sub-section. The purpose of the evaluation (p. 2) is consistent with the UNFPA Evaluation Policy (2024), aiming to promote accountability, support evidence-based decision-making, and inform implementation. It mentions learning on what works, what does not, for whom, under what circumstances, and why (p. 2), and explicitly states that the evaluation is intended to empower communities and stakeholders (p. 2). The purpose also specifies generating 'real-time evidence on how LNOB and RFB approaches are being implemented ...' (p. 2). This is further elaborated under the objectives (p. 2), which include informing the upcoming Latin America and Caribbean Regional Programme and the UNFPA Strategic Plan 2026–2029. Evaluators might consider including this rationale under "purpose," as it clarifies why the evaluation is needed at this point and how the results will be used. However, this does not affect the score, as the information also fits appropriately under "objectives." The key intended users of the evaluation are identified on p. 3.
Question 5.	**Are the objectives and scope of the evaluation clear and realistic?**		
i	Clear and complete description of the objectives of the evaluation, including reference to any changes made to the objectives included in the inception report (if applicable).	Yes	The evaluation objectives are clearly described on p. 2. They support the purpose of the evaluation. They align with the objectives presented in the ToR and Inception Report (p.1 of IR).
ii	Clear and relevant description of the scope (e.g. thematic, geographic, and temporal) of the evaluation, covering what will and will not be covered, as well as, if applicable, the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political, humanitarian or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Yes	The scope is clearly and explicitly stated on p. 2-3. This includes a description of the temporal, geographic and thematic scope. The rationale for the geographic scope is briefly described with sign-posts to the sampling strategy in Section 3.3.4. This section also states that the evaluation does not evaluate financial data relating to LNOB programming due to lack of access to the data in UNFPA's 'new financial system' (p.3).
SECTION D:	**EVALUATION DESIGN AND METHODOLOGY (weight 20%)**	89%	**Comments on Rating**
Question 6.	**Are the selected evaluation questions and evaluation criteria appropriate for the purpose of the evaluation and is there clear justification for their use?**		
i	Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Yes	OECD/DAC evaluation criteria are identified against each of the evaluation questions (p.13-14 and Annex 1). These include Relevance (EQ1), Effectiveness and Sustainability (grouped together as EQ2), effectiveness and humanitarian-development-Peace (HDP) nexus (EQ3), Coherence (EQ4) and Efficiency (EQ5). These are as presented in the Inception Report. The evaluation questions and sub-questions, which are framed as assumptions, are presented in Table 1 (p.13-14). They are appropriate for meeting the purpose and objectives of the evaluation.
ii	Evaluation matrix clearly presents the evaluation criteria used as well as the corresponding evaluation questions, assumptions, indicators/ lines of inquiry, and methods for data collection, including relevant data sources.	Partially	The Evaluation Matrix is included in Annex 1. It clearly presents each question against the OECD-DAC criteria. Each of the five main questions are supported by assumptions. For each assumption, there is a column titled 'indicators/benchmarks' that describes the types of evidence that will be used to verify each assumption. Each assumption is also accompanied by sources of data, data collection and data analysis methods. The matrix also includes links to the evaluation theory of change (ToC) and Area of Inquiry (AoI) in the ToR. To fully meet this criterion, evaluators can be encouraged to specify the data sources, particularly for KIIs. For example, they specify 'KIIs at global, regional and country levels', but it is not clear who these will involve. If it is the case that all KIIs and FGDs will be asked questions then this can be mentioned at the top to avoid repetition. Another suggested improvement is for indicators to be worded as benchmarks or to clarify how judgments will be made. Several of the indicators are worded as 'Evidence of...' but it is not clear if there is a threshold for 'sufficient evidence'. If this links to triangulation of data (e.g. Table 3 on p.20) then evaluators can sign-post accordingly.
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 relevant parts of the results chain-being tested by the evaluation, through a ToC, results chain or logical framework. This should be presented in both a visual (graphic or tabular) and accompanying narrative format.	Yes	There is a clear description of the Theory of Change (ToC) in Section 3.2 (p.11). This is accompanied by a visual of the ToC (Figure 4, p.12). The ToC was developed by the evaluation team for the purpose of the evaluation (p. 11). As such, it clearly outlines the OECD/DAC criteria and corresponding evaluation questions in blue dots number 1-5.
