UNFPA Evaluation Quality Assessment Grid

Version: May 2024

REPORT RATING SUMMARY		
Overall Rating	67%	Satisfactory
● ● ● ● ● Excellent	5	
● ● ● ● Highly Satisfactory	4	
● ● ● – Satisfactory	3	**The report meets UNFPA/UNEG standards for evaluation reports, but some indicators are inadequately addressed or missing. Decision makers may use the evaluation with some confidence.**
● ● – – Fair	**2**	
● – – – Unsatisfactory	1	

REPORT DETAILS	
Title of the evaluation report	Formative Evaluation of the Arab States Regional Programme 2022-2025
Region	ASR
Country	Regional
Year of report	2024
Business Unit/programme country *(managing evaluation)*	ASRO
Date of assessment review (dd/mmm/yyyy)	October 14, 2024
Name of assessment review firm	IOD PARC
CLASSIFICATION OF EVALUATION REPORT	
Primary SDG(s) covered *(list provided below)*	**1, 3,5 10, 13, 16, 17**
UNFPA Strategic Plan areas covered *(lists provided below)*	
Three transformative results	yes
Six outputs	yes
Six accelerators	yes
Organizational effectiveness and efficiency	Yes
Humanitarian evaluation	No
Evaluation evaluand *(e.g. country programme/intervention/policy/thematic area)*	Strategy
Evaluation type *(e.g. formative, summative, developmental)*	Formative
Geographic scope *(e.g. global, regional, national)*	Regional

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.*

Overall, the evaluation has some good quality elements and potentially very useful findings. What significantly weakens the report, however, is the limited reference to supporting or corroborating evidence that would assure the reader that the analysis presented is well-considered and grounded in triangulated data, rather than solely reflecting the opinions of a few stakeholders or the evaluators themselves. Additionally, several potentially valuable findings are not fully explored and are addressed only briefly, which could lead to inaccurate conclusions. The report's forward-looking focus, as noted in section 1.3 para 32, is a strategic choice but may have contributed to a lack of clarity, particularly in assessing progress against indicators or targets. Moreover, the absence of a clear "big picture" analysis—connecting individual data points (such as specific country examples) to the broader regional context—detracts from the overall coherence. In many instances, findings resemble a list of activities that may have been internally reported, without sufficient effort to examine the outcomes or impact of those activities among diverse stakeholders. Some specific strengths are as follows:
• The exec. summary is comprehensive and covers all elements of the evaluation effectively.
• The report is reasonably well laid out and formatted, per UNFPA guidelines, with visuals adding to the overall presentation.
• A clear description of the purpose, objective, scope and background/context.
• The report is thorough in its addressing of all EQs and subquestions, despite being slightly lengthy.
• Acceptable use of multiple methods for data collection such as interviews, group discussions, secondary survey data, document review for a good mix of primary/secondary and qualitative/quantitative data sources.
• Conclusions and recommendations are well-grounded in the evaluation findings.

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.*

Most significantly, the evaluation required a thorough review of the analysis under all findings to ensure that both the quantity and quality of the supporting evidence were sufficient and clear to the reader. Triangulation between sources was crucial to enhance the credibility and reliability of the findings and sub-findings. Additionally, many potentially interesting findings needed further unpacking or analysis—some of which were noted in the in-line commentary, although a review by the evaluation manager could have identified more. Other, more specific areas for improvement include the following:
• Some acronyms and abbreviations lack clear explanations.
• Formatting inconsistencies and minor grammar, syntax, and proof-reading errors throughout the document. The report should have been edited in line with UNFPA style guidelines.
• The evaluation matrix introduces an additional layer of assumptions that (a) do not analyse the sub-questions in any greater detail (they are a simplification) and not noted in the main report and thus not clearly reflected in the analysis.
• The description of the methodology in the main report and annexes is overly dense and the language not straightforward or concise, masking key weaknesses/inconsistencies with the description of the approach, such as around the use of AI (and the ethical considerations in its use) and lack of details on sampling approaches.
• The writing style in the report is challenging in places, with over-use of passive voice, superlative language/use of buzzwords.
• The correlation between the findings and the recommendations provided is not always strong with some potentially useful sub-findings not analysed fully nor reflected in the conclusions/recommendations.
• In the annexes, much of the text has the incorrect tense (future, rather than past or present).

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). *Note: YES - the executive summary is within the indicated maximum page limit. PARTIAL - the executive summary exceeds the maximum page limit by 1 to 2 pages. NO - the executive summary exceeds the maximum page limit by more than 2 pages.*	Yes	The structure of the Executive Summary is sound, and it is of a sufficient length (7 pages) to be a standalone document. It is clearly presented, although the "findings" are more-or-less restatements of the narrative accompanying the conclusions (which, as noted below, are in turn summaries of the findings, so apposite in this context). There are some style/grammar/spelling/punctuation errors which should be addressed as a priority. The summary (as with the main report) requires a thorough proofreading by a professional English-language editor to bring it to an acceptable standard.

