

UNFPA Evaluation Quality Assessment Grid

Version: August 2025

REPORT RATING SUMMARY		
Overall Rating	93%	Highly Satisfactory
● ● ● ● ●	Excellent	5
● ● ● ●	Highly Satisfactory	4
● ● ● -	Satisfactory	3
● ● - -	Fair	2
● - - -	Unsatisfactory	1

REPORT DETAILS	
Title of the evaluation report	Mid-Term Evaluation of the UNFPA Supplies Partnership (2021-2030)
Region	Global
Country	Global (54 countries across UNFPA six regions)
Year of report	2025
Business Unit/programme country (managing evaluation)	Independent Evaluation Office
Date of assessment review (dd/mm/yyyy)	9/10/2025
Name of assessment review firm	IOD PARC
CLASSIFICATION OF EVALUATION REPORT	
Primary SDG(s) covered (list provided below)	3, 5, 17
UNFPA Strategic Plan areas covered (lists provided below)	
Three transformative results	1) Ending unmet need for family planning; 2) Ending preventable maternal deaths; and 3) Ending
Six outputs	
Six accelerators	All
Organizational effectiveness and efficiency	No
Humanitarian evaluation	Yes
Evaluation evaluand (e.g. country programme/intervention/policy/thematic area)	Programme
Evaluation type (e.g. formative, summative, developmental)	Summative and formative
Geographic scope (e.g. global, regional, national)	Global
EQA Summary: The rater will provide top line issues for this evaluation relevant for feedback to senior management (strengths and weaknesses), summarizing how the evaluation report meets or fails to meet all criteria. As relevant, the rater will highlight good practice/added value elements and the level of complexity of the evaluation. The rater should also highlight how cross-cutting issues were addressed in the report. Considerations of significant constraints (e.g. humanitarian crisis or political turmoil) should also be highlighted here.	
<p>This is a highly-satisfactory Mid-Term Evaluation of the UNFPA Supplies Partnership (2021-2030). The key strengths and areas for improvement are highlighted below:</p> <p>Key strengths</p> <ul style="list-style-type: none"> • Executive Summary: The Executive Summary is well-written and can serve as a standalone document that is useful for informing decision-makers. It is five and a half pages and includes all necessary components of an evaluation report. • Stakeholder mapping: A stakeholder map is included in the Annex of the main report, which is good practice. It clearly identifies a wide range of stakeholders including UNFPA, other UN agencies, government ministries, global partnerships, civil society, donors and rights holders. This serves as a good foundation for the sampling strategy. In addition, the mapping also outlines the purpose of each stakeholder group's engagement in the evaluation and how they are expected to use the evaluation report, although this was not systematically done (please see Annex 3, Table 2). • Purpose, objectives and scope: The evaluation report clearly describes the purpose, objectives and scope of the evaluation. Each is presented separately, and the information is explicit rather than inferred. • Evaluation matrix: The Evaluation Matrix is clear, well-structured and comprehensive. It provides an explanation for each evaluation question and makes explicit linkages to the theory of change (ToC). This is particularly valuable, as assessing the ToC is a key objective of the evaluation. Overall, the matrix is very well presented and can be considered an example of good practice. • Data analysis: An area of good practice is the rubric developed to assess the strength of evidence (Table 7 in Section 3.4.4). It would strengthen the report to see the results of this assessment in the findings section. • Conclusions: The conclusions are clearly articulated and provide a layer of analysis that goes beyond a summary of the findings. It is evident that the evaluators have stepped back to reflect and distil the findings. • Recommendations: Recommendations are clearly formulated and are derived from the conclusions. They are also useful and actionable – organised into three categories: strategic, programmatic and operational. Each recommendation is supported by 'operational implications'. <p>Key areas for improvement</p> <ul style="list-style-type: none"> • Data analysis: While the report cites contribution analysis in the methodology, its application is not explained. There is no reference in the data analysis section to contribution analysis, including how it was implemented, how the contribution story was developed, or how it was tested. • Theory of change (ToC): Overall, relevant assumptions are identified in the ToC, but improvements can be made to more comprehensively cover all three areas - causality, context, and stakeholder characteristics. • Ethical considerations: The evaluation report includes a dedicated section on ethical considerations for the evaluation. However, it does not detail the processes for ensuring data privacy, confidentiality, and safeguards. <p>Suggestions for future evaluators: The rater will identify key suggestions to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, examples will be cited to assist evaluation managers in overseeing future evaluations.</p>	

Recommendations <ul style="list-style-type: none"> • Data analysis: Where contribution analysis is applied, it would be helpful to outline the steps taken to develop the contribution story. Moreover, Section 3.1.1 does not appear to consider other influencing factors, which is a core element of contribution analysis. If an adapted version of contribution analysis was used, it would be helpful for evaluators to describe this explicitly. • Theory of change: The assumptions identified in the theory of change can be further strengthened by more comprehensively addressing causality, context, and stakeholder characteristics. • Ethical considerations: It is important to include data collection protocols i.e. preambles or instructions, that outline how data collection is designed to ensure confidentiality and consent is obtained. Such protocols would help demonstrate how evaluators actively recognised and addressed potential power imbalances during individual and collective data collection (for example, focus groups) to ensure equitable participation and consider the voices of all relevant stakeholders (p. 21). In addition, evaluators are also encouraged to provide informed consent forms, where used. 			