ii	Critical assumptions underlying the ToC, results chain or logical framework are clearly identified, including assumptions about causal relationships, contextual factors and stakeholder characteristics.	Partially	The ToC identifies assumptions about contextual factors, which are presented in the box on the left-hand side. These include, for example, 'favourable political, socio-economic context; functioning civil society; donor prioritization...' These assumptions are expected to be tested through the evaluation. On the opposite side of the ToC diagram are internal assumptions, such as those relating to the organisation. To some extent, the organisational ones include assumptions about some of the stakeholders, although only in a limited way. Assumptions about causal relationships or causal mechanisms are not explicitly reflected; only those on stakeholder characteristics and contextual factors are articulated in the ToC diagram . Since some of these are being examined through the evaluation questions, as assumptions, evaluators may wish to represent them within the ToC. This could be achieved, for example, by numbering the assumptions to better situate them within the ToC and describing the assumptions about the causal pathways in the narrative. Similarly, external assumptions can be linked to the relevant pathways.

iii	Comprehensive assessment of the logic, internal coherence and plausibility of the ToC, results chain or logical framework, and if required, it is refined or revised by the evaluators.	Partially	As no formal Theory of Change (ToC) existed for the LNOB Operational Plan, the evaluation team developed one during the course of the evaluation. Therefore, the report was not well-placed to provide a critical analysis of the ToC. However, the evaluation reconstructed the ToC through a critical analysis of documents and KIIs. The constructed ToC outlines the intended results (outputs, outcomes, and transformative results), depicts some causal linkages - primarily through arrows connecting the results, and identifies a number of assumptions, mostly related to contextual factors. To fully meet this criterion, evaluators are encouraged to expand the ToC by incorporating a clear problem statement, a high-level summary of activities, and more explicit assumptions, particularly those concerning causal relationships and stakeholder characteristics.
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 approach and design are clearly outlined in Section 3 (p. 10 onwards), with additional detail provided in Annex 5. It is a formative evaluation, informed by both feminist and transformative approaches. The design is theory-based, mixed-methods, participatory, and utilisation-focused. Given the evaluation's focus on the LNOB and RFB principles, the approaches are not only well articulated but also appropriate and relevant for addressing the evaluation's purpose, objectives, and scope. The report also states that no artificial intelligence was used during the evaluation process (p. 26).
ii	Data sources are all clearly described and are relevant and robust; these would normally include qualitative and quantitative sources.	Yes	This evaluation employs a mixed methods approach (p. 10). The data sources are described on p. 14 onward. They include qualitative and quantitative documents (p. 15 and Annex IX), involving 242 sources (p.17); semi-structured key informant interviews (p. 16 and Annex VI); online survey (p.16); and focus group discussions (p.16). Figure 6 provides a visual of the number of sources, which are disaggregated by the types of informants (p. 17). This is good practice that other evaluations can be encouraged to adopt.
iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or if not, provide reasons for this).	Yes	The sampling strategy is described in Section 3.3.4 (p. 18-19). Purposive sampling was used to select regional, thematic and country case studies. The selection criteria for country case studies is provided on p. 18. The selection criteria for thematic case studies are also described on p. 18. The selection of the Latin America and Caribbean country studies is described towards the end of the paragraph on the country case studies (p. 18). The evaluation report also describes the sampling strategy for KII and FGDs (p. 18-19), online survey (p. 19), and sampling of documents and data (p. 19). This includes a description of the stakeholder mapping (provided in Annex VII) and the process for selection at the global, regional, and country level. One suggestion to strengthen the transparency of the sampling is for evaluators to provide the criteria for how they identified key informants with UNFPA regional and country offices, but in a way that mitigates potential bias.
iv	Methods allow for rigorous testing of the theory of change, results chain or logical framework (e.g. methods help to understand the causal connections, if any, between outputs and expected outcomes.	Yes	The methods provide a basis for rigorous testing of the Theory of Change (ToC).
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	The evaluation report states that no artificial intelligence (AI) was used (p. 25). How the evaluation matrix is used for data collection and analysis is described in Section 3 (p. 11 onward). A short version of the evaluation matrix is presented in the main body of the report (p. 13) with the full version included in Annex 1. The evaluation report provides a clear and complete description of data analysis methods in Section 3.4 (p. 19 onward). This includes a description of the data coding and processing of primary and secondary data at global, regional, country, and thematic (p. 20). It also describes qualitative and quantitative data analysis. What is good practice is the rating on 'strength of evidence' rating based on the quality and quantity of evidence in the process of triangulation (p.20). There is a helpful description of the evaluation team's introduction of an LNOB transformative continuum (p.21). However, it was difficult to apply, as planned (p.22 and p.41). Other areas of analysis include a power dynamics lens as well as gender and human rights (p.24).
vi	Clear and complete description of limitations and constraints faced by the evaluation in its data collection and analysis, along with the mitigation measures implemented by the evaluation to address these limitations, where feasible.	Yes	There is a dedicated section on limitations and mitigations in Section 3.10 (p. 26). These are clear and complete. There is a table that provides the 'anticipated limitation' with corresponding 'anticipated mitigation measure' and 'actual limitation and mitigation' (p. 26-27). The evaluation report also describes the 'unanticipated challenges' in this section, describing what they were and mitigating steps.