	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	All elements are present, and are well summarised - not overly detailed, but presenting just enough information to provide the reader with a comprehensive overview of the context, the evaluation itself, and the findings.
	iii	Includes all significant information in a concise yet clear manner to understand the theme, intervention, programme, project and the evaluation.	Yes	Yes, the section includes all significant information and summarises well the different sections of the report to present the reader with a coherent, concise, well-structured summary that (notwithstanding issues with the quality/quantity of corroborating evidence noted below) is likely useful to UNFPA.
SECTION B:		**BACKGROUND (weight 5%)**	80%	**Comments on Rating**
Question 2.		**Is the evaluand (i.e. intervention/policy/thematic area etc. that is to be evaluated) and context of the evaluation clearly described?**		
	i	Clear description of the evaluand (e.g. intervention), including: geographic coverage, implementation period, main partners, cost/budget, and implementation status.	Yes	Section 1.3 offers a clear and comprehensive overview of the ASR regional programme, covering key elements such as the time period, partners, and budget. However, the paragraph on the programme's implementation status at the time of the evaluation (paragraph #32) is notably brief but this is justified in the text, in line with the formative nature of the evaluation. It simply notes the main intervention areas and any progress or notable gaps, without much detail.
	ii	Clear description of the context of the evaluand (e.g. economic, social and political context, relevant aspects of UNFPA's institutional, normative and strategic framework, cross cutting issues such as gender equality and human rights, disability and LNOB dimensions) and how the context relates to the evaluand (e.g. key drivers and challenges that affect the implementation of the intervention/policy/thematic area	Partially	There is a quite comprehensive summary of the overall regional context provided in section 1.2. It covers some of the major socio-economic and political aspects of the region, with good examples drawn from individual countries and focusing on those that specifically relate to the UNFPA mandate. However, there are some areas for improvement.\nFor example, in the discussion of the economic background to the region, the formulation is poor - the report speaks of "national" poverty lines/threshold, but is referring to the region. Further, using regional averages for poverty (para 7) is problematic, given that the region is home to some very wealthy countries (UAE, Bahrain, Kuwait) - it would make more sense to provide a more nuanced analysis on individual countries in this context. Regional averages would make more sense in other areas (such as women's empowerment, harmful practices etc.) but even still, the enormous disparities in the region may skew the analysis.\nFurthermore, leaving no one behind considerations are not sufficiently discussed even though the regional programme was designed to reach the furthest left behind aligned with the UNFPA global strategy.\nIn addition, many of the statements are unreferenced and require footnotes to explain the source.\nFinally, at least some of the text in the section is copied/pasted directly from source documentation (e.g. para 15, pg. 4) - this is not good practice and should be addressed.
	iii	Linkages drawn between the evaluand and the ICPD benchmarks and SDGs relevant targets and indicators.	Yes	The description of UNFPA responses in the ASR clearly links the global UNFPA strategy (the 3TRs) to specific SDGs in section 1.3 (para 24) and also notes realization of the ICPD programme of action as an aspiration for UNFPA. This is linked to the ASRO programme strategy in subsequent paragraphs. No further details are provided on specific SDG targets or the ICPD benchmarks but this is understandable, considering the nature of the evaluand (a regional strategy).
Question 3.		**Are key stakeholders clearly identified and analysed?**		
	i	Clear identification of key stakeholders which should include implementing partner(s), development partners, rights holders, and duty bearers among others; and of linkages between them (e.g., stakeholder map).	Yes	There is a brief description of the key stakeholders in the regional strategy presented in section 1.3. Paras 29 and 30 cover the institutional stakeholders, essentially listing the main partners and noting the broad areas where they might contribute. However, this appears to be primarily focused on Output 1 of the SP (policy and accountability), with more limited reference to the other output areas. In addition, Section 1.2 provides useful information on the status of the rights-holders that UNFPA works with and for.
	ii	Stakeholders are analysed to understand their specific rights, duties, needs, interests, concerns, and potential impact on the evaluand.	Partially	As described above, the description of stakeholders is the background section is brief - there is no substantive analysis or description of their interaction with UNFPA beyond noting some general areas in which they cooperate with UNFPA at the regional level.
SECTION C:		**EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)**	100%	**Comments on Rating**
Question 4.		**Is the purpose of the evaluation clearly described?**		
	i	Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.	Yes	The purpose is well and concisely described in the Evaluation Objectives and Scope section (1.1). This has been taken directly from the TOR, so there is no deviation. Given the completion of the current strategic plan period, it is clear that the evaluation is now required and will provide useful and important evidence/information regarding the realization of the previous strategy and that UNFPA will use the findings to build the future strategy. The audience for the evaluation is also stated under the "Objective" subheading - this should really have its own subheading.
Question 5.		**Are the objectives and scope of the evaluation clear and realistic?**		
	i	Clear and complete description of the objectives of the evaluation, including reference to any changes made to the objectives included in the ToR (if applicable).	Yes	The evaluation objective, directly taken from the IR, is clearly presented in the first section of the report. This is helpful as it allows the reader to immediately grasp the purpose and scope of the evaluation, as well as its intended outcomes.
	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	As with the purpose and objective, section 1.1 has a clear and concise description of the scope, comprising the thematic, temporal and geographic elements. These are clearly delineated and it is equally clear that there are no exclusions. A good practice is the noting (in the thematic scope) that the evaluation is formative ("forward-looking") in order to maximise the benefit to the planned strategy. This is expanded on in section 1.3 (para 32) which notes that less emphasis was given to evaluation of the "implementation status and progress of indicators" of the RP - some consideration to noting this under the scope might be given, to ensure comprehensiveness.
SECTION D:		**EVALUATION DESIGN AND METHODOLOGY (weight 20%)**	60%	**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?**\n\n*Note: UNFPA evaluation standards refer to the OECD/DAC criteria such as: relevance, coherence, effectiveness, efficiency and sustainability (not necessarily applicable to all evaluations) and, for country programmes that include circumscribed and limited humanitarian and/or emergency interventions, the criteria of coverage and connectedness.*		