SECTION RATINGS			
SECTION A:	EXECUTIVE SUMMARY (weight 5%)	100%	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 well-written and can serve as a standalone document that can be useful for informing decision-making. At five and a half pages, it meets the required length.
	ii Includes all necessary components of the evaluation report, including: (1) overview of the context and intervention, (2) evaluation purpose, objectives and intended users, 3) scope and evaluation methodology, (4) summary of most significant findings, (5) main conclusions and (6) key recommendations	Yes	The Executive Summary includes all necessary components of the evaluation report. This includes: (1) Overview of the context and intervention (provided under 'Background' on p. 10); (2) Evaluation, purpose, objectives and intended users (under 'Purpose and scope of the evaluation on p. 10); (3) Scope and evaluation methodology (under scope and methodology on p. 10); (4) Summary of the most significant findings (under 'Main findings on pp. 10-13); (5) Main conclusions (pp. 13-14); and (6) Key recommendations (pp. 14-15).
	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.	Yes	It includes all significant information needed to understand the evaluand, evaluation methodology and results. It is written in a clear and concise manner.
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	<p>A description of the UNFPA Supplies Partnership phase III (the evaluand) is provided in section 2.4 (pp. 5-8). This includes the geographic coverage - 54 countries across Africa, Asia, the Pacific, and Latin America and the Caribbean (p. 5), with further description of the country selection provided in section 2.4.2 (p. 6). The implementation period is 2021-2030 (p. 5). In addition, the preceding section describes the evolution of the Supplies Partnership from 2004 onward, showing the work that this third phase builds on.</p> <p>The main partners are presented in Table 2, which summarises the governing bodies of the UNFPA Supplies Partnership (p. 7). It involves the main constituents of the partnership, including donors, governments, civil society, and strategic partners. A summary of donor contributions and budget utilised for 2021-2023 is provided (pp. 7-8), along with the projected funding gap for 2026-2030 and the risks of a funding shortfall.</p> <p>As all the required components are included in the Evaluation Report, it is rated as met. One suggestion for evaluation teams to consider is incorporating the description of the Supplies Partnership Phase II from the Inception Report (pp. 11-16 of the IR) to the Annex of the main report. The IR offers a more in-depth description of the intervention's components and rationale in a way that provides a clearer picture. While this may not be necessary for most primary users, it would be helpful for secondary audiences.</p>
	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	<p>The context of the Supplies Partnership is provided in Section 2 (pp. 2-5), which presents the global context for reproductive health and the global supply chains. While it acknowledges relevant social, economic and political factors, it is relatively limited. The Inception Report (IR), however, offers a clearer description of the key drivers affecting access to family planning services. In addition, the evolution of the programme in the IR (pp. 11-16) also provides a helpful context and rationale for the intervention, particularly in relation to governance and the renewed emphasis on country ownership.</p> <p>With regard to the institutional context, the evaluation report highlights links to UNFPA's Strategic Plan and UNFPA's Family Planning Strategy (pp. 5-6), and references core principles such as Leaving No One Behind (LNOB), the Human Rights-Based Approach (HRBA) and the Gender-Transformative Approach (GTA) (p. 6).</p> <p>Cross-cutting issues are also identified in the background section, which highlights disproportionately affected stakeholder groups, including 'adolescents, unmarried women, the urban poor, people with disabilities and sex workers' (p. 4). The "last mile," a key element of both the Partnership and the evaluation, is also mentioned. To enhance clarity, evaluators are encouraged to provide a description of what the "last mile" means in this context, either within the body of the report or in a footnote. However, this does not affect the rating.</p>
	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	The background section outlines the links between the Partnership and the Sustainable Development Goals (SDGs) and ICPD (p. 6). The SDG targets mentioned are 3.3, 3.7, 3.8, 5.6, 17.16 and 17.17. Specific indicators are also not explicitly mentioned but data related to unmet need for family planning is provided (p. 3). The Inception Report (IR) provides more detail including unmet needs at global and regional level (pp. 3-4 of IR).