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.	Yes	Ethical considerations are described in Section 3.9 (p. 25). The report makes explicit reference of the evaluation's compliance with the UNEG Ethical Guidelines for Evaluation and the UNEG Code of Conduct for Evaluation. It also provides contextualised reference for how the evaluation considered and/or applied ethical principles including integrity, independence and impartiality, transparency, accountability, credibility, respect, beneficence, data privacy and confidentiality, and safeguard mechanisms for respondents (p. 25-26).
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.	Yes	There is a clear description of ethical issues and considerations in Section 3.9 (p. 25), which are contextualised for this evaluation. The dedicated sub-section on ethical considerations is considered good practice. The Inception Report provides more information on ethical considerations for different phases of the evaluation e.g. design, data collection, analysis, conclusions, and recommendations (p.35 of IR). Ethical measures and considerations for ensuring data privacy and confidentiality are mentioned on p.26 and elaborated on in the Inception Report (p.34 of IR). Data collection tools are included in Annex 2. They also include the protocols used during data collection to obtain consent during the KIIs and the FGDs. However, even with the explanation on how consent was obtained, one of the tools (online survey) did not include a section for participants to provide their consent to participate.
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.	Yes	The establishment of a Steering Committee composed of individuals with lived experience i.e. people who identify with various 'left behind' factors or who work closely with such groups (p.23 of the Annex), is innovative. Annex 3 sets out the rationale, responsibilities, composition, and modes of engagement of the Steering Committee. It would be valuable to hear more about how members were selected, whether and how they were compensated for their time, and what barriers or lessons were encountered. While this does not affect the score, it would be interesting to learn more, as evaluators and commissioners may consider replicating it.
SECTION E:	**EVALUATION FINDINGS (weight 25%)**	100%	**Comments on Rating**
Question 11.	**Do the findings clearly and adequately address all evaluation questions and sub-questions?**		
i	Findings are presented clearly and provide sufficient levels of evidence to systematically address all the evaluation's questions.	Yes	The findings are presented clearly and systematically (p. 30 onward). The structure broadly follows the structure of the Evaluation Matrix, starting with the OECD-DAC criteria and evaluation question. For each evaluation question are a series of 'findings statements' and each statement includes links to the assumption numbers. Typically each finding links to one or two assumptions, the assumptions are relevant to the evaluation questions. This makes it easy to cross reference between the findings section and Evaluation Matrix, and to ensure that all evaluation questions are addressed. All findings statements are supported with sufficient levels of evidence.
ii	Explicit use of the evaluand's theory of change, results chain, logical framework in the formulation of the findings.	Not Rated	This criterion is not rated as the UNFPA Evaluation Handbook 2024 does not require evaluations to make reference to the ToC in the formulating of the findings in the narrative or its structure. However, the assumptions for each evaluation question are answered systematically.
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. This is evidenced in the references provided in the footnotes as well as in Annex 6 and Annex 9. The evaluation report indicates triangulation by referring to data sources within the text and also as footnotes. Findings often include primary data sources and secondary data sources e.g. 'UNFPA regional and country level KIIs' (footnote 152), 'KII UNFPA LACRO, CSO Costa Rica, CSO Panama, IPs Peru, CSO LAC, participants Intercultural Dialogue for Recognition and Inclusion' (footnote 176), and 'Evaluation of the 6th UNFPA Iran country programme 2017-2021' (footnote 199). In some cases, it would be valuable to understand the type of stakeholder category at regional and country-level. It is possible that these were not included so that sources are not identifiable. What evaluators can consider including in the evaluation report is their triangulation rating, based on the rubric developed and described on p. 20, for each finding. This provides transparency of the strength of evidence while ensuring confidentiality. Furthermore, data is provided at output and outcome level, which is most evident in the findings section related to evaluation questions under effectiveness and sustainability. Findings are also corroborated by multiple data sources or methods. For example, the footnote for Finding 6 states that evidence is supported by global and regional KIIs and document review, as well as thematic and two country case studies. Furthermore, the supporting evidence for Finding 6 also provides references in the footnotes that indicate multiple data sources and data collection methods used.