	i Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Yes	The report includes the list of finalised evaluation questions derived from the original set in the TOR. These have been arrived at via a clearly-documented process with the evaluation management, which is positive to see. One relatively minor issue is that the list of questions presented in para 35 is quite duplicative with the list of evaluation questions and sub-questions on the next page. Suggest simply removing this to avoid the repetition (and making sure the evaluation criteria are included in the table). Beyond this, the evaluation questions appear relatively robust to meeting the objective of the evaluation and are clearly aligned with six OECD/DAC evaluation criteria, although EQ1 might also include the efficiency criterion as it reflects technical and operational support to Country Offices (COs).
	ii Evaluation matrix clearly presents the evaluation criteria used as well as the corresponding evaluation questions, indicators, lines of inquiry, benchmarks, assumptions, source of data, methods for data collection and analysis, and/or other processes from which the analysis can be based, and conclusions drawn.	Partially	The evaluation matrix (included in annex 3) does clearly present the evaluation questions, criteria and subquestions, although there is some discrepancy between the list of subquestions at the start of section 2.1 (part of 1.2 is missing in the matrix). Further, many of the subquestions are, in fact, multiple subquestions - e.g. 1.1 has two components and should be two separate subquestions, 1.2 has three. As a best practice, each of these should be divided into distinct sub-questions for greater clarity and focus. Further, the formulation of the evaluation matrix after the subquestions is questionable. Each EQ is assigned assumptions to be tested. Normal good practice, as per the standard evaluation matrix in the UNFPA Evaluation Handbook, is that each subquestion will give rise to one or more assumptions, which in turn will be parsed into indicators or benchmarks. In this matrix, however, assumptions are listed underneath the suite of EQ/subEQs and are very few. For example, EQ1, with two subquestions (in actuality 4-5) has one assumption associated with it, and four rather basic indicators. For EQ2, there are six subquestions (in actuality 10-12 separate areas of inquiry) but only five assumptions to be tested. This formulation appears to substantially simplify the analytical approach, rather than systematically break down ALL of the subquestions into measurable or evaluable elements as is normal practice. As the matrix was designed at inception, it is assumed that this formulation was approved by the evaluation management, but it is not standard evaluation practice. Finally, insofar as the assumptions are presented, they do include illustrative sources of information and data collection methods (although the "deep dive" cited as a method is not a typical evaluation methodology, so is questionable), but the effectiveness of the analytical framework in ensuring the evaluation meets its objective is doubtful.
Question 7.	**Is the theory of change, results chain, logical framework, or equivalent framework well-articulated?**		
	i Clear description of the intervention's intended results, or of the parts of the results chain that are applicable to, or are being tested by, the evaluation.	Not Rated	The evaluation report specifically notes (in section 1.3, para 32) that due to its future-looking and formative nature, less emphasis was given to evaluation of the "implementation status and progress of indicators" of the RP. This is also noted in section 2.2, para 37 which notes that the evaluation did not assess progress towards the 3TRs "nor use a theory of change". Therefore this section is not rated.
	ii Causal relationships between the various elements (e.g. outcomes, including the three or relevant Transformative Results, outputs) of the theory of change, results chain or logical framework are presented in narrative and/or graphic form).	Not Rated	As above, the evaluation explicitly notes it did not analyse the relationships between results elements, specifically the 3TRs, although they are core elements of at least two of the evaluation questions (#s 2 and 3).
	iii Comprehensive analysis and assessment of the theory of change, results chain or logical framework, and if requested in the ToR, it is retrofitted/reconstructed by the evaluators.	Not Rated	As noted, the evaluation explicitly did not use a TOC or results chain due to (a) the formative nature of the evaluation and (b) the findings of the mid-term review of the RP which found that the original TOC was still relevant (discussed in section 2.2 para 37).
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.	Partially	While the outline of the methods presented in the main report is useful and should be robust if applied correctly, the supporting detail in the annexes is not adequate. The evaluation methods are summarised in brief in the main report section 2. This is supported by more comprehensive annexes (2, 4) which repeat much of the conceptual background and list out the data collection locations and tools. There are some issues with the methods. Firstly, the language used in describing the approaches and tools is dense, which makes the methods difficult to understand by the reader. Secondly, a description of the methods and tools on pg. 10 of the annexes notes the adoption of "a sequenced approach", but then that "methods will be executed concurrently" - these two statements are contradictory. Thirdly, there are various representations of the conceptual approach, including the use of two figures (figures 6 & 7 in Annex 2) which do not simplify, rather add to the complexity (for the latter) or appear simply irrelevant (for the former). There are other inconsistencies within the annexes or between the annexes and the main report. Finally, the use of AI is noted as supporting the evaluation process in data transcription and analysis, although the specific tool has not been noted (as should be the case). Much reference is made to ensuring compliance with ethical safeguards in this regard but few details on this are provided (i.e. HOW and WHAT safeguards were applied).
	ii Data sources are all clearly described and are relevant and robust; these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR).	Partially	The sources are described reasonably well in the main report, with the supporting detail in the annexes providing further information. The use of KIIs, FGDs (actually group interviews) and desk review documentation, if applied correctly and to the sample of stakeholder noted in the stakeholder map (as appears to be the case from the list of stakeholders consulted in the annexes), are relevant and should lead to robust findings.
	iii Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or if not, provide reasons for this).	Partially	The stakeholder map is contradictory with the text, which states that no rights-holders were sampled. However, the stakeholder map does seem to incorporate these elements, including "Recipients of project interventions in multiple locations (SRHR; GBV case management, CP, and others)," and even features a specific category for "refugees and IDPs." This highlights an inconsistency in the mapping process. The sampling process for case study countries is only partially described - criteria are noted, but the description is not clear and it is not evident from the text how the criteria were applied or who applied them (it is implied that it is the evaluation team, but it would have been more appropriate for the evaluation reference group to be significantly involved here).
	iv Methods allow for rigorous testing of the theory of change, results chain or logical framework (e.g. methods help to understand the causal connections, if any, between outputs and expected outcomes (3TRs).	Not Rated	As noted above, testing of the TOC and/or results framework (i.e. indicators) was explicitly excluded from this evaluation, despite its claims to be a "theory-based evaluation".