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	<p>The main report identifies stakeholder categories, including UNFPA, other United Nations agencies, government ministries, global partnerships, civil society, and donors (p. 2). It also signposts to a stakeholder map included in the Annex of the main report, which is considered good practice.</p> <p>The stakeholder mapping is provided in Annex 3.1. It lists the broader stakeholder groups, such as UNFPA, and includes a supporting column that specifies sub-groups. Rights holders (noted as 'end-users/communities') are also identified, including women of reproductive age, adolescents, men and marginalised groups. The only exception is academic institutions, for which sub-groups are not specified.</p> <p>The introductory paragraph to Annex 3 explains that, at the country level, discussions were held with UNFPA country office focal points to compile a full list of stakeholders. However, this list is not reflected in the stakeholder mapping. It also notes that the stakeholder mapping was adjusted during data collection to "ensure all relevant perspectives were captured" (Annex 3, introductory paragraph).</p> <p>In addition, the evaluation team provides a mapping of partnerships, programmes and key actors involved in family planning and reproductive health commodity security, which were identified during inception (Table 3 of Annex 3).</p> <p>Overall, the stakeholder map provides a clear identification of key stakeholders, providing a good basis for sampling.</p>
	ii Stakeholders are analysed to understand their interests and needs, power and influence, and potential impact on the evaluand.	Partially	<p>The stakeholder map provided in Annex 3 provides some analysis of some of the stakeholders. However, this is primarily limited to their engagement in the evaluation (see column titled 'purpose of engagement in the evaluation' in Table 2 of Annex 3). This information is not provided for all stakeholder groups. In addition, only two stakeholder groups (UNFPA and 'Other UN agencies') include a description of intended use of the evaluation.</p> <p>To fully meet this criterion, evaluators can be encouraged to consider and/or document their analysis in the following areas:</p> <ol style="list-style-type: none"> 1. Interests and needs: Understanding stakeholders' concerns, priorities, and how the intervention may impact them; 2. Power and influence: Assessing the level of power or influence stakeholders have over the intervention and how their support or opposition could affect its success; and 3. Potential impact: Analysing how stakeholders may affect or be affected by the intervention, both positively and negatively.
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	<p>The purpose of the evaluation is clearly stated and described in sub-section 1.1 (p. 1). It is aligned with the purpose outlined in UNFPA's Evaluation Policy (2024), particularly in aiming to generate evidence-based insights to guide the Partnership through the remainder of phase III (p. 1). The report also clarifies its intended use i.e. to inform "programme priorities, strategic and operational frameworks, and guiding principles" (p. 1).</p> <p>Although the report does not explicitly explain why the evaluation is needed at this point, this is implicit in its status as a mid-term evaluation and that it is intended to guide the remainder of the current phase.</p> <p>The key intended users are identified in sub-section 1.3 (pp. 1-2), with references to the stakeholder map in Annex 3, which further specifies the intended users among the primary audiences.</p>
Question 5.	Are the objectives and scope of the evaluation clear and realistic?		
	i Clear and complete description of the objectives of the evaluation, including reference to any changes made to the objectives included in the inception report (if applicable).	Yes	<p>The objectives of the evaluation are provided in sub-section 1.1 (p. 1). Where the purpose and objectives are presented in the same sub-section, they are distinct. The objectives are clear and are consistent with those presented in the Inception Report (IR) (p. 1 of IR).</p>
	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	<p>The scope of the evaluation is defined in sub-section 1.2 (p. 1). It outlines the temporal scope (2021 to 2024), the geographic scope (54 partner countries, with the number of countries listed by region and accompanied by a map of participating countries in figure 1), and the focus areas (technical interventions across the four Strategic Objectives and their associated outcome areas). There are no changes to the scope compared with what was presented in the Inception Report (IR) and Terms of Reference (ToR), except that it specifies 2024 as the endpoint, whereas the IR refers to '2021 to present.' The report does not mention any areas that fall outside the scope.</p>
SECTION D:	EVALUATION DESIGN AND METHODOLOGY (weight 20%)	92%	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	<p>The evaluation questions and underlying assumptions are presented in Section 3.2 (pp. 15–16). Each main evaluation question is aligned with the relevant OECD/DAC criteria, namely: Relevance; Effectiveness; Coverage; Efficiency; and Coherence (pp. 15–16). Assumptions for each evaluation question are clearly identified. The Inception Report (IR) provides a rationale for the selection of the evaluation criteria, which includes the humanitarian criterion of coverage.</p> <p>In addition, the evaluation considers two key cross-cutting issues within the evaluation questions: gender equality and social inclusion, and humanitarian response (pp. 16–17). Overall, the evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation.</p> <p>The Inception Report (IR) provides a rationale for the selection of the evaluation criteria, which includes the humanitarian criterion of coverage. It would be helpful for this information to be incorporated into the main report.</p>
	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.	Yes	<p>The evaluation matrix is provided in Annex I. It is clear, very well structured, and includes all six key elements: evaluation criteria, evaluation questions, assumptions, indicators, data collection methods, and data sources.</p> <p>It is organised by the main evaluation question (EQ). For each EQ, the relevant OECD/DAC criterion or criteria are identified, along with the area of interest and linkages to the theory of change. This linkage is helpful as assessing the adequacy of the theory of change is a key objective of the evaluation. The rationale for each EQ is also provided, summarising the focus of the analysis. The key assumptions for each EQ are presented, and each assumption is supported by corresponding indicators for verifying the assumptions, data collection methods, and data sources. The data sources specify the categories of the relevant informants.</p> <p>Overall, this is a very well-presented matrix and can be considered an example of good practice.</p>
Question 7.	Is the theory of change, results chain, logical framework, or equivalent framework well-articulated?		