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 supported by the evidence presented. These include both positive and negative findings or, more nuanced findings. For example, finding 9 and finding 10 correspond with assumptions 3.1 and 3.2 for evaluation question 3. The findings statement is both positive and negative - *"Finding 9. There is evidence of a clear intention to understand both continuing and existing vulnerabilities does not consistently translate into fully addressing the intersectional needs of those furthest behind in complex humanitarian situations and newly emerging vulnerabilities within UNFPA humanitarian response programming. This, however, does not consistently translate into fully addressing the intersectional needs of those furthest behind in complex humanitarian situations'* (p. 52-53). This statement is supported by relevant evidence with corresponding references in the footnotes.
iii	Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings analyse the logical chain (progression -or lack thereof - from outputs to high level results).	Yes	Causal factors leading to the achievement or non-achievement of results are clearly identified. This means not just stating what worked or not but discussing and analysing the factors contributing to results, or lack thereof. This particularly evident in the findings section related to the OECD-DAC criteria on effectiveness i.e. EQ2 and EQ3, but evident in other questions as well. For example, to answer EQ2, the report identifies some of the factors that have led to the operationalisation of LNOB principles. For instance, Box 4 on p.40 provides an example of the use of intersectional analysis and context-specific data, allowing the programme to go 'beyond demographic generalisations to effectively reach those furthest behind' (p.40). Another example from Türkiye is offered on the same page. The following page also identifies factors that hinder the translation of 'rhetoric into practice' (p.40) such as awareness of prioritisation tools provided by HQ, scale of need, adaptation to local context and more.
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	Under evaluation question 5, assumption 5.5 states that 'UNFPA has the capacity to monitor, collect and disaggregate its data and results to assess existing inequalities and ensure UNFPA reaches specific groups who are the furthest behind' (p.14). As such, Finding 18 assesses monitoring of LNOB (p. 80-81). The section begins with a short description of what UNFPA seeks to monitor at output, outcome, and goal levels. It then goes on to describe the challenges faced, particularly in understanding intersectionality of multiple levels of vulnerability as well as capacity issues. The report also describes how these challenges limit the organisation's ability to demonstrate outcomes and impact and also to assess its normative role (p.81). Overall, this assessment is adequate for the scope of this evaluation, although a closer look to understand the specific challenges at country, regional, and global level may be required to tackle them effectively.
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 judgements of the evaluand.	Yes	The conclusions are very clearly formulated. They are not a summary of findings but provide an additional layer of analysis that draw on the key findings. Each conclusion is expressed as a concise and impartial judgement, supported by a more detailed narrative that includes clear references to the findings section through the use of relevant findings numbers. The statements are balanced, neither overly positive nor negative, and the accompanying narrative helps substantiates them.
ii	Conclusions are well substantiated and derived from findings.	Yes	As mentioned above in Q14i, the conclusions are well-supported by the analysis. It provides the number of the findings for each conclusion, making it easy to trace. In addition, the narrative supports the conclusion statements. The statements are presented as substantiated judgements. Each conclusion is linked to more than one finding.
Question 15.	**Are lessons learned identified?**		
i	Lessons learned are derived from the findings and are well substantiated with practical, illustrative examples.	Not Rated	The Inception Report mentions lessons learned in relation to analysis and findings, but there is no explicit request in the ToR for a dedicated section on lessons. However, an objective of this evaluation is to 'capture good practices and generate knowledge...'. This is evident in the analysis and the formulation of the conclusion. In addition, the evaluation team have presented 'good practices' in boxes in the findings section, relevant to the evaluation question e.g. *'Box 3. Good practice: Applying the LNOB Operational Plan in humanitarian settings - Afghanistan and Whole of Syria'* (p.32). There are a total of six such boxes, drawn from the country case studies, which are helpful illustrative examples of good practice/lessons.
ii	Lessons learned are clearly presented and provide actionable insights on the positive aspects of the evaluand as well as any areas of improvement.	Not Rated	Lessons learned are not explicitly requested as a standalone section in the inception report of this evaluation. It is mentioned more broadly under 'analysis of data' (p.33 and p.35).