	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	The main report provides a basic description of the methods of analysis (para 46), but, as with much of this section, substitutes high-sounding and repetitive jargon for clear and concise description (e.g. "an iterative and multi-phased approach"). The annex merely expands on this trend, with few concrete details on, for example, the coding process for qualitative data, and repetition of what is said in the narrative in the supporting visual graphics. Once the overly-complex language is parsed, a lack of useful details emerges. The evaluators do note the use of AI in the data analysis, though (as noted above) do not specify the AI package (e.g. ChatGPT, MaxQDA, Atlas.ti) that is used, so despite numerous assurances of compliance with UNFPA evaluation standards around use of AI, it is not possible for the reader to find out how, or to what extent, such compliance will be achieved.
	vi	Clear and complete description of limitations and constraints faced by the evaluation in its data collection and analysis, including gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators (as feasible).	Partially	There is a useful section on limitations presented in the main report (which notes additional details in Annex 7, but this is incorrect - there are no additional details and Annex 7 is an evaluability assessment). This presents four potential limitations, mitigation strategy employed with the ultimate outcome on the evaluation also described. For the limitations presented, this is done well. However, a key and obvious limitation not noted is the absence of the voices of rights-holders in the evaluation. This should have been addressed.
Question 9.		**Are ethical issues and considerations described?** **The evaluation should be guided by the UNEG ethical standards for evaluation. As such, the evaluation report should include:**		
	i	Explicit and contextualized reference to the UNEG obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) and/or UNEG Ethical Principles.	Partially	The report clearly notes compliance with the UNFPA Evaluation Policy and the UNEG Ethical Guidelines (2020 version, linked to in a footnote). The short section makes explicit reference to the UNEG obligations, but they are listed without being contextualised. No further details on these are presented in the annexes.
	ii	Clear description of ethical issues and considerations (e.g. respect for dignity and diversity, fair representation, confidentiality, and avoidance of harm) that may arise in the evaluation, safeguard mechanisms for respondents (e.g. parental consent forms for adolescents, compliance with codes for vulnerable groups; WHO standards of safe data collection on GBV and ethical considerations in the use of AI as applicable (e.g., transparency of use, explainability, privacy, data protection, accuracy, human rights). If AI is used in the evaluation, there should be transparency and disclosure on the ethical and responsible use of AI in the report.	Partially	As no rights-holders were (apparently) involved in the evaluation, many of the fundamental ethical considerations in relation to doing no harm etc. do not apply to the same level of rigour as an evaluation that did include these stakeholders. The short section (2.4, paras 48-50) in the main report provides a useful and clear description of the primary ethical considerations that would pertain in an evaluation of this nature. Annex 2 provides some additional detail (and repeats much of what is in the main report). The primary data collection tools (annexed) include appropriate consent-seeking language in their preamble. The ethical issues around use of AI are noted and the report states that the evaluators adhered to ethical considerations in its use. However, much of the relevant text is simply repetition of the same thing, with little actual further details on how such compliance was achieved. This is discussed at length above.