	i Clear description of the intervention's intended results, or of the 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	<p>There is a clear description of the Supplies Partnership's intended results and the specific components being examined in the evaluation in Section 3.1 (pp. 9–13). As this is a theory-based evaluation, the intervention is presented through a Theory of Change (ToC). The ToC is presented as a visual (Figure 6, p. 12) with an accompanying narrative (pp. 9–13).</p>
	ii Critical assumptions underlying the ToC, results chain or logical framework are clearly identified, including assumptions about causal relationships, contextual factors and stakeholder characteristics.	Yes	<p>Critical assumptions underlying the Theory-of-Change (ToC) are identified in the visual itself (see the assumptions key for Figure 5, p. 12) and they cover all three areas - causality, context, and stakeholder characteristics. Although relevant assumptions are identified, there is an imbalance in coverage, with a primary focus on causal relationships (for example, see assumption 1, 2, 3, 4, 6 and 7) and stakeholder characteristics (for example, see assumption 6, 7, 10 and 11). Furthermore, many aspects are overly consolidated into a single, complex assumption.</p> <p>Overall, relevant assumptions are identified but improvements can be made to more comprehensively address all three areas and to better delineate them, where feasible.</p>
	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.	Yes	<p>The evaluation report provides a fairly comprehensive assessment of the theory of change, refining it from the version presented in the ToRs (see Annex A of ToR). The report describes the process of refining the ToC in section 3.1.4 (p. 9), which involved desk review, stakeholder feedback, and validation. It also describes the main purpose for reconstructing the ToC on p.10. What is particularly helpful is the links between the causal assumptions of the ToC with the areas being assessed in the EQs (see Evaluation Matrix in Annex 1 and p. 18 of the IR).</p>
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	<p>This evaluation is designed as a theory-based using a contribution analysis approach, with the description and rationale provided in sub-section 3.1.1 (p. 9). There are some questions about how contribution analysis was employed but this is covered further below under Q8, criterion v of this grid. The evaluation employs a mixed-methods approach, incorporating both quantitative and qualitative data (p. 17). Quantitative methods included closed survey questions, financial data analysis, and outcome trend analysis. Qualitative methods included key informant interviews (KIIs), group discussions, document reviews, open-ended survey questions, field visits, and observations (p.17). The evaluation design and methods are appropriate and sufficiently robust to address the evaluation's purpose, objectives, and scope. Artificial intelligence (AI) was not used in the evaluation process.</p>

	ii Data sources are all clearly described and are relevant and robust; these would normally include qualitative and quantitative sources.	Yes	<p>As noted above, the evaluation employed a mixed-methods approach (p. 17). Secondary data sources, described on p. 17, include a wide range of programme, strategic, and operational documents at global, regional, and country levels, as well as evaluation reports, national plans and programmes, and other relevant literature. The evaluators also reviewed national data on reproductive health and family planning. In total, 194 documents were consulted, with the full list provided in Annex 10.</p> <p>Primary data sources are described on p. 18 and include key informant interviews (258 informants), group discussions at the country level with facility staff, national health officials, and other key organisations, as well as a survey. The list of 258 consulted informants is provided in Annex 9. Observations are briefly mentioned on p. 17 but are not described. It would be beneficial to elaborate on these, particularly if they were documented and informed the analysis; otherwise, it may be appropriate to remove the reference. Overall, while the data sources are adequately described.</p>
	iii Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or if not, provide reasons for this).	Yes	<p>Section 3.4.3 provides a summary of the sampling strategy across four areas: (i) country case studies; (ii) key informant interviews; (iii) group discussions; and (iv) the survey. The level of detail provided for most components is appropriate given the purposive nature of the sampling. The sampling for both the survey and the key informant interviews/group discussions is well-defined. Annex 3.4 clearly describes the survey as a purposive sample drawn from distinct stakeholder categories (noting the relationship to the Supplies Partnership and desired sample sizes per group). Similarly, the sampling for interviews and group discussions is grounded in the stakeholder map (which serves as the sampling frame), ensuring data collection targets the most influential stakeholders. Annex 3.2 provides detailed information on the selection of the country case studies, which involved a two-step process with defined selection criteria. However, it is unclear how the initial list of 22 countries, which met the criteria, was subsequently narrowed down to the final case study list, as the report refers only to a broad 'consultative process with UNFPA' without detailing the final criteria or rationale for exclusion. However, this minor omission does not affect the rating.</p> <p>To meet the highest level of transparency, it is important for evaluators to provide a clear description of all aspects of the sampling process, including the criteria guiding consultations with UNFPA.</p>