SECTION G:	**EVALUATION RECOMMENDATIONS (weight 15%)**	100%	**Comments on Rating**
Question 16.	**Are recommendations well-grounded and articulated?**		
i	Recommendations are clearly formulated and logically derived from the conclusions.	Yes	The recommendations are clearly formulated. Each recommendation provides the links to the conclusions through numbering. For example, recommendation 1 links to conclusion 1, 2, 5, and 6. Each recommendation specifies the key actions to be taken.
ii	Recommendations are useful and actionable for primary intended users. Specific guidance is provided for its implementation (e.g. strategic or programmatic level, suggested actions/operational implications, and, responsible actors), as appropriate.	Yes	The recommendations are useful and actionable, with key actions identified to support the implementation of the recommendations. A lead unit - either team or division, is targeted for each recommendation. In addition, other units are also identified to support implementation. A timeframe is provided, although it is not required, as it will be specified in the management response. However, this does not affect the scoring.
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 of developing the recommendations is provided at the start of Section 6 (p. 89). They were developed by the evaluation team and refined through multiple consultation with the Steering Committee and Evaluation Reference Group.
iv	Recommendations are prioritized based on their importance, urgency, and potential impact.	Yes	Recommendations are prioritised based on a three-tiered system i.e. 'high', 'medium' and 'low'. In addition, a timeframe for implementation (due date) is suggested, even though this is not required.
SECTION H:	**REPORT STRUCTURE AND PRESENTATION (weight 5%)**	75%	**Comments on Rating**
Question 17.	**Does the evaluation report include all required information?**		
i	Opening pages include: Name of evaluation and/title of evaluation, timeframe of the evaluation, date of report, location of evaluand, names and/or organization(s) of the evaluator(s), name of the evaluation manager, name of organization commissioning the evaluation, list of the evaluation reference group members, acknowledgements, table of contents (including, as relevant, tables, graphs, figures, annexes), and list of acronyms/abbreviations.	Partially	The opening pages include the following components: Title of evaluation; date of the report; name of the evaluation manager; name of the organisation commissioning the evaluation; list of the evaluation reference group members; acknowledgements; table of contents (including Table, Figures, Boxes and Annexes); and a list of acronyms. It does not provide the following components: timeframe of the evaluation or the location. What is helpful is that it provides a glossary of key terms, which can be noted as good practice.
ii	Annexes include, if not in body of report: terms of reference, evaluation matrix, list of respondents, results chain/ToC/logical framework, list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology (e.g. inception report), additional details on ethical considerations, (country) case study reports, thematic papers or "deep dives."	Yes	The Annexes include the Terms of Reference (Annex 10); Evaluation Matrix (Annex 1); List of respondents (Annex 6); Data collection instruments (Annex 2); and List of documentary evidence (Annex 9). The ToC is presented in the main body of the report (p.11) and the list of sites is provided in the main report, such as the sampling. Other information included in the Annex is the Steering Committee engagement (Annex 3); Feminist principles in evaluation - lessons learned (Annex 5); Stakeholder mapping (Annex 7). The Annex also lists the Inception report in Annex 8, although it is blank in the version shared. However, it was shared a separate file for this assessment. Case studies are not included in the Annex but as standalone documents.
Question 18.	**Is the report logically structured and of reasonable length?**		
i	The report has a logical structure that is easy to identify and navigate (for instance, with numbered sections, clear titles, well formatted).	Partially	The report is well-structured, with a clear and logical flow that makes it easy to navigate. Section and sub-section titles are clearly formatted and consistently numbered, and the overall formatting is coherent throughout. The table of contents is comprehensive and contributes to ease of reference. However, one area for improvement relates to the table of contents for figures. It appears that the page numbers have not been updated. For instance, Figure 12 does not appear on p.43 and Figure 14 is not on p.58, as indicated. While this issue does not significantly hinder navigation, since the clarity of the figure titles still makes them easy to locate, correcting the page references would strengthen the overall usability of the report.
ii	Structure and length accords to UNFPA guidelines for evaluation reports; it does not exceed number of pages that may be specified in ToR.	Partially	The length of this report is 93 pages in total, excluding the Executive Summary, Annexes and Case Studies. This exceeds the 90-page page limit for centralised thematic evaluations commissioned by the IEO.
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	This report is well written, engaging and easy to understand. It is generally free from grammar, spelling and punctuation mistakes.
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.	Yes	This is a lengthy report that is fairly text heavy. However, evaluators have made an effort to break up the text through the use of visual aids using tables, figures and boxes. It therefore meets the minimum criteria for visual aids being used across the evaluation report. The figures and tables are clearly labelled and easy to understand. The figures are often introduced in the narrative to orient the reader e.g. Figure 13 (p. 49) and Figure 14 (p. 63).