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	In as much as AI has been used in the evaluation for data transcription and analysis, innovative practice has been incorporated. However, as noted above, the report is absent any further specifics on how the AI was used, what the actual AI model or software package that uses AI is, so it is not possible to determine whether the application was truly innovative, or the evaluators used it for basic editing tasks, which, while still innovative, is becoming rapidly mainstream.
SECTION E:		**EVALUATION FINDINGS (weight 25%)**	50%	**Comments on Rating**
Question 11.		**Do the findings clearly and adequately address all evaluation questions and sub-questions?**		
	i	Findings are presented clearly and provide sufficient levels of evidence to systematically address all the evaluation's questions	Partially	Despite the large number of subquestions to be addressed by the evaluation, a strength of this report is that the evaluators systematically seek to address all of the evaluation questions and subquestions as laid out in the evaluation matrix (or at least in the table in section 2.1 - there are some minor discrepancies between that and the matrix, noted above). Each findings subsection begins with a list of the questions to be addressed, which is very useful for readability. It is clear that the authors have sought to systematically work through the strategic areas, EQs and sub-questions and present data under each (and should be commended for this). Findings are listed clearly and well numbered and there is extensive discussion under each, although one issue is that some of the findings are very long and, in reality, contain multiple finding (e.g. Finding 4 really has 3-4 separate findings within one). In terms of the sufficiency of evidence under each area, the report has been less successful. The report does discuss achievements and presents evidence across findings and specific programmatic areas, but the evidence base is thin in places. This is discussed further under 12ii below.
	ii	Explicit use of the evaluand's theory of change, results chain, logical framework in the formulation of the findings.	Not Rated	The theory of change/results chain has been explicitly excluded from the evaluation, due to a desire to focus on the formative rather than summative aspects.
Question 12.		**Are evaluation findings derived from credible data sources as well as a rigorous data analysis?**		
	i	Evaluation uses credible forms of qualitative and quantitative data. It presents both output and outcome-level data as relevant to the evaluation framework. Triangulation is evident using multiple data sources.	Partially	The report effectively utilizes all planned data sources, both qualitative and quantitative, in formulating its findings. It references qualitative data gathered from interviews conducted across the region, although this is not always done consistently or sufficiently (as noted below). Additionally, the report combines this qualitative data with secondary data from documentary reviews, albeit with some issues in referencing. It also makes good use of quantitative data derived from the Mid-Term Review (MTR) of the regional strategy, particularly from a survey conducted for that review. Overall, this approach is commendable. However, the analysis tends to focus primarily on basic output-level discussions rather than providing a deeper evaluation. As highlighted below, there is a tendency to list activities extensively, whether at the regional or, to a lesser extent, country level, without adequately analyzing their contribution to broader regional outcomes. While there are occasional references to specific planned outcomes or related indicators—such as the regional indicator for innovation mentioned in section 3.2.4 (para 118)—it remains difficult to understand how the various activities and outputs collectively contribute to achieving these regional outcomes. There are plenty of positive examples of good use of data and triangulation. For example, under EQ2, (section 3.2, para 90 onwards), despite some initial unreferenced assertions (paras 87-89) the evaluators combine analysis of CPDs for articulation of UNFPA accelerators with primary data from country-level stakeholders and secondary survey data from the MTR to present a relatively robust finding.