	iv Methods allow for rigorous testing of the theory of change, results chain or logical framework (e.g. methods help to understand the causal connections, if any, between outputs and expected outcomes).	Yes	<p>As a theory-based evaluation, the methods outlined allow for testing the causal connections depicted in the theory of change. This is supported by the evaluation team's reconstruction of the ToC, which makes explicit some of the underlying assumptions and causal connections between inputs, outputs and expected outcomes.</p>
	v Clear and complete description of the methods of analysis, including explainability and full disclosure of the use of AI in the evaluation process, if applicable.	Partially	<p>Data analysis is described in Sections 3.4.4 and 3.4.5 (pp. 19–20), and the report notes that AI was not used in the evaluation (p. 1). The data analysis sub-section states that data was coded using the evaluation matrix (p. 19), and that findings were mapped against the evaluation assumptions. While there is a general overview of the process, the application of contribution analysis is not explained. There is no reference in the data analysis section to contribution analysis, how it was implemented, or how the contribution story was developed or tested. In addition, the section signposts Annex 3 for further detail, but no additional information was found in the annex.</p> <p>To fully meet this criterion, evaluators can be encouraged to more clearly describe how the evaluation matrix was used for data processing, analysis, and interpretation. It would also be helpful to outline the steps taken to develop the contribution story, in line with the contribution analysis approach described in Section 3.1.1 (p. 9). Moreover, Section 3.1.1 does not appear to consider other influencing factors, which is a core element of contribution analysis. If an adapted version of contribution analysis was used, it would be useful for evaluators to describe this explicitly.</p> <p>An area of good practice is the rubric developed to assess the strength of evidence (Table 7). While this does not affect the score, the section (3.3.4, p. 19) could be further strengthened by explaining how the strength-of-evidence ratings were applied. For example, whether 'poor' evidence was excluded or included with qualifications. To enhance transparency, it would also be useful to indicate the strength of evidence for each finding in the findings section.</p>
	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	<p>There is a dedicated section on limitations in Section 3.8 (pp. 21–22). The limitations are clearly described, with several additional limitations identified beyond those presented in the Inception Report. Each limitation is accompanied by a corresponding mitigation measure implemented by the evaluation team. While these are sufficient to meet the criterion, to further strengthen this section, evaluators can consider applying a Red/Amber/Green (RAG) rating to each limitation and/or risk, followed by an indication of the residual risk or limitation after mitigation measures were applied. This would enhance transparency and provide a clearer assessment of the extent to which each limitation may still affect the findings.</p>
Question 9.	Are ethical issues and considerations described? The evaluation should be guided by the UNEG ethical standards for evaluation. As such, the evaluation report should include:		

	i Explicit and contextualized reference to the UNEG obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) and/or UNEG Ethical Principles.	Yes	<p>The evaluation report explicitly references UNEG Ethical Guidelines for Evaluation and the UNEG Code of Conduct for Evaluation, and the evaluation's adherence to these ethical and professional standards. It also provides brief, yet contextualised, references to key ethical principles such as independence, impartiality, integrity, credibility, and accountability (Section 3.7, p. 21).</p> <p>For example, regarding credibility, the evaluators contextualise their adherence by stating that they applied a 'rigorous methodology, maintained impartiality, conducted robust quality assurance, ensured transparency, and engaged stakeholders' (p. 21). They further note that their 'findings were evidence-based, with supporting data triangulated across multiple sources' (p. 21).</p> <p>In some areas, the discussion is more high-level. For instance, with independence and impartiality, evaluators note that 'findings and recommendations are evidence-based and unbiased' (p. 21). This would merit elaboration on steps taken to ensure that the evidence was unbiased. Nevertheless, the evaluators have made a clear effort to contextualise ethical principles to the evaluation.</p>
	ii Clear description of ethical issues and considerations (e.g. respect for dignity and diversity, fair representation, confidentiality, and avoidance of harm) that may arise in the evaluation, safeguard mechanisms for respondents (e.g. parental consent forms for adolescents, compliance with codes for vulnerable groups; WHO standards of safe data collection on GBV) and ethical considerations in the use of AI as applicable (e.g., transparency of use, explainability, privacy, data protection, accuracy, human rights). If AI is used in the evaluation, there should be transparency and disclosure on the ethical and responsible use of AI in the report.	Partially	<p>The evaluation report includes a dedicated section on ethical considerations for the evaluation, which is considered good practice (Section 3.7, p. 21). This section includes a description of areas such as respect, beneficence, data privacy and confidentiality, and safeguard mechanisms for respondents (p. 21). It is noted that artificial intelligence was not used in the evaluation (p. 21).</p> <p>While this information is useful, the report does not detail the processes for ensuring data privacy, confidentiality, and safeguards. Evaluators are encouraged to describe these processes, either in the main body of the report or in the Annexes.