SECTION I:	**CROSS-CUTTING ISSUES (weight 10%)**	93%	**Comments on Rating**
Question 20.	**Are cross cutting issues - in particular, human rights-based approach, gender equality, disability inclusion, LNOB - integrated in the core elements of the evaluation (e.g. evaluation design, methodology, findings, conclusions and recommendations)?**		
i	Evaluation's data collection methods designed to capture the voices/perspectives of a wide range of stakeholders including rights holders, marginalized and vulnerable persons, young people, people with disabilities, migrants or refugee populations, indigenous communities, and other persons that are often left behind.	Yes	The evaluation's data collection methods are designed to capture the voices and perspectives of a wider range of stakeholders - this is seen in the stakeholder mapping provided in Annex 7. While this is an evaluation of the integration of LNOB principles and reaching the furthest behind, rather than of specific impact on distinct population groups, the evaluators have sought to capture the perspectives of rights holders. This has mainly been through FGDs on p.17. The data collection tools are made available in Annex 2, with protocols for how the FGDs were conducted (p. 20 of Annex). Moreover, this evaluation involved individuals with an expertise or lived experienced as left-behind in the Evaluation Steering Group, which can be taken as good practice.
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	As this evaluation is specifically focused on a cross-cutting issue, with all evaluation questions and assumptions related to LNOB, this criterion is rated as 'yes'.

iii	Data is disaggregated by population groups (e.g. persons with disability, age, gender, etc.) where there are implications related to UNFPA's portfolio/interventions for these population groups; differential results are assessed (distribution of results across different groups), as feasible.	Yes	The evaluation provides disaggregated data by population groups, where relevant. However, the analysis could be strengthen through further disaggregation or clearer references to focus group discussions (FGDs) in the findings section. This is particularly as 306 individuals with lived experience were invited to participate in the evaluation, but their perspectives are limited in the main report. This said, the case studies, such as the one on the Latin America and Caribbean Regional Office (LAC) region, provide clearer references to FGD discussions disaggregated by country and stakeholder category such as indigenous women, adolescent girls, LGBTQIA+ persons, and persons with disabilities. In addition, as this is an evaluation of a normative principle, it is rated as meeting the criterion.
iv	Intersectional lens is applied in the data analysis, looking at various and multiple forms of exclusion and discrimination (and how they overlap with each other) and how this may impact the performance or results of the evaluand.	Yes	An intersectional lens is applied in the data analysis, looking at various and multiple forms of exclusion and discrimination. The report sets out UNFPA's intersectional approach in the context section of the report (p.9). Intersectionality is also explicit in one of the assumptions for EQ1. This is discussed more thoroughly in Section 4.9 (p. 30 onward) but also as part of other evaluation questions.
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, conclusions and recommendations all address cross-cutting issues, more specifically on LNOB as this is the focus of the evaluation.
vi	Inclusion of young people in the evaluation team and/or Reference Group.	Yes	This is rated as being met as the IEO confirm that young evaluators were included in the evaluation team. However, it would be helpful if this were made more explicit as it is difficult to determine from the title of the roles alone.
Question 21.	**Does the evaluation meet UN SWAP evaluation performance indicators?**	7	**Comments on Rating**
i	GEEW is integrated in the evaluation scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Satisfactorily integrated	While the evaluation objectives, scope and questions do not explicitly ask about HR and GE, they are incorporated through the LNOB lens. For this reason, this criterion is rated as 'satisfactorily integrated'.
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Fully integrated	The methodology for this evaluation employs a mixed-methods approach that is appropriate for evaluating LNOB. In addition, a diverse range of data sources were used with clear description of the process for triangulation. The methods described indicate meaningful participation of underrepresented groups. The sampling frame considers the diversity of stakeholders. In addition, ethical standards were considered through the evaluation and stakeholder groups were reported to be treated with integrity and respect for confidentiality.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	The evaluation conclusions and recommendations reflect a gender analysis. In addition, the evaluation team have adopted a 'gender plus' approach and gender discussions are evident throughout the findings section. The evaluation has a background section on intersectional analysis. The findings reflect a gender analysis although it could improve in how it demonstrates triangulation of the voices of different groups. The findings, conclusions and recommendations explicitly address gender and human rights dimensions. However, the recommendations are less explicitly about addressing GEWE issues but this is appropriate this for this evaluation.

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