	ii	Findings are clearly supported by the evidence presented, both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Partially	The evaluation is a mix of positive and poor performance in this regard. There are some quite questionable evidence/sub-findings, for example citing the distribution of dignity kits and the rollout of mobile clinics as RO achievements (these are surely fully the responsibility of the CO - the RO contribution to these is not made clear). In many cases, evidence is not presented against assertions of achievement. For example, under 3.1, the statement that "when the CO reaches out for technical documents or funding assistance, ASRO responds promptly" is not supported by any evidence for how the evaluators know this. Another example is under finding 3.2.3, related to the second UNFPA accelerator (HR and Gender Transformative approaches). The finding states (in part) that both HR and Gender Transformation are "taboo" in many countries in the region. However, the evidence presented beneath does not explain this in any way (it discusses maternal health, which is only peripherally related, and goes on to cite positive country-based examples of gender work), and indeed there is plenty of secondary evidence available via a cursory search to contradict, at the very least, part of the assertion (related to HR). A final, significant, example is in section 3.4.1 (para 175) which asserts that "unusually high" numbers of UNFPA staff are under (unspecified) internal investigation by OAIS. This most definitely needs to be explained further (bit is not). If true, it clearly implies that the "best possible" candidates are NOT hired, contrary to the finding statement immediately above. While many assertions are indeed supported by a source (although the report will often refer to "COs" as sources, rather than specific key informants, even anonymised), there are many examples of unattributed or inadequately evidence assertions throughout the report. It frequently reads like a list of activities or achievements, with little in-depth analysis of what these mean or where they are derived from, nor by recourse to a reflection on performance targets or benchmarks. Another example is under 3.2 (para 95, where the report asserts that "ASRO enjoys a strong relationship with the League of Arab States" but no evidence is presented to corroborate this - despite having conducted interviews with representatives of the LAS which could be used to triangulate with perspectives from UNFPA and the secondary evidence of joint activities. The report then goes on to discuss examples of CO relationships with individual governments - but how this might relate to a regional strategy is not clear.
	iii	Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings analyse the logical chain (progression -or not- from outputs to high level results).	Partially	Although this is not an output/outcome/impact evaluation, the report examines the extent to which 'enablers' facilitated the implementation of the accelerators and strategic shifts. The enablers were applied to explore issues that can explain what worked and did not in operationalizing the shifts and accelerators. However, in some cases, the analysis insufficiently unpacks potentially interesting sub-findings or evidence that could provide useful insights for UNFPA moving forwards. For example, under EQ2 (section 3.2, para 89) the report notes that "advancing innovations for population data has been slow, with inconsistencies in data collection methods" - this could have been explored with reference to individual countries and linked to any UNFPA regional strategy to harmonise population data strategies across countries (if it exists - if not, then inconsistencies can hardly be cited as a lack of progress). The area is explored further in section 3.2.7, but still causal factors are very lightly touched on).
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	The ToR did not require an assessment of the RBM/MEAL components of the regional programme because it was aligned with the global Strategic Plan evaluation
SECTION F:		**EVALUATION CONCLUSIONS (weight 10%)**	75%	Comments on Rating
Question 14.		**Do the conclusions clearly present an unbiased overall assessment of the evaluand?**		
	i	Conclusions are clearly formulated and present unbiased summative statements that respond to the evaluation questions.	Yes	The nine conclusions are clearly and concisely stated, with additional narrative explanation below that summarises the findings on which they are based (with the specific findings related to each itemised). They also clearly point the way for future actions, as presented in the recommendations and note the specific recommendations to which they pertain. However, the conclusions tend to be somewhat vague and lack the necessary detail to enhance their applicability.
	ii	Conclusions are well substantiated and derived from findings and add deeper insight and analysis beyond the findings.	Partially	While the conclusions effectively reflect the findings of the analysis, they are well-founded and accurately summarize both the achievements and challenges, as previously noted. However, they are based on the analysis presented in the findings section, which has some significant shortcomings and thus they should be treated with caution. Further, they are simply a summary of the main findings, and do not offer and more significant insight. Notably, they miss the "big picture" of how the findings point towards achievement of the Regional Programme under evaluation. In at least one case (conclusion 7), the evidence base for part of the conclusion narrative is questionable, notably around UNFPA leadership deficits - this is only barely reflected in the findings, with very little (or no) corroborating evidence provided in the relevant section (3.4.1, para 177).
Question 15.		**Are lessons learned identified? [N/A if lessons are not referenced or requested in ToR]**		
	i	Lessons learned are derived from the findings and are well substantiated with practical, illustrative examples.	Not Rated	Firstly, there is no standalone lessons-learned section in the report. This has not been explicitly noted in the TOR as being required. Nonetheless, the evaluation subquestions include provision for lessons - specifically under EQ1.2 "What are important good practices, bottlenecks and lessons learned?". The analysis framework and data collection tools explicitly integrate this. As such, the report does note quite a number of lessons or good practices under the relevant section (3.1.2) - many of these are indeed derived from specific examples from country level.
	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	The section most akin to a 'lessons-learned' (3.1.2) is relatively clearly presented, and does provide good information on good practices and challenges faced that could be leveraged for action in other contexts. Some of the statements suffer from a key weakness of the overall analysis - poor attribution of the sub-findings to individuals or sources of evidence, which makes it challenging to determine (a) the veracity or robustness of the statements, and (b) the extent to which the practices could (or should) be replicated or scaled. Given the formative nature of the evaluation, it would have benefitted from a standalone Lessons/Good Practices section with specific guidance on the nature of the lessons/practices being sought (e.g. specifically for the RO, by sectoral area/TR/accelerator etc.).
SECTION G:		**EVALUATION RECOMMENDATIONS (weight 15%)**	75%	Comments on Rating
Question 16.		**Are recommendations well-grounded and articulated?**		