</p> <p>Additionally, the data collection protocols/guidance, aside from the survey questionnaire, are not included in the Annex. Annex 3.3 is titled 'Key Informant Interview Protocol', but it does not provide preambles or instructions for data collectors regarding confidentiality and consent. Such protocols would help demonstrate how evaluators actively recognised and addressed potential power imbalances during individual and collective data collection (for example, focus groups) to ensure equitable participation and consider the voices of all relevant stakeholders (p. 21). In addition, evaluators are also encouraged to provide informed consent forms, where used.</p> <p>Furthermore, the evaluation report does not present how ethical considerations were addressed throughout the evaluation process from the design phase to the reporting, dissemination, and facilitation of use phases. This needs to be inferred from the text, and it would be better to present a table outlining how ethical issues were addressed in each evaluation phase.</p>
Question 10. Does the evaluation incorporate innovative practice that adds value to the evaluation process?			
	i Innovation practice is used to improve the quality of evaluation process. This could include efforts to optimize the evaluation process (e.g., use of AI or new technology for data gathering, content analysis, outcome harvesting among others), or components introduced to enhance inclusion and participation in the evaluation processes (e.g. a youth steering committee), or ways of sharing of evaluation results.	Not Rated	The evaluation applies methodology that is thoughtful and relevant for the evaluation, but is not considered innovative practice. In addition, innovation or innovative practices were not required as per the Inception Report, which is why this criterion is not rated.
SECTION E: EVALUATION FINDINGS (weight 25%)		88%	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	<p>While the findings section is not structured by evaluation question, this is acceptable as it is an IEO-led evaluation and does not affect the rating. The logic of the structure is explained in a box (unnumbered) at the beginning of Section 4 (p. 23). The findings are organised into six sections, most of which correspond to at least one of the strategic objectives. This alignment makes it easy to understand how the findings relate to different parts of the Theory of Change, which supports one of the evaluation objectives (p. 16). In addition, a column indicates the assumption numbers (linked to each evaluation question) covered within each section.</p> <p>Nevertheless, aspects of the guidance in the Evaluation Handbook are still applied with each section beginning with a summary box followed by the numbers of the assumptions assessed. The findings are numbered and broadly correspond to the assumptions. For example, Finding 1 aligns with Assumption 1.1, while Finding 4 aligns with Assumptions 1.4 and 6.3. However, this requires careful reading and cross-referencing to confirm that all assumptions are fully addressed. Overall, the main evaluation questions are covered, and Annex II presents the evaluation matrix alongside the supporting evidence, which enhances transparency and traceability.</p>
	ii Explicit use of the evaluand's theory of change, results chain, logical framework in the formulation of the findings.	Not Rated	While explicit use of the Theory of Change (ToC) is not required, the report nonetheless ensures traceability, particularly as the evaluation questions are broadly aligned with the strategic objectives identified in 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.	Partially	<p>The evaluation presents credible qualitative and quantitative data. The full list of secondary sources is provided in Annex 9, and the list of key informants is included in Annex 6. In addition, each country case study annex contains a list of persons consulted and a bibliography. Triangulation is evident, as each finding is supported by data from multiple sources, as demonstrated by the footnotes in nearly all pages of the findings section. These sources include different types of documents (e.g. evaluation reports, strategic plans, annual and programme reports) and multiple data collection methods (e.g. secondary sources, KIs, and surveys).</p> <p>However, there is scope for improvement. For example, references to KIs while included are limited. In addition, divergent views and the strength of evidence are not always made explicit. In Section 4.2.1.2 (p. 35), for instance, challenges are presented, but it would help to indicate whether particular stakeholder groups raised them more than others. While this is only one example, similar improvements could be made across the section. This said, it should be noted that some case studies provide more sources for primary data (especially references to KIs).</p> <p>Another consideration is to include the strength-of-evidence rating for each finding (as described under data analysis, p. 20), which would also improve transparency and credibility.</p>
	ii Findings are clearly supported by the evidence presented, both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Yes	The report contains 21 finding statements that present a mix of positive, negative and nuanced findings. Each finding statement is followed by supporting evidence. However, because the report is not structured according to the Evaluation Matrix, it is less straightforward to cross-reference findings with the corresponding evaluation assumptions. To enhance traceability, one suggestion is to explicitly signpost each finding to the relevant assumptions. Nonetheless, it is evident that some of the assessment indicators from the matrix were applied when analysing the assumptions.
	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	The evaluation report identifies and discusses the causal factors leading to, in this case, the strategic objectives. For example, section 4.4 discusses the causal pathways leading to Strategic Objective no. 2 (p. 46 onward). It explores the enabling factors as well as challenges. Similarly, as another example, section 4.5 discusses the causal pathways leading to strategic objective no. 3 (p. 52 onward). In this case, looking at Partnership's role in increasing and diversifying financial and programmatic contributions and prioritising reproductive health. This section provides a description of the current state of play and also discusses the challenges.
Question 13.	Does the evaluation assess and use the intervention's Results Based Management elements?		
	i Assessment of the adequacy of the intervention's planning, monitoring, and reporting system (including completeness and appropriateness of results/performance framework - including vertical and horizontal logic, M&E tools and their usage) to support decision-making.	Not Rated	Assessment of the intervention's planning, monitoring, and reporting system is not mentioned in the Inception Report.
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 clearly articulated and provide a layer of analysis that goes beyond a summary of the findings. It is evident that the evaluators have stepped back to reflect and distil the findings. There are 11 conclusions, organised into five broader clusters, such as Evolution of the Design (Cluster A), Integration and Coordination (Cluster C), Financial Sustainability (Cluster D), and so on. Each conclusion presents a concise and impartial judgement, followed by a more detailed narrative. Some conclusions would benefit from being slightly longer, as a single sentence sometimes misses nuances that are then captured in subsequent sentences (see, for example, conclusion 2 and 7). However, this does not affect the overall rating. Each conclusion includes signposts to the relevant evaluation question number and the corresponding section in the findings.
	ii Conclusions are well substantiated and derived from findings.	Yes	The conclusion statements are supported by a narrative that summarises the analysis of the evidence. Each conclusion includes a signpost to the relevant evaluation question number and section of the report. One suggestion is that, as numbered findings are provided in the report, evaluators can consider using the numbers for sign-posting rather than the evaluation questions and sections within the report. However, this does not affect the score.
Question 15.	Are lessons learned identified?		
	i Lessons learned are derived from the findings and are well substantiated with practical, illustrative examples.	Not Rated	A section on lessons learned is not explicitly requested in the Inception Report.
	ii Lessons learned are clearly presented and provide actionable insights on the positive aspects of the evaluand as well as any areas of	Not Rated	A section on lessons learned is not explicitly requested in the Inception Report.
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	Recommendations are presented in Section 6 (pp. 78-81). They span four pages in total, which falls within the required maximum 6 pages length in the Evaluation Handbook (2024). The recommendations are clearly formulated in unambiguous language. The recommendation is clearly stated at the top, followed by the rationale for the recommendation. They provide the links to the relevant conclusion/s. They also provide the specific actions to be taken, which are labelled as 'operational implications'.

	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. They are organised into three categories: strategic, programmatic and operational. For each recommendation, a list of supporting actions titled 'operational implications' is provided. Each recommendation is directed at a specific unit, such as SRHB or SCMU. While this evaluation provides sufficient guidance for action, it may also be beneficial to identify both the lead unit and the supporting units at the level of recommendations and operational implications. Evaluators can consider aligning the structure of the recommendations, to the extent possible, to the requirements of the management response template for centralised evaluations, including 'lead implementation unit', 'due date', and 'other units involved in implementation'. However, while these suggestions can clarify and strengthen evaluation reports, they do not affect the score.
	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	Although the process for developing the recommendations is not described in Section 6 (p. 78), it is outlined in Section 3.5 under 'Generation of Recommendations' (p. 20). The evaluation report states that the evaluation team and the evaluation manager engaged in a sense-making process to formulate initial ideas for recommendations. The final recommendations were then developed through a co-creation process with members of the Evaluation Reference Group (ERG) and the Partnership team. It may be worth noting that, while the ERG members are listed in the opening pages of the report, it is not clear which members of the Partnership team participated in the co-creation of the recommendations.
	iv Recommendations are prioritized based on their importance, urgency, and potential impact.	Yes	All six recommendations are assigned a 'high' priority rating. While this is acceptable, it may be useful to consider whether applying the same rating to all recommendations provides sufficient differentiation to help the intended audience prioritise actions when taking them forward. In such cases, it may be helpful to include timelines ('due date', as per the management response requirements for centralised evaluations).