	i	Recommendations are clearly formulated and logically derived from the findings and/or conclusions.	Partially	The ten recommendations are quite clearly formulated and are generally well-aligned with the purpose of the evaluation and a reasonable reflection of the findings, particularly the sub-recommendations (the recommendations statements themselves are short - this is good practice when accompanied by sufficiently detailed action points). This said, the recommendations do not cover all notable elements of the findings. For example, the LNOB issues discussed under 3.2.5 are not reflected in the recommendations (though are briefly noted in the conclusions). Some more reflection on them may be warranted. Of note is a certain amount of overlap between the recommendations section and section 3.6.3 (Key Considerations for the Next Regional Programme) - this section includes, essentially, a list of recommendations and is therefore quite duplicative with the dedicated recommendations section. It is a design issue, rather than one inherent to the report, and should have been addressed at inception.
	ii	Recommendations are useful and actionable for primary intended users. Specific guidance is provided for its implementation (e.g. actions, deadlines, responsible actors), as appropriate.	Yes	The recommendations are useful and (when taken with the actions) are adequately actionable, with specific guidance provided within the action points. As noted below, ASRO and COs are collectively deemed responsible for all recommendations, which should be nuanced further (e.g. by assigning responsibility for individual actions). Further, the recommendations are grouped into "Strategic" and "Programmatic" - some of the "Strategic" recommendations (e.g. around staff recruitment/retention = #s 3, 5, 6) are more in line with an "Operational" category - consideration to reorganising according to this category is recommended.
	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 recommendations section notes that they were validated by the Evaluation Reference Group and Regional Office staff via a co-creation workshop. The ERG included three external stakeholders, but it is not noted whether these participated in the recommendations workshop.
	iv	Recommendations are clearly articulated and prioritized based on their importance, urgency, and potential impact.	Partially	The preamble to each of the ten recommendations includes responsibility for the recommendation, related conclusion and prioritisation, but no timeframe. However, the responsibility for the sub-recommendations is allocated for all to "ASRO and COs" - there are clearly nuances to each of the action points that make them more suitable to each of these stakeholders, which could have been explored further.
SECTION H:		**REPORT STRUCTURE AND PRESENTATION (weight 5%)**	75%	**Comments on Rating**
Question 17.		**Does the evaluation report include all required information?**		
	i	Opening pages include: Name of evaluation and/title of evaluation, timeframe of the evaluation, date of report, location of evaluand, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents (including, as relevant, tables, graphs, figures, annexes)-; list of acronyms/abbreviations.	Yes	All of the specified elements in the criterion are present, although the list of ERG members (normally with the evaluation managers/implementers) is not included here (nor in the annexes, other than as part of the list of key informants). One issue on the opening pages is that a page number is not normally included on the cover page. The page numbering can be customised to hide the first page, and set the numbering to start at "0".
	ii	Annexes include, if not in body of report: terms of reference, evaluation matrix, list of respondents, results chain/ToC/logical framework, list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology (e.g. inception report), case study reports.	Yes	There is a full suite of annexes present with all of the required elements plus the additional country briefs.
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 structure is sound, with fully numbered sections and subsections. The formatting is largely good, but there are some section/subsection headings at ends of pages (that should start on the following page). Those noted by the reviewer have been highlighted in the reviewed report which accompanies this template.
	ii	Structure and length accords to UNFPA guidelines for evaluation reports; it does not exceed number of pages that may be specified in ToR. *Note: Maximum pages for the main report, excluding executive summary and annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations and 50 for other types of evaluations)*	Partially	The main report is 75 pages, which is slightly outside the guidance provided (no page length is specified in the TOR).
Question 19.		**Is the report well presented?**		
	i	Report is easy to understand (written in an accessible way for the intended audience) and generally free from grammar, spelling and punctuation errors.	Partially	While the writing is generally acceptable and understandable, there are still a variety of challenging formulations and some basic errors across the text. Many of these have been noted via in-line commentary, but specific examples include: - Some acronyms are not spelled out at first usage. - There are a variety of incorrect formulations, grammar/syntax/style mistakes - Over-usage of passive voice ("it has been reported" etc.) - Incorrect use of numerals (1,2,3...) vs. number spelling (one, two, three etc.) as per the UN/UNFPA style guidelines. - Inappropriate/excessive use of Capitalization of some words, positions etc. - see the UNFPA style guide for direction. - Excessively superlative/hyperbolic language (e.g. "The methodology is intricately crafted to harmonise with the overarching objectives..." instead of "The methodology is designed to meet the objectives". - Tenses are incorrect, particularly in the annexes (present rather than past tense).
	ii	Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labelled, and referenced in text.	Partially	The report relies principally on qualitative data, so there is little scope for the use of infographics and charts. However, it does make good use of charts (primarily in the background/context, but also to present some secondary quant data from the MTR survey), as well as the highlighting of key findings in boxes, which enhances readability. Some of the charts are formatted poorly (E.g. figure 1 - country names are not on one line) and the page numbers of the figures and table are not included in the Table of Contents (MS Word offers this functionality). Finally, for one figure (e.g. figure 13) there is little explanation of what the data means, and for one (figure 10), the meaning is very obscure ("partnerships to not continue doing business as usual").
SECTION I:		**CROSS-CUTTING ISSUES (weight 10%)**	57%	**Comments on Rating**
Question 20.		**Are cross cutting issues - in particular, human rights-based approach, gender equality, disability inclusion, LNOB - integrated in the core elements of the evaluation (e.g. evaluation design, methodology, findings, conclusions and recommendations)?**		