SECTION H:	REPORT STRUCTURE AND PRESENTATION (weight 5%)	100%	Comments on Rating
Question 17.	Does the evaluation report include all required information?		
	i Opening pages include: Name of evaluation and/title of evaluation, timeframe of the evaluation, date of report, location of evaluand, names and/or organization(s) of the evaluator(s), name of the evaluation manager, name of organization commissioning the evaluation, list of the evaluation reference group members, acknowledgements, table of contents (including, as relevant, tables, graphs, figures, annexes), and list of acronyms/abbreviations.	Yes	The opening pages includes nearly all the required elements. The only missing element is the timeframe of the evaluation. However, the timeframe of the evaluated intervention and date of the report is 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), additional details on ethical considerations, (country) case study reports, thematic papers or "deep dives."	Yes	The annexes include nearly all of the required elements. It would be helpful for evaluators to include the group discussion questions and observation checklists, or other tools used for observations during field visits. Evaluators are also encouraged to include the full data protocols i.e. the preambles/guidance for how the data collection was conducted, as these help demonstrate how the evaluation was conducted, as noted above under ethical considerations (see Q9, criterion ii of the grid).
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. All main sections are numbered and clearly titled, and clear headings and consistent formatting are used throughout. The table of contents is comprehensive, enabling intended users to navigate the report with ease. The only exception is the annexes, which appear to compile separate documents into one volume. Evaluators can be encouraged to provide continuous page numbering for the annexes to improve navigation. However, as these are supplementary documents, it does not affect the score. Overall, the sections of the report flow in a logical and coherent order.
	ii Structure and length accords to UNFPA guidelines for evaluation reports; it does not exceed number of pages that may be specified in ToR.	Yes	The length of the report is 81 pages, excluding the opening pages and Executive Summary. This falls with the 90-page limit for 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	The report is written in a clear and accessible way. 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	The report makes frequent use of visual aids, including maps, tables and figures, to convey key information. These are clearly presented, well labelled, and appropriately referenced in the text. They are also listed in the table of contents. There are a few minor exceptions, such as boxes or tables that are not referenced in the text (e.g. Figure 7, p. 17) or boxes that are unnamed or untitled (e.g. the first box on p. 23). However, these are few and minor and do not warrant a 'partial' rating.
SECTION I:	CROSS-CUTTING ISSUES (weight 10%)	85%	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.	Partially	This is rated partial as the full data collection instruments have not been included in the Annex. However, the 'key informant interview protocol' in Annex 3.3 includes a question that asks how UNFPA support affects left-behind groups (for example, see Q5 on p. 11, Q7 on p.13, and Q13 on p. 18). In addition, the survey includes questions about last-mile delivery and reaching remote health facilities, however, the questions are limited (Annex 3.4).
	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	Cross-cutting issues are integrated into at least half of the evaluation questions. For example, evaluation question 1 considers the Partnership's contribution to addressing the needs of women in its design; evaluation question 2 examines availability and choice for women, including marginalised groups and those in humanitarian settings; and evaluation question 3 focuses on ensuring that reproductive health commodities reach the last mile. While the criterion is rated as being met, it is worth considering for future evaluations, a focus on environmental standards. Given that this evaluation concerns supplies, this would enable a closer look at environmental considerations such as the extent to which the Partnership's adoption of measures to minimise the carbon footprint of procurement, shipment and last-mile distribution, and of local suppliers.
	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	There is limited disaggregation by population groups, although this is not required for this evaluation. However, the incorporation of country-level findings from the seven case studies is helpful. The report meets the minimum requirement to disaggregate stakeholders by gender, which is presented in a table in the main report by data source (p. 18).
	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 evaluation.	Not Rated	It is not clear whether an intersectional lens was applied in the data analysis. If this was done, particularly in relation to LNOB and 'the last mile', it would be helpful for the evaluators to make this more explicit. However, as this is not a requirement for the evaluation, the score is not rated.
	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 address cross-cutting issues such as LNOB and reaching the last mile.
	vi Inclusion of young people in the evaluation team and/or Reference Group.	Yes	The ToR does not explicitly request the inclusion of a young evaluator (see the proposed team composition on p. 39 of the ToR). However, the team included a junior evaluator to support data collection and analysis (data analyst researcher).
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	The evaluation includes questions that explicit focus on human rights and gender equality (see, for example, EQ1 and EQ2). The evaluation also includes a gender and rights-based analysis, as well as indicators to ensure that GEEW-related data will be collected (see, for example, indicators for assumption 1.4).
	ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated	There is some evidence that a gender-responsive methodology is considered in the Inception Report, but this is less evident in the main report. The data sources are diverse and the evaluation draws on the use of mixed methods data collection, which is appropriate for evaluating GEEW considerations. Ethical standards are contextualised, although they could be strengthened by incorporating more information such as protocols and/or preambles for the conduct of data collection.
	iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Fully integrated	The findings, conclusions and recommendations reflect a gender analysis, particularly on LNOB and reaching the last mile, which is appropriate 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