	i	Evaluation's data collection methods designed to capture the voices/perspectives of a wide range of stakeholders including right holders, marginalized and vulnerable persons, young people, people with disabilities, migrants or refugee populations, indigenous communities, and other persons that are often left behind.	Partially	Firstly, the report is a high-level evaluation of a regional strategy, which, by its nature, has less scope for involvement of rights-holders. Indeed, in the methods section, the report notes "it was agreed that it would be less beneficial to engage with end beneficiaries (such as GBV survivors and refugees) hence this category of rights holders were not included in the stakeholders' map". However, the stakeholder map does appear to include these: "Recipients of project interventions in multiple locations (SRHR; GBV case management, CP, and others" - there is even a specific category "refugees and IDPs". There is inconsistency here which needs to be addressed and the rationale for absence of rights-holder voices clearly explained.
	ii	Evaluation questions address cross cutting issues, such as human rights-based approach, gender equality, disability inclusion, LNOB, social and environmental standards as appropriate.	Yes	The evaluation questions clearly include gender transformation and LNOB as specific subquestions under the UNFPA accelerators. Under these (and specifically LNOB) issues of disability and other vulnerability (such as age) are explored.
	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).	No	There is no disaggregation of stakeholders by gender. This should have been incldued in the report. However, with a focus on institutional strategy, the evaluation did not collect primary or secondary data in relation to population groups, and thus disaggregation was not relevant.
	iv	Intersectional lens is applied in the data analysis, looking at various and multiple forms of exclusion and discrimination (and how they overlap with each other) and how this may impact the performance or results of the evaluand.	Partially	There is very brief explicit acknowledgement of the intersectional nature of the work that UNFPA undertakes and that intersecting vulnerabilities affect target populations. This is noted under the "Megatrends" analysis (3.3.1) in relation to climate change specifically, and again in relation to overall achievement of the 3TRs (3.6.1). However, the analysis simply notes the presence of such overlapping issues, without exploring it in any detail. The conclusions, again, note this (conclusions 8 and 9) but without any detail, and the recommendations do not address the issue.
	v	Findings, conclusions and recommendations, address cross-cutting issues such as equality and vulnerability, disability inclusion, leave no-one behind, social and environmental as relevant.	Partially	As noted above, under the UNFPA accelerators analysis (and specifically LNOB) issues of disability and other vulnerability (such as age) are explored. Issues of human rights and social/environmental standards are less explicitly covered, although environment is addressed via climate change as part of the "megatrends" analysis. As noted above, the report could go further to address these issues, for example around LGBTQ+ inclusion - which is a notably challenging issue in the region, but with some positive examples in the region (specifically Lebanon) from which other countries could learn (and the RO provide an important facilitation role).
	vi	Inclusion of young people in the evaluation team and/or Reference Group [N/A if not requested in ToR]	Yes	The evaluation included a "young and emerging evaluator" on the team, noted in the cover pages.
Question 21.		**Does the evaluation meet UN SWAP evaluation performance indicators?** *Note: this question will be rated according to UN SWAP standards with detail provided below*	5	**Comments on Rating**
	i	GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Fully integrated	Gender equality/equity, as one of the fundamental principles of UNFPA was fully included in the evaluation scope at the outset and is integral to the evaluation questions (either directly via exploration of gender transformation or indirectly via the RP outputs and 3TRs). This was carried over to the evaluation analysis via the evaluation matrix.
	ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Partially integrated	The evaluation methodology does not provide extensive detail on the level of gender-responsiveness of the evaluation. The description of the methods merely notes that data collection methods will "integrate[e] gender considerations". The gender integration outlined in the Inception Report is similarly brief/generic in this regard (and notes that all data will be disaggregated by sex and "analyse[ed] through a gender lens", but does not specify how this will be done). The tools themselves do not specifically mention "gender", although clearly explore this dimension via achievement of the 3TRs, the RP outcomes and the UNFPA accelerators.
	iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Partially integrated	Again, the conclusions and recommendations do not explore gender dimensions in any explicit detail, relying more on the indirect reference to the 3TRs and other SP/RP elements. Gender is only mentioned once in the recommendations (#9) - in the context of mainstreaming social norms into gender (and youth) programming.

i GEEW is integrated in the evaluation scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.
a. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?
b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?
c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?
d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?

ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.
a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex?
b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations?
c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?
d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?
e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality?

iii The evaluation findings, conclusions and recommendations reflect a gender analysis.
 a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality?
b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable?
c. Are unanticipated effects of the intervention on human rights and gender equality described?
d. Does the evaluation report provide specific recommendations addressing GEWE issues, and priorities for action to improve GEWE or the intervention or future initiatives in this area?

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

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

